The Economics of Language

International analyses

Barry R. Chiswick and
Paul W. Miller

Routledge Studies in the Modern World Economy

The Economics of Language

This book is primarily concerned with the determinants of dominant language proficiency among immigrants and other linguistic minorities, and the labour market consequences of this proficiency. Chiswick and Miller present a selection of essays from their Economics of Language research programme, published from 1988 to 2005.

Concentrating on the United States, Canada, Australia, Israel and Bolivia, the authors use economic theory to develop models of the determinants of dominant language proficiency. They use econometric techniques to test these models and estimate the magnitude of the effects. The research demonstrates that dominant language proficiency is greater, with more exposure to the dominant language, with greater efficiency in acquiring dominant language skills and the greater economic benefits from proficiency. The effects of dominant language proficiency on labour market outcomes are also estimated.

This book will be invaluable for anyone engaged in the contemporary issues regarding the adjustment of immigrants in their host economy and society, as well as students of economics, sociology and linguistics. It will also be invaluable for those interested in bilingualism and the assimilation of ethnic/linguistic minorities.

Barry R. Chiswick is Distinguished Professor at the University of Illinois at Chicago.

Paul W. Miller is Professor of Economics at the University of Western Australia.

Praise for *The Economics of Language*

This book contains an impressive collection of essays on the economics of language. Barry Chiswick and Paul Miller have addressed a wide range of areas on both the determinants of language proficiency, and the consequences of language proficiency for labour market outcomes. They offer a broad and international perspective. The book provides careful and insightful analysis on issues relating to language and immigration and will be a great resource for researchers and graduate students working in this area.

Professor Christian Dustmann, Department of Economics and Centre for Research and Analysis of Migration (CReAM), University College London

Barry Chiswick and Paul Miller have contributed greatly to an interdisciplinary understanding of the mechanisms, conditions, and consequences of language acquisition and the processes of migrant integration as a whole. Their work is a break-through in many respects.

Professor Hartmut Esser, Department of Sociology, University of Mannheim

In *The Economics of Language: International Analyses*, Barry Chiswick and Paul Miller have put together their finest articles on the topic. The book is nicely organized around important topics. The theoretical foundations for their empirical work come out strongly, reminding us that it is exposure, efficiency and economic incentives that matter for determining language acquisition. The book will prove of great value to researchers on language issues, showcasing in a most useful way the contribution of economists.

Harry A. Patrinos PhD, Lead Education Economist, The World Bank, Washington DC

Barry Chiswick is one of the most enlightened and insightful economics researchers alive today. Together with his co-author, Paul Miller, he has discovered the enormous importance *language* plays on the map of socio-economics in terms of jobs, immigration patterns and advancement in immigrant-receiving countries. This book is obligatory reading for all concerned with immigrants and the socio-economic aspects of language.

Professor Adam Makkai, Professor of English and Linguistics, University of Illinois at Chicago, Founding Executive Director of LACUS, The Linguistic Association of Canada and the United States

For too long, economics has omitted language, whether language in general or, as in the path breaking work of Chiswick and Miller, languages plural: how a Greek migrant to Australia or a Polish migrant to America gets along by becoming fluent. And linguistics too often omits the economics, provided here in lucid and cosmopolitan bulk. Chiswick and Miller ask how much language matters to immigrant lives. It matters a lot, though in surprising ways. *The Economics of Language* tells the story of the flight from Babel with theoretical depth and quantitative precision.

Deirdre McCloskey, Distinguished Professor in the Departments of Economics, English and History at the University of Illinois at Chicago

Routledge Studies in the Modern World Economy

1 **Interest Rates and Budget Deficits**
A study of the advanced economies
Kanhaya L. Gupta and Bakhtiar Moazzami

2 **World Trade after the Uruguay Round**
Prospects and policy options for the twenty-first century
Edited by Harald Sander and András Inotai

3 **The Flow Analysis of Labour Markets**
Edited by Ronald Schettkat

4 **Inflation and Unemployment**
Contributions to a new macroeconomic approach
Edited by Alvaro Cencini and Mauro Baranzini

5 **Macroeconomic Dimensions of Public Finance**
Essays in honour of Vito Tanzi
Edited by Mario I. Blejer and Teresa M. Ter-Minassian

6 **Fiscal Policy and Economic Reforms**
Essays in honour of Vito Tanzi
Edited by Mario I. Blejer and Teresa M. Ter-Minassian

7 **Competition Policy in the Global Economy**
Modalities for co-operation
Edited by Leonard Waverman, William S. Comanor and Akira Goto

8 **Working in the Macro Economy**
A study of the US labor market
Martin F. J. Prachowny

9 **How Does Privatization Work?**
Edited by Anthony Bennett

10 **The Economics and Politics of International Trade**
Freedom and trade: Volume II
Edited by Gary Cook

11 **The Legal and Moral Aspects of International Trade**
Freedom and trade: Volume III
Edited by Asif Qureshi, Hillel Steiner and Geraint Parry

12 **Capital Markets and Corporate Governance in Japan, Germany and the United States**
Organizational response to market inefficiencies
Helmut M. Dietl

13 **Competition and Trade Policies**
Coherence or conflict
Edited by Einar Hope

14 **Rice**
The primary commodity
A. J. H. Latham

15 **Trade, Theory and Econometrics**
Essays in honour of John S. Chipman
Edited by James C. Moore, Raymond Riezman, James R. Melvin

16 **Who benefits from Privatisation?**
Edited by Moazzem Hossain and Justin Malbon

17 **Towards a Fair Global Labour Market**
Avoiding the new slave trade
Ozay Mehmet, Errol Mendes and Robert Sinding

18 **Models of Futures Markets**
Edited by Barry Goss

19 **Venture Capital Investment**
An agency analysis of UK practice
Gavin C. Reid

20 **Macroeconomic Forecasting**
A sociological appraisal
Robert Evans

21 **Multimedia and Regional Economic Restructuring**
Edited by Hans-Joachim Braczyk, Gerhard Fuchs and Hans-Georg Wolf

22 **The New Industrial Geography**
Regions, regulation and institutions
Edited by Trevor J. Barnes and Meric S. Gertler

23 **The Employment Impact of Innovation**
Evidence and policy
Edited by Marco Vivarelli and Mario Pianta

24 **International Health Care Reform**
A legal, economic and political analysis
Colleen Flood

25 **Competition Policy Analysis**
Edited by Einar Hope

26 **Culture and Enterprise**
The development, representation and morality of business
Don Lavoie and Emily Chamlee-Wright

27 **Global Financial Crises and Reforms**
Cases and caveats
B.N. Ghosh

28 **Geography of Production and Economic Integration**
Miroslav N. Jovanović

29 **Technology, Trade and Growth in OECD Countries**
Does specialisation matter?
Valentina Meliciani

30 **Post-Industrial Labour Markets**
Profiles of North America and Scandinavia
Edited by Thomas P. Boje and Bengt Furaker

31 **Capital Flows without Crisis**
Reconciling capital mobility and economic stability
Edited by Dipak Dasgupta, Marc Uzan and Dominic Wilson

32 **International Trade and National Welfare**
Murray C. Kemp

33 **Global Trading Systems at Crossroads**
A post-Seattle perspective
Dilip K. Das

34 **The Economics and Management of Technological Diversification**
Edited by John Cantwell, Alfonso Gambardella and Ove Granstrand

35 **Before and Beyond EMU**
Historical lessons and future prospects
Edited by Patrick Crowley

36 **Fiscal Decentralization**
Ehtisham Ahmad and Vito Tanzi

37 **Regionalisation of Globalised Innovation**
Locations for advanced industrial development and disparities in participation
Edited by Ulrich Hilpert

38 **Gold and the Modern World Economy**
Edited by MoonJoong Tcha

39 **Global Economic Institutions**
Willem Molle

40 **Global Governance and Financial Crises**
Edited by Meghnad Desai and Yahia Said

41 **Linking Local and Global Economies**
The ties that bind
Edited by Carlo Pietrobelli and Arni Sverrisson

42 **Tax Systems and Tax Reforms in Europe**
Edited by Luigi Bernardi and Paola Profeta

43 **Trade Liberalization and APEC**
Edited by Jiro Okamoto

44 **Fiscal Deficits in the Pacific Region**
Edited by Akira Kohsaka

45 **Financial Globalization and the Emerging Market Economies**
Dilip K. Das

46 **International Labor Mobility**
Unemployment and increasing returns to scale
Bharati Basu

47 **Good Governance in the Era of Global Neoliberalism**
Conflict and depolitization in Latin America, Eastern Europe, Asia and Africa
Edited by Jolle Demmers, Alex E. Fernández Jilberto & Barbara Hogenboom

48 **The International Trade System**
Alice Landau

49 **International Perspectives on Temporary Work and Workers**
Edited by John Burgess and Julia Connell

50 **Working Time and Workers' Preferences in Industrialized Countries**
Finding the balance
Edited by Jon C. Messenger

51 **Tax Systems and Tax Reforms in New EU Members**
Edited by Luigi Bernardi, Mark Chandler and Luca Gandullia

52 **Globalization and the Nation State**
The impact of the IMF and the World Bank
Edited by Gustav Ranis, James Vreeland and Stephen Kosak

53 **Macroeconomic Policies and Poverty Reduction**
Edited by Ashoka Mody and Catherine Pattillo

54 **Regional Monetary Policy**
Carlos J. Rodríguez-Fuentez

55 **Trade and Migration in the Modern World**
Carl Mosk

56 **Globalisation and the Labour Market**
Trade, technology and less-skilled workers in Europe and the United States
Edited by Robert Anderton, Paul Brenton and John Whalley

57 **Financial Crises**
Socio-economic causes and institutional context
Brenda Spotton Visano

58 **Globalization and Self Determination**
Is the nation-state under siege?
Edited by David R. Cameron, Gustav Ranis and Annalisa Zinn

59 **Developing Countries and the Doha Development Round of the WTO**
Edited by Pitou van Dijck and Gerrit Faber

60 **Immigrant Enterprise in Europe and the USA**
Prodromos Panayiotopoulos

61 **Solving the Riddle of Globalization and Development**
Edited by Manuel Agosín, David Bloom, George Chapelier and Jagdish Saigal

62 **Foreign Direct Investment and the World Economy**
Ashoka Mody

63 **The World Economy**
A global analysis
Horst Siebert

64 **Production Organizations in Japanese Economic Development**
Edited by Tetsuji Okazaki

65 **The Economics of Language**
International analyses
Barry R. Chiswick and Paul W. Miller

The Economics of Language

International analyses

Barry R. Chiswick
and
Paul W. Miller

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2007
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
270 Madison Ave, New York, NY 10016

Routledge is an imprint of the Taylor & Francis Group, an informa business

This edition published in the Taylor & Francis e-Library, 2007.

"To purchase your own copy of this or any of Taylor & Francis or Routledge's
collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk."

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
A catalog record for this book has been requested

ISBN 0–203–96315–6 Master e-book ISBN

ISBN10: 0–415–77181–1 (hbk)
ISBN10: 0–203–96315–6 (ebk)

ISBN13: 978–0–415–77181–8 (hbk)
ISBN13: 978–0–203–96315–9 (ebk)

Contents

To our spouses, Carmel and Tram
and our children,
Abraham and Benjamin
and
Erin and Andrew

Acknowledgements

The authors and publisher wish to thank the following who have kindly given permission for the use of copyright material.

The American Enterprise Institute for Public Policy Research, Washington, D.C. for article: 'Language in the Immigrant Labor Market', in Barry R. Chiswick (ed.), *Immigration, Language and Ethnicity: Canada and the United States*, Washington: American Enterprise Institute, 1992, pp. 229–296, and 471–476.

Blackwell Publishing for articles: 'Language Skill Definition: A study of legalized aliens', *International Migration Review*, **32**(4), Winter 1998, pp. 877–900; 'Schooling, Literacy, Numeracy and Labour Market Success' (Barry R. Chiswick, Yew Liang Lee and Paul W. Miller), *Economic Record*, **79**(245), June 2003, pp. 165–181; 'Do Enclaves Matter in Immigrant Adjustment?' *City and Community*, **4**(1), March 2005, pp. 5–35; 'The Effect of Linguistic Distance and Country of Origin on Immigrant Language Skills: Application to Israel' (Barry R. Chiswick, Michael Beenstock and Gaston L. Repetto), *International Migration*, **39**(3), September 2001, pp. 33–60.

Elsevier for articles: 'Language and Labor Supply: The role of gender among immigrants in Australia', *Research on Economic Inequality*, **5**, 1994, pp. 153–189; 'The Linguistic and Economic Adjustment of Soviet Jewish Immigrants in the United States, 1980–2000' (Barry R. Chiswick and Michael Wenz), *Research in Labor Economics*, **24**, 2006, pp. 179–216; 'Earnings in Canada: The roles of immigrant generation, French ethnicity, and language', *Research in Population Economics*, **6**, 1988, pp. 183–228.

Multilingual Matters for article: 'Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages', *Journal of Multilingual and Multicultural Development*, **26**(1), 2005, pp. 1–11.

Population Association of America for article: 'A Model of Destination-Language Acquisition: Application to male immigrants in Canada', *Demography*, **38**(3), August 2001, pp. 391–409.

Springer-Verlag for articles: 'Language Choice Among Immigrants in a Multi-Lingual Destination', *Journal of Population Economics*, **7**(2), 1994, pp. 119–131; 'Immigrant Earnings: Language skills, linguistic concentrations and the business cycle', *Journal of Population Economics*, **15**(1), January 2002, pp. 31–57; 'Language Skills and Earnings Among Legalized Aliens', *Journal of Population Economics*, **12**(1), February 1999, pp. 63–89; 'Ethnic Networks and Language Proficiency Among Immigrants', *Journal of Population Economics*, **9**(1), March 1996, pp. 19–35.

Springer Science+Business Media, Inc. for article: 'Parents and Children Talk: English language proficiency within immigrant families' (Barry R. Chiswick, Yew Liang Lee and Paul W. Miller), *Review of Economics of the Household*, **3**(3), 2005, pp. 243–268.

The University of Chicago Press for articles: 'The Endogeneity Between Language and Earnings: International analyses', *Journal of Labor Economics*, **13**(2), April 1995, pp. 246–288; 'Speaking, Reading, and Earnings Among Low-Skilled Immigrants' (Barry R. Chiswick), *Journal of Labor Economics*, **9**(2), April 1991, pp. 149–170; 'Indigenous Language Skills and the Labor Market in a Developing Economy: Bolivia' (Barry R. Chiswick, Harry A. Patrinos and Michael E. Hurst), *Economic Development and Cultural Change*, **48**(2), January 2000, pp. 349–367. © The University of Chicago Press.

Preface

The papers included in this volume range in publication date from 1988 to 2005. We are indebted to many individuals who served as research assistants and who provided us with helpful comments (including journal editors and anonymous journal referees), to numerous funding sources, to statistical agencies for generously making data available to the research community, and to our institutional affiliations (both regular and visiting) over these two decades. The specific acknowledgements are reported in each of the separate essays. We would, however, like to thank here IZA—Institute for the Study of Labor (Bonn), the Smith-Richardson Foundation, and the Institute of Government and Public Affairs, University of Illinois for the funding that made this volume possible, Derby Voon who assisted us in the preparation of this volume for publication, and our co-authors on some of these papers: Michael Beenstock, Michael Hurst, Yew Liang Lee, Harry A. Patrinos, Gaston L. Repetto, and Michael Wenz.

Barry R. Chiswick
Chicago, Illinois

Paul W. Miller
Perth, Western Australia
August 2006

Introduction

Genesis, Chapter 11, begins:[1]

> Everyone on earth had the same language and the same words. And as they migrated from the east, they came upon a valley in the land of Shinar and settled there. They said to one another, 'Come, let us make bricks and burn them hard.'—Brick served them as stone, and bitumen served them as mortar.—And they said, 'Come let us build a city, and a tower with its top in the sky, to make a name for ourselves; else we shall be scattered all over the world.' The LORD came down to look at the city and tower that man had built, and the LORD said, 'If, as one people with one language for all, this is how they have begun to act, then nothing that they may propose to do will be out of their reach. Let us, then, go down and confound their speech there, so that they shall not understand one another's speech.' Thus the LORD scattered them from there over the face of the whole earth; and they stopped building the city. That is why it was called Babel, because there the LORD confounded the speech of the whole earth; and from there the LORD scattered them over the face of the whole earth. (Genesis, 11, 1–9)

The economics of language

The story of the Tower of Babel in Genesis provides an explanation of the diversity of languages. It also underlines the power of language and the consequences of the multiplicity of languages that are not mutually intelligible. The linguistic heterogeneity reduced communication and work efficiency, thwarting attempts to build a tower to reach into the Heavens.

While the Tower of Babel story tells of the determinants and consequences of linguistic heterogeneity, the approach taken here is the reverse. In the Biblical story a common language becomes many, and the people are scattered across the face of the globe. In most of the essays that follow the reverse is considered. 'Scattered' peoples are brought together as immigrants in a destination or host country, and the process by which they acquire the common language of the destination is studied. In addition, the labor market consequences of acquiring the primary language of the host country is also studied. These two issues are inter-related since the incentive to acquire the host language, and hence the speed and completeness of doing so, is greater if the benefits from doing so are greater.

In the 20 essays presented in this volume the theoretical and quantitative (empirical) tools that are the standard stock and trade of economics are applied to increase our understanding of the determinants and consequences of language proficiency. The two fundamental questions addressed in these chapters are: What are the *determinants* of language proficiency, especially among immigrant and other linguistic minorities? What are the *consequences* of language proficiency for labor market outcomes, such as earnings and employment?

The approach taken is derived from economics. It is assumed that people have scarce resources (including time) and that they want to maximize their economic wellbeing, subject to the constraints imposed by these scarce resources. It is also assumed that language skills are a form of 'human capital'. Language skills satisfy the three requirements for something to be considered human capital. First, it must be productive. The essays that follow will demonstrate the productivity of language skills in the labor market through enhancing earnings and employment. Presumably, better destination language skills make it easier to find a job that is a good match to one's skills, and improved proficiency in the destination language makes a person more productive on the job, both in doing the job *per se* and in interacting with others. Although we have not found appropriate data to test the hypothesis, it is reasonable to assume that language skills are also productive in consumption and social activities. Just ask anyone who has been a tourist in a country in which they cannot speak the local language.

Second, language skills are acquired at a cost. The cost may be the time and effort the immigrant puts into learning the language, whether in school, by watching soap operas on television, by struggling to communicate with locals or merely learning-by-listening. The cost also includes formal institutional costs for classroom learning (teacher salaries, classrooms, etc.), as well as books and newspapers used to improve skills. Indeed, even the acquisition of language skills among infants is not without cost since it is crucially dependent on parents or other caregivers devoting their time and energy to talking with and otherwise interacting with the infants.

Finally, and most crucially for a resource to be 'human capital', it must be embodied in the person. A person can be separated from the land, truck or IBM stock that he or she owns, but this cannot be done for skills, such as language skills. Moreover, with the abolition of slavery, people cannot be bought and sold, even though nearly all of us 'rent' our labor resources to our employers.

In addressing the issue of the determinants of host country or destination language skills we have found that a useful approach is to consider three fundamental variables. Our three 'E's of language proficiency are *Exposure* to the destination language, *Efficiency* in the acquisition of the destination language, and *Economic Incentives* to acquire the language. Many empirical variables are used to represent the three E fundamental concepts.

Greater detail on these is provided in the separate essays, and so an example of each will suffice here. An important exposure variable is how long the person has lived in the destination. Efficiency refers to the ability to convert exposure into language proficiency. An efficiency variable is age at the onset of learning the language: a four-year-old can learn a ‘foreign’ language much faster than a forty-four-year-old with the same exposure. Finally, economic variables are the costs and benefits from learning the language. For example, someone would be much more willing to put the time, energy and other resources into learning another language if the person expects to be in the destination for two decades rather than for two weeks.

Structure of the book

The essays in this volume were selected from a larger set of our research papers on the economics of language.[2] The selection criteria included our desire to present a broad range of topics and countries studied.

Theory can tell us what might be, but empirical analysis is needed to tell us what is, to test hypotheses, and to measure the magnitude of effects. Each of the 20 essays in this volume involves both theory and measurement (empirical analysis). The data analyses support the hypotheses developed in the theoretical discussions. When this is not the case it is so indicated and the analysis is pushed further to try to provide a better understanding of what is happening and why. Yet, a reader not well versed in the statistical techniques can still profit from the flavor of the analysis by focusing on the discussion in prose.

The subtitle of this volume is ‘International Analyses’. The analyses reported here are for five countries on four continents: the United States, Canada, Australia, Israel and Bolivia. There are several reasons for this range of countries. One is that no data set is ideal. Some have a richer set of questions on some issues related to language, while others have a richer set for other issues. Thus, the studies included here use a wide variety of data sets, crossing international boundaries. Moreover, if all of the studies were limited to one country, we would not know if the findings were specific to that country or whether they were robust findings that transcend national boundaries and institutions. Are the findings sensitive to institutions that are specific to a country or are they generalizable? One such institution is destination country language. For example, are findings for the U.S., Australia and English Canada obtained because of the unique role of English as an international language? The analyses for French Canada, Israel and Bolivia suggest that the patterns are largely universal rather than country specific.

Although many of the essays were originally published in economics journals, others were published in journals in other disciplines, including demography, immigration, linguistics and sociology. Most of the analyses are for speaking skills because this seems to be the favorite language question in censuses and surveys. Yet, literacy (reading and writing) and numeracy (literacy in mathematics) are studied where possible. Most of the analyses are for an economy in which a single language dominates the labor market, although Canada is an exception and the role of linguistic enclaves is studied in several essays. Most of the analyses focus on the individual, but in one the focus is on the linguistic consequences of the interactions between and among parents and children. Most of the studies are based on samples of the population as a whole or of immigrants as a whole, although some focus on specific segments of the population (e.g., illegal aliens, Soviet immigrants, indigenous people). Although many of the studies focus on adult men, several are concerned with women and children as well. Some of the studies are exclusively on immigrants, but many others include the native born.

Thumbnail sketch of findings

Part I includes five papers under the theme 'The Determinants of Language Proficiency'. A theoretical model is developed to explain destination language proficiency among immigrants based on human capital theory and the immigrant adjustment process. The model is found to be very robust for explaining destination language proficiency among immigrants across destinations. Using data for the U.S. and Canada and the model based on the three fundamental variables, *Exposure*, *Efficiency* and *Economic Incentives*, the analyses show that among immigrants destination language proficiency increases with: duration in the destination, educational attainment, living in an area where few speak the origin language, coming from a country that is more distant geographically and linguistically, and a younger age at immigration among both immigrants in general, and for the U.S. the low-skilled illegal aliens who were the beneficiaries of the 1986 amnesty. Proficiency is greater among immigrants from a former British or American colony or dependency, and is lower among refugees. Immigrants to Canada tend to settle in Quebec and adopt French if they come from a Romance language country, otherwise they tend to settle elsewhere in Canada and speak English.

An analysis of linguistic patterns within the family reveals that, other things being equal, greater proficiency by one parent or one child for unmeasured reasons enhances the proficiency of all other family members. The effect of the mother's proficiency on children's proficiency is greater than that of the father's proficiency.

Part II includes five essays on 'The Effects of Language Proficiency on Labor Market Outcomes'. The analyses are for males and females, for speaking, literacy and numeracy skills and for the United States, Australia, and Israel. The labor market outcomes considered are earnings, employment, unemployment and labor force participation. The findings reported above on the determinants of language proficiency are reconfirmed in the papers that include this analysis.

In addition to the usual findings regarding the determinants of earnings, there is a highly significant and large effect of language proficiency (fluency, literacy and numeracy) on both earnings and employment for both men and women. Arriving in a destination during a period of high unemployment is associated with lower earnings, but this 'scarring' effect diminishes with duration in the destination. Earnings are found to be lower among immigrants living in areas with many others from their country or language of origin. Earnings are also lower for those with interrupted stays in the destination.

In the three essays in Part III there is a greater focus on 'The Interaction of Language and Earnings Among Immigrants'. Alternative statistical techniques are employed to disentangle the joint effects of earnings and language on each other. The first essay focuses on Australia, and then compares these results to findings for the U.S., Canada and Israel. Other variables being equal, earnings are over 15 percent greater when those proficient in the destination language are compared to those who are not proficient. The estimates imply a high rate of return on an investment in destination language proficiency, especially for more highly educated immigrants who receive a larger than average increase in earnings.

Two studies in this section focus on specific immigrant groups in the U.S. at opposite ends of the educational spectrum. One is low-skilled (low education) illegal aliens apprehended in the interior of the U.S. The other is the highly-educated Soviet Jewish refugees. Both groups experience improvements in English language skills with duration in the U.S., although the effect is larger for the Soviet Jews, and among both groups education is associated with greater proficiency. The low level of schooling and low English language proficiency retards the earnings progress of the illegal aliens. The data suggest that for the illegal aliens English reading proficiency has a larger impact on earnings than English speaking skills, and as was found in studies of the general population, other variables being equal, Hispanic immigrants are less proficient in English than other immigrants. Soviet Jews experience very rapid improvements in English skills and in earnings from an initially very low base that is consistent with their being highly-educated refugees.

'Language and Earnings Among the Native Born' is the primary theme in the three studies in Part IV, using data on the U.S., Canada and Bolivia. Proficiency in the dominant language in the labor market, English in the U.S., English/French in Canada, Spanish in Bolivia, enhances earnings even among the native born. Bilingualism offers mixed results. Proficiency in an immigrant language as well as English is associated with lower earnings among the native born in the U.S. In Canada, proficiency in an immigrant language is also associated with lower earnings, but English–French bilingualism is associated with higher earnings. The research on Bolivia focuses on indigenous (Indian) language speakers who live in the city and the significant return to them from Spanish language proficiency.

Part V is 'Language, Networks and Enclaves' and includes two studies of immigrant adaptation in the U.S. and Australia. Enclaves and ethnic networks are both a blessing and a curse for immigrants. They facilitate the initial adjustment, both social and in the labor market. They retard, however, linguistic (speaking, reading and writing) and labor market upward mobility. Using the concept of 'ethnic goods' (goods, services and networks that are particularly valued by an ethnic/immigrant group), the analysis shows why ethnic enclaves or ethnic concentrations are formed, why they can persist for a long time, and their adverse effect on improvements in language proficiency and earnings. Indeed, even controlling for the person's own characteristics, including their own language proficiency, living in an ethnic enclave is associated with lower earnings.

The two essays in the final section, Part VI, 'Linguistic Distance', return to a theme developed and discussed in earlier chapters in this volume, namely an efficiency variable in the model for the determination of destination language proficiency—the 'distance' between the immigrant's origin language and the language of the destination. Korean and Japanese are clearly more 'distant' from English than are French and German.

The analysis for Israel shows that there are separate effects of country of origin and language of origin on immigrant's Hebrew language proficiency. The English-mother tongue speakers have lower proficiency in Hebrew, other variables being equal, perhaps because it is an international language for business and science and it is an important second language in Israel. The Arabic-mother tongue speakers have greater proficiency in Hebrew, perhaps because Arabic is linguistically close to Hebrew. Immigrants from dual-language countries of origin are more proficient in Hebrew than those from single language countries.

The last chapter in the volume develops and tests a quantitative measure of the linguistic distance between English and nearly all of the languages coded in the U.S. dicennial census, except for American Indian languages. The tests performed in this chapter indicate that linguistic distance is very important for explaining immigrant English language proficiency in the U.S. This measure has also been successfully applied in Australia and Canada. The methodology could be applied to the development of measures of linguistic distance for other languages.

In summary, taken as a whole, these studies for the U.S, Australia, Canada, Israel and Bolivia indicate that human capital, demographic, linguistic and other factors representing exposure, efficiency and economic incentive concepts can go far to explain the destination language proficiency of immigrants. Dominant language proficiency, whether measured by fluency or literacy skills (as well as numeracy skills) is an important determinant of labor market outcomes for immigrants and the native born. Among whom one lives is important—both the family and the local environment (enclave or linguistic concentration) matter for language skills and labor market outcomes. Considering the favorable effects on employment and earnings, there appears to be a high rate of return to investments in dominant language skills, especially for younger and more highly educated immigrants.

The analyses also demonstrate the importance of testing for robustness through analyses of a variety of data sets, using different questions on language, for different countries, and for both the population as a whole and for selected demographic groups.

Note

1 *Tanakh: The Holy Scriptures*, Philadelphia: Jewish Publication Society, 1985, pp. 16–17.
2 All chapters were co-authored by Chiswick and Miller (and others as indicated in the Contents), except for Chapters 10, 12, 13, 16 and 19, which were written by Barry Chiswick as a sole author or with the collaborators listed where relevant in the Contents.

Part I

The determinants of language proficiency

1 A model of destination-language acquisition

Application to male immigrants in Canada[1]

Language skills play an important role in determining immigrants' social and economic status.[2] Knowing the reasons for immigrants' language skills is important for understanding the determinants of their economic well-being, as well as other aspects of economic, political, and social life in the destination. The identification of the groups "at risk" of lacking proficiency in an official language can provide a basis for designing more effective public policies regarding immigration, language training, the labor market, and immigrants' social and political integration.

In this chapter, in which we examine immigrants' destination-language proficiency, we develop a model of investment in destination-language skills. Among other implications, the model permits insights into the channels through which country of birth influences this proficiency. We test the model using the 1991 Census of Canada.[3] Canada, a major immigrant-receiving country, draws its immigrants from a wide array of countries of origin and linguistic origins.

We demonstrate that a systematic economic model can be developed and applied successfully to analyzing the language practice of immigrants in Canada.[4] Variables reflecting exposure to English or French, efficiency in language acquisition, economic incentives for acquiring destination-language skills, and wealth for financing investments in language skills are all relevant. In addition, we show that models with behavioral interpretations based on birthplace-related characteristics are statistically significant, important in magnitude, and intuitively more appealing, and can be nearly as successful in statistically explaining language proficiency as can a set of birthplace dichotomous variables.

A model of language attainment

The model of language attainment is based on the assumption that language skills are a form of investment in human capital. Immigrants who are not already proficient in the dominant destination language(s) make optimal investments in acquiring the dominant language. Investments in language skills may be made before or after immigration, and those skills affect the choice of destination (Chiswick and Miller 1994a).

Language as human capital

Language skills satisfy the three requirements for human capital. First, they are productive in that they may increase earnings in the labor market or decrease costs of consumption (prices) by lowering the costs of communication with others.[5] One aspect of the productivity of language skills is that they increase the productivity of other forms of human capital. For example, the labor market productivity of a worker with professional skills and dominant language skills is greater than that of an otherwise comparable professional who lacks proficiency in the dominant language. That is, a complementarity exists between language skills and other forms of human capital. (For English-speaking countries, see Chiswick and Miller 1995, 2002, 2003; for Israel, see Berman et al. 2000.) Second, language skills also are embodied in the person and, like other forms of human capital, cannot be separated from the person. Finally, language skills are created at a "sacrifice": this sacrifice takes the form of the time and out-of-pocket resources devoted to language acquisition by those making the investment and by others who may be financing the investment in these skills. Thus language skills are a form of human capital.

One can imagine individual immigrants acting as if they are making optimal investments in destination-language proficiency. They invest up to the point where the marginal rate of return from their investment is just equal to the marginal interest cost of the resources they invest. A schematic representation is presented in Figure 1.1 (Becker and Chiswick 1966), where the marginal rate of return from the investment is the demand curve for investment in language skills (D). This demand curve is higher, the lower the cost of obtaining language skills (whether forgone earnings or out-of-pocket costs), and the greater the monetary and nonmonetary benefits from investments in language skills. The cost of acquiring destination-language skills is lower, the greater the immigrant's exposure to the destination-language either before or after immigration. The costs also are lower if the immigrant is more efficient in converting exposure into language acquisition. The benefits from destination-language acquisition are greater, the larger the returns in the labor market through higher wages or greater employment (lower unemployment) and the lower the costs of consumption among those with greater proficiency. The benefits are also greater, the larger the nonmonetary returns in the form of consumption benefits or from greater participation in the cultural, social, or political life of the destination.

The demand curve for funds for investment in language skills slopes downward because marginal rates of return on the investment decline with a higher level of investment. This decline occurs in part because more profitable investments are made sooner rather than later if investments are

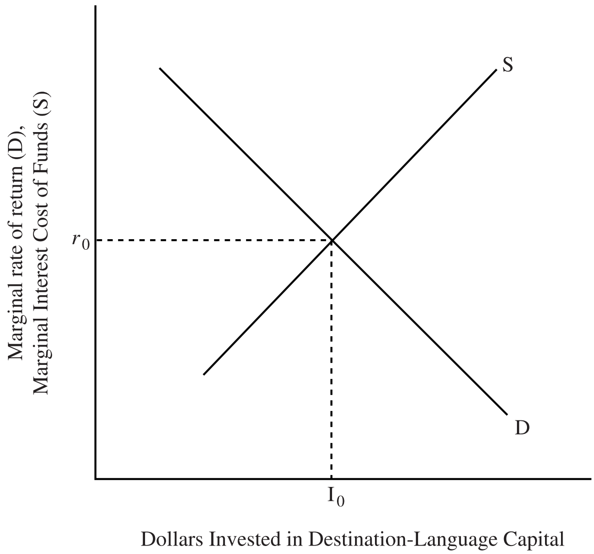

Figure 1.1 Schematic representation of supply and demand for funds for investment by immigrants in destination-language capital.

Note: I_0 is the optimal investment in destination-language capital; r_0 is the marginal rate of return at optimal investment.

independent of each other. It also declines because forgone earnings (i.e., the opportunity cost of time devoted to the investment) increase with greater prior investments. Finally, the remaining working life becomes shorter, the more time devoted to language training.

The marginal interest cost of funds for the investment is the supply curve for investment in language skills (S). This supply curve is lower or farther to the right, the greater the wealth and the greater the access to other resources for financing the investment (Figure 1.1). Those with greater wealth can obtain investment funds at a lower borrowing or self-financing cost than those who are poorer. The supply curve rises upward or slopes positively because investors use cheaper sources of funds before using more expensive sources.

A rising marginal interest cost of funds (supply schedule) and a falling marginal rate of return (demand schedule) from additional investments imply that if investments are made, a stable equilibrium level of funds (I_0) will be devoted to investment in language skills (Figure 1.1). This translates into an optimal level of destination-language proficiency for the individual. It is not possible to measure dollars invested in destination-language proficiency, the schedules for the marginal interest cost of funds, nor the marginal rate of return from the investment. It is possible, however, to measure destination-language proficiency (LANG).

Determinants of language proficiency

The above discussion suggests that a reduced-form equation could relate immigrants' destination-language proficiency to factors that shift the supply and demand curves for investment, namely exposure, efficiency, and economic factors affecting the demand for investment in language skills, and wealth or access to funds for investment in language skills affecting the supply of investment funds. Thus the reduced-form language proficiency equation can be written as

$$\text{LANG} = f\,(\text{Exposure, Efficiency, Economic Incentive, Wealth}).$$

Exposure Exposure to the destination language can occur before or after immigration. The model is trivial if immigrants come from an origin in which the primary language is the same as the primary language used in the destination. Yet even for immigrants from countries in which the destination language is not the dominant language, there may be various levels of exposure in the origin before immigration. Thus the characteristics of the country of origin may matter.

An important country characteristic is the extent to which the destination language is used in the origin country. Such use is more likely if the country of origin is or was a dependency of the destination. Thus pre-immigration exposure to French is more likely among immigrants from the Ivory Coast; similiarly, exposure to English is more likely among immigrants from Ghana, in contrast to (say) immigrants from Angola. Thus the colonial past of the origin is a relevant characteristic.

Post-immigration exposure to the destination language can be decomposed into two elements: time units of exposure and intensity of exposure per unit of time. Time units of exposure merely measure the number of years since immigrating to the destination. Because investments in language skills tend to be more profitable if made just after arrival, and because improvements can be expected with practice, with the biggest improvements in the earliest years, the effect of duration or years since migration will not be linear. A quadratic specification would allow for the improvement in language skills for each additional year in the destination to diminish with additional years since migration.

The intensity of exposure per unit of time in the destination can be related to "neighborhood" characteristics and to "family" characteristics. An immigrant who is not proficient in the dominant language can be more successful in avoiding exposure to and practice of the dominant language if this individual lives in an area where many others use his or her origin language (Chiswick and Miller 1996). The issue is not the use of the origin language per se, but rather the ease of avoiding the dominant language. Depending on the size of the linguistic community, the origin language may be employed in community interactions and activities, newspapers, radio, and even cable TV.

A larger linguistic community can support a linguistic-enclave labor market; working in such an enclave would reduce exposure to the destination language. These effects can be measured by the proportion of the population of the area, regardless of nativity, that speaks the immigrant's origin language. Ethnic networks have been shown to be quite important in recent Canadian research (see, for example, Guindon and Poulin 1998).

The family is an intense environment in which the immigrant operates. Language practice within the family will influence proficiency in the destination language. Marriage before immigration is more likely to be to a spouse from the same country of birth and linguistic background, and the origin language is more likely to be used at home. The result would be less proficiency in the destination language. If the immigrant married a person not proficient in the immigrant's origin language, there would be greater use of, and proficiency in, the destination language. Thus the effect of marriage per se before immigration is likely to be a lower level of proficiency, but the effect of post-immigration marriage is ambiguous.

Children could exert several partially offsetting effects on parents' language skills. First, their greater ability to learn new languages and their intensive exposure to the destination-language in school tend to hasten their linguistic adjustment (Long 1990; Newport 1990; Service and Clark 1993). By bringing home their improved destination-language skills, children serve as role models and "teachers" for their parents.

Second, because children, whether native- or foreign-born, are more likely than their parents to acquire proficiency in the destination language, they can serve as translators.[6] This role for children would detract from the parents' potential destination-language proficiency. Because this role as translator is more important for consumption and home production activities than for labor market activities, it would be expected to exert a larger negative effect on the mother's destination-language proficiency than on the father's.

Third, children also affect labor supply, particularly that of their mothers. Insofar as investments in destination-language skills are made for their labor market benefits and insofar as exposure to the dominant language in the labor market enhances proficiency, children would detract from parents' proficiency. Thus children would be expected to exert a less positive or more negative effect on their mother's proficiency than on their father's.

Finally, parents concerned with transmitting the cultural identity and practices of the country of origin to their children may retain the language of origin in the home. Insofar as the parents use the origin language at home for their children's benefit, their own proficiency in the destination language is diminished.

This analysis of the effects of children on parents' language skills suggests that the sign of the effect is ambiguous. It also suggests that the effects differ for the mother and the father: the effect of children on parents' language skills would be less positive or more negative for the mothers than for the fathers.

Efficiency Immigrants may differ in their efficiency in acquiring the destination language. Efficiency refers to the extent of improvement in destination-language skills per unit of exposure. One of the most important variables influencing efficiency is age at migration: younger people have a far greater capacity for learning a new language than do older individuals (Long 1990; Newport 1990; Service and Clark 1993). At least in part this seems to be a biological process: as a person ages, the brain loses some of its flexibility in adapting to new languages. Thus we expect that destination-language proficiency will decline with an older age at immigration, all else being equal, but without any sharp breaks at particular ages.

School attainment also may affect efficiency. Those with higher levels of schooling may have a greater ability to learn, and this learning ability for school subjects may carry over to languages. Those with more schooling also may have learned more about their origin language, including its structure and grammar, and may be more proficient in their origin language. A fuller or deeper understanding of one's own language may facilitate the learning of other languages. Moreover, if the destination language is an international tongue (such as English or French), those with higher levels of schooling in the origin may have received more exposure to the destination language in school before immigration. In some countries, for example, the study of English or French is mandatory at higher levels of schooling.[7] Finally, among immigrants who have not completed their schooling, post-migration schooling may be associated with higher levels of destination-language proficiency because proficiency is required for entry into school and because schooling in the destination language enhances proficiency.

The difficulty in learning a destination language depends in part on the person's origin language. A Chinese speaker would find it more difficult than a Spanish speaker to learn French because the differences between the languages are so much greater. That is, the "linguistic distance" between Chinese and French is greater than between Spanish and French. The greater the linguistic difference between the destination and the origin language, the lesser an immigrant's efficiency in learning the destination language.

Linguists' approach to language differences is based on their concern about the roots or evolution of languages. For example, as the *Cambridge Encyclopedia of Languages* explains, "The main metaphor that is used to explain the historical relationships is that of the language *family* or *family tree*" (Crystal 1987:292, emphasis in the original; also see Grimes and Grimes 1993). The *Encyclopedia* includes an extensive discussion of families of languages (Crystal 1987:283–340). Yet it includes only two brief paragraphs in a sidebar on "interlingual distance":

> The structural closeness of languages to each other has often been thought to be an important factor in FLL (foreign language learning). . . . However, it is not possible to correlate linguistic difference and learning difficulty in any straightforward way, and even the basic task of

> quantifying linguistic difference proves to be highly complex, because of the many variables involved.
>
> (Crystal 1987:371)

As shown in Chiswick and Miller (1998a) and below, this interpretation is too pessimistic.

Immigrants differ in their reasons for migrating. Those who move primarily for improved labor market opportunities for themselves are known as economic migrants. Those who move primarily to accompany or join another family member, as is often the case for women and more so for children, are known as tied movers. Those who move because of a fear or perception of persecution or discrimination due to their race, ethnicity, political orientation, or social class are called refugees. Ideological migrants are those who move for nationalistic or political reasons to a destination that better satisfies these objectives, even if they do not fear or perceive short-term discrimination in the origin.

Refugees, tied movers, and ideological migrants are less likely than economic migrants to be favorably self-selected for labor market success in the destination because, by definition, such success is less important in their migration decision. Because part of the adjustment is linguistic, this implies that those who are not economic migrants are likely to be less able in learning a foreign language: that is, they would be less efficient in learning the destination language. Moreover, refugee status is often unanticipated, and refugees may have had less time than economic migrants to plan or prepare for the move. Thus we would expect refugees, tied movers, and ideological migrants to be less proficient in the destination language than otherwise similar economic migrants.

Economic incentives The economic incentives for destination-language proficiency depend on the increment in wages due to becoming proficient and on the expected duration of employment or duration of stay in the destination. One might wish to set up a simultaneous system of equations in which language skills are a function of expected increments in wages, and wages are a function of language skills. This arrangement is not feasible for econometric reasons, primarily the difficulty in developing identifying variables. To some extent, level of schooling may reflect a wage increment incentive effect because the economic returns to destination-language proficiency increase with the level of schooling. (For analyses for several countries, see Chiswick and Miller 1995, 2002, 2003.)

Another relevant economic incentive variable would be the likelihood of return migration: that is, the expected future duration in the destination labor market. The geographic distance of the country of origin from the destination is relevant here. The greater cost of migrating due to greater distance has two interrelated effects. First, all else being equal, immigrants will tend to be more favorably self-selected, and hence of greater ability (efficiency), the greater the geographic distance (Chiswick 1999). Second, distance is associated with a lower expectation of return migration and hence with a greater incentive to invest in destination-specific skills, including language skills.

An immigrant obtaining destination-language skills also receives "consumption benefits," defined broadly. These may take the form of lower prices through a more efficient search (or a broader market) for market goods and services, and greater participation in the social, political, and cultural life of the destination country. If education increases the demand for social, political, and cultural incorporation with the destination, then those with higher levels of education would have a greater economic incentive for obtaining destination-language skills.

Wealth Finally, one must consider the supply side of the market for funds for investment in human capital. Variables that lower the interest cost of funds—that is, that shift to the right the supply curve of funds for investment—encourage greater investments in destination-language proficiency. Access to the capital market is important. Greater wealth, which lowers the interest cost of funds, encourages investment in language skills and hence enhances language proficiency (Grenier and Vaillancourt 1983). In the absence of a more direct measure, greater wealth may be associated with a higher level of schooling.

The overall model Thus the model generates a conceptual equation (with hypothesized signs in parentheses):

> LANG = *f* [Age at migration (–), Years since migration (+), Education (+), Married before migration (–), Married after migration (?), Children (?), Linguistic Distance (–), Geographic Distance (+), Minority Language Concentration Index (–), Refugee (–), Colony (+)].

The estimating equation, of course, includes a residual to account for purely random unmeasured individual characteristics that affect language proficiency. These include factors such as innate language ability and personality traits (see, for example, Chastain 1975; Gardiner 1990; Svanes 1987).

The data and empirical implementation

The primary data set used in this research is the 1991 Census of Canada, Public Use Microdata File (Individuals), 3% simple random sample of the population. Because each observation has the same weight (33.3), we compute the analyses unweighted (or self-weighted). These data are supplemented from various sources. The construction of each of the variables used in the empirical analysis is described in detail in the appendix. The Public Use Microdata File (Individuals) released from the 1991 Census contains information on the speaking knowledge of the official languages (English and French), knowledge of other languages, the language usually used at home, and mother tongue.[8]

The language variables can be used to create a trichotomous dependent language variable (see appendix). The first language state is the inability to conduct a conversation in either of the official languages of Canada, English and French (L1). In the second language state, the immigrant is able to conduct a conversation in one or the other of the two official languages of Canada, but usually speaks a nonofficial language at home (L2). This state may be considered a partial shift away from the language of the country of origin. In the third language state, the immigrant speaks English and/or French, and usually speaks one or both of these official languages at home (L3). When an immigrant comes from a country in which English and French are not spoken, and when neither language is the respondent's mother tongue, language state L3 represents a complete shift from the language of the country of origin to an official language of Canada.

The statistical analysis is limited to adult foreign-born males age 25 to 64. We limit our analysis to adult males who are not aged in order to focus on individuals likely to be fully attached participants in the labor market. The inclusion of females and of males over 65 would necessitate expanding the analysis to include labor supply considerations that may be determined endogenously with destination-language skills.[9] Moreover, because of likely gender differences in the determinants of language proficiency, a simple dichotomous variable would be inadequate; we would need to compute separate equations by gender (Chiswick and Miller 1998a; Stevens 1994). This process would take our analysis in directions that are beyond its scope.

Those born in identifiable English-speaking countries are excluded from the analysis because they tend to arrive fully fluent in English. These immigrants include persons from the United States and the United Kingdom, and those born in "Central America, Caribbean, Bermuda, and South America" with an English mother tongue.[10]

Among adult male immigrants in Canada from non-English-speaking countries, about 5% cannot speak either English or French (L1): 4% in Quebec and 5% in "English" Canada. About 95% can speak English or French: of this group, about half in Quebec and about half in English Canada usually speak English or French at home, and about half usually speak neither language at home. Thus slightly fewer than half of these immigrants usually speak one of the official languages at home.

The Census provides basic information on immigrants' characteristics, as described in the appendix. The variables for age, educational attainment, period of residence in Canada, marital status, and province/metropolitan area are standard. Unfortunately, there are no data that permit us to determine whether the current marriage took place before migration or to learn the number of children for adult males.

As indicated earlier, country of birth may play a key role in the empirical application of the human capital model of language skills. Chiswick and Miller (1992), for example, included eight country-of-birth dummy variables in the analysis of official-language fluency in Canada in the 1981 Census; the majority of these were highly significant and were associated with sizable effects on the degree of language fluency. Yet dichotomous variables for country of birth may be regarded as a measure of what we do not know: that is, of the relevant unmeasured behavioral variables that underlie the country-of-birth effects.

The five variables discussed in the previous section that reflect dimensions of country of birth are geographic distance, linguistic distance, the minority language concentration index, refugee status, and former colony of an English- or French-speaking country. The incorporation of these variables into an analysis for Canada is a major, innovative contribution of this chapter.

The geographic distance variable reflects the cost, in money and time, of moving from the origin to the destination, as well as the cost of a return migration. In studies of English-language fluency in the United States, Espenshade and Fu (1997) and Chiswick and Miller (1998a) used a similar variable to capture the effects associated with propensities for initial migration (self-selection) and return migration that vary by distance of the country of origin. In this study we measure this variable as the distance, in thousands of miles, between the major city in the country of origin and the closest major gateway city (Toronto or Vancouver) into Canada. In analyses for Quebec Province, we use Montreal as the gateway city.

A second innovation in this study of Canada is a direct examination of the impact of "linguistic distance" on official-language fluency. The greater the linguistic distance of the origin language from the destination language, the lesser the proficiency in the destination language. Because languages are multifaceted, linguists have not been able to develop a measure of linguistic distance (Crystal 1987: 292, 371). "Family trees" for languages reflect linguists' perceptions or hypotheses regarding the origins of languages; they do not indicate the difficulty, for the speaker of one language, in learning another.

Chiswick and Miller (1998a) developed an index of "linguistic distance" based on the degree of difficulty, among Americans who are native English speakers, in learning foreign languages (see appendix). This index is developed from a set of language learning scores (LS) (see Hart-Gonzales and Lindemann 1993). A low value of the score indicates a high degree of difficulty (e.g., Cantonese LS = 1.25); a high value indicates a low degree of difficulty (e.g., Dutch LS = 2.75). Symmetry is assumed; if it is difficult for speakers of language A to learn language B, it is also difficult for B speakers to learn language A. In this study of language in Canada, we assume that the linguistic distance (LD) index developed for English in the United States can be applied as well to English in Canada. In the empirical application, linguistic distance is measured as the reciprocal of the language score: LD = 1/LS. Thus a higher value for LD denotes a greater distance between English and the origin language.

The approach to the study of language in the United States must be modified somewhat here because Canada is a dual-language country. We are not aware of any comparable linguistic distance index for French, and thus pursue three alternatives. First, we allow the effect of the LD measure to vary between Quebec and English Canada. Second, we use a separate measure for countries speaking a Romance language to capture the linguistic distance from French. This is a Romance-/non-Romance-language dichotomous variable used in the analysis for Quebec. Third, we conduct separate analyses for Quebec and for English Canada.

A third innovation and another constructed variable is the minority language concentration index. This is measured as the percentage of the population (native-born and foreign-born, male and female) age 18 to 64 in the respondent's region of residence that reports the same mother tongue as the respondent.[11] The region is defined as the person's Census Metropolitan Area (CMA) or, for those not living in a CMA, the rest of the province.

Finally, we create two dichotomous variables that describe different characteristics of the country of origin. The precision of these variables is restricted by the limited amount of detail on country of origin provided in the Canadian Public Use Sample. "Refugees" are identified as those reporting Vietnam or the USSR as their country of birth. Those born in former "colonies" of the United States, the United Kingdom, or France are identified through their birth in south Asia, Vietnam, Africa, Hong Kong, or the Philippines.

Table 1.1 lists means and standard deviations for the variables in the study of language. These data are for the 3% sample of foreign-born males from non-English-speaking countries, age 25 to 64, who report valid information on each of the variables used in the analysis. Data are reported for the total sample, and separately for Quebec and English Canada (other than for the Atlantic provinces). For this group, the mean age is 44 for Canada. On average, almost one-half of these years have been spent in Canada: the mean duration of residence is 20 years. The mean educational attainment is 11.6 years. Quebec accounts for 17% of the sample, the prairie provinces for 13%, British Columbia for 15%, and Ontario for 54%. The immigrants' major birthplace regions are "other" Europe (24%), Italy (14%), south Asia (8.3%), and Africa, Germany, and Portugal (6% for each).

We find few differences between the immigrant populations of Quebec and of English Canada, other than for the birthplace distributions. There are fewer immigrants from Germany, Poland, and the countries or regions of Asia (other than Vietnam), and more immigrants from Italy, Africa, and Central and South America in Quebec than in English Canada. Most striking is the greater propensity for immigrants from Romance-language countries and former French colonies to live in Quebec, presumably because of the smaller linguistic and cultural distance (Chiswick and Miller 1994a).

Table 1.1 Means and standard deviations of variables, male immigrants from non-English-speaking countries: 1991 Census of Canada

Variable	*Total Sample*		*English Canada*[a]		*Quebec*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Age	44.02	10.91	44.04	10.93	43.86	10.81
Age at Migration	24.18	11.46	24.00	11.51	25.30	11.06
Education Level	11.61	4.19	11.61	4.11	11.57	4.57
Period of Residence	19.84	12.64	20.05	12.73	18.57	11.96
Atlantic Provinces	0.006	0.08	—[b]		—[b]	
Quebec	0.170	0.38	0.000	0.00	1.000	0.00
Ontario	0.544	0.50	0.660	0.47	—[b]	
Prairie Provinces	0.127	0.33	0.155	0.36	—[b]	
British Columbia	0.153	0.36	0.185	0.39	—[b]	
Lives in CMA	0.872	0.33	0.861	0.35	0.949	0.22
Married	0.815	0.39	0.819	0.38	0.794	0.40
Miles Between Canada and Origin[c]	4,985	1,468	5,039	1,443	4,686	1,726
Linguistic Distance	0.504	0.11	0.507	0.11	0.493	0.10
Minority Language Concentration	2.253	2.43	2.401	2.54	1.611	1.62
Refugee	0.048	0.215	0.050	0.218	0.042	0.200
Colony	0.253	0.434	0.265	0.442	0.199	0.400
Italy	0.135	0.34	0.126	0.33	0.181	0.39
Germany	0.062	0.24	0.070	0.26	0.022	0.15
Portugal	0.057	0.23	0.059	0.24	0.050	0.22
Poland	0.043	0.20	0.048	0.21	0.021	0.14
USSR	0.015	0.12	0.017	0.13	0.006	0.07
Other Europe	0.244	0.43	0.236	0.42	0.255	0.44
Middle East	0.048	0.21	0.039	0.19	0.091	0.29
Southern Asia	0.083	0.28	0.093	0.29	0.037	0.19
Hong Kong	0.044	0.20	0.051	0.22	0.007	0.09
China	0.050	0.22	0.057	0.23	0.022	0.15
Philippines	0.031	0.17	0.036	0.19	0.006	0.07
Vietnam	0.034	0.18	0.033	0.18	0.036	0.19
Other Asia	0.040	0.20	0.041	0.20	0.036	0.19
Africa	0.062	0.24	0.052	0.22	0.113	0.32
C. and S. America	0.053	0.22	0.040	0.20	0.116	0.32
Sample Size	32,168		26,484		5,483	

Notes
a Excludes the Atlantic provinces. See text for explanation.
b Variable not relevant.
c Distance variable for Quebec defined with reference to Montreal.

Source: 1991 Census of Canada, Public Use Microdata File (Individuals).

Empirical analysis

Two statistical techniques can be applied: multinomial logit analysis and ordered logit analysis. Ordered logit analysis is the more restrictive of the two: it treats the three categories as ordered, moving from a lower to a higher level of proficiency in one of the official languages. The three categories in the language variable are not strictly ordered in principle, although this appears to be the situation in practice. Ordered logit, however, assumes that the proportional odds of going from category 1 to category 2 are the same as from going from category 2 to category 3. Yet the hypothesis that the ordered odds are the same is not consistent with the data. For the total sample, the chi-square test statistics for the test of the proportional odds assumption in the ordered logit models corresponding to those presented in Tables 1.2, 1.3, and 1.5 are 357.8, 617.6, and 644.2 respectively. These have an asymptotic chi-square distribution with 10, 24, and 15 degrees of freedom, and thus indicate that the proportional odds assumption is not appropriate for these data and models.

For these reasons, we employ the more flexible but somewhat more complex multinomial logit technique and use the logit coefficients to obtain predicted probabilities for various values of several explanatory variables.

The analysis of the results of a multinomial logit model can focus on the estimated coefficients (which inform about the impact of variables on the log-odds), on the associated marginal effects on predicted probabilities, or on predicted distributions across the language categories. In the present discussion we first consider only the sign and statistical significance of the estimated coefficients. When we discuss the final model and when we examine the birthplace coefficients, to simplify the exposition, we present predicted probabilities of being in each of the language states for a range of characteristics.

The estimated coefficients in the multinomial logit model give the partial effects of the explanatory variables on the log of the odds (often referred to as the "log-odds") of being in the second (L2) or third (L3) language state relative to being unable to speak either English or French (L1). A positive coefficient for log (L2/L1) means that the explanatory variable increases the probability of being in L2 relative to being in L1. The relevant log-odds for L3 relative to L2 are easily calculated from log (L3/L2) = log (L3/L1) – log (L2/L1).

One may compute the partial derivatives of the probability of being in L3, L2, or L1 with respect to the explanatory variables. For dichotomous variables with large coefficients, these can be poor approximations of the true partial effects because the partial derivative concept refers to infinitesimal changes rather than to discrete changes. For this reason we prefer predicted probabilities.

We estimate three models of language practice for Canada as a whole and separately for English Canada and Quebec. The first model contains only variables for region of residence and personal characteristics other than birthplace (Table 1.2). The second augments the first core specification with

Table 1.2 Estimates of logit model of language practice, 25- to 64-year-old male immigrants from non-English-speaking countries: 1991 Census of Canada

	Total Sample		*English Canada*[a]		*Quebec*	
	Log(L2/L1)	*Log(L3/L1)*	*Log(L2/L1)*	*Log(L3/L1)*	*Log(L2/L1)*	*Log(L3/L1)*
Constant	1.107	0.182	0.909	–0.097	3.750	3.801
	(5.46)	(0.85)	(4.26)	(0.43)	(3.32)	(3.35)
Age at Migration	–0.051	–0.091	–0.048	–0.091	–0.068	–0.099
	(16.56)	(27.42)	(14.57)	(25.30)	(8.02)	(11.02)
Educational Attainment	0.236	0.369	0.245	0.370	0.186	0.345
	(33.70)	(48.47)	(32.23)	(44.42)	(10.08)	(17.65)
Period of Residence (PER)	0.119	0.219	0.120	0.228	0.118	0.196
	(13.24)	(23.00)	(12.26)	(21.94)	(5.06)	(8.09)
$PER^2/100$	–0.087	–0.087	–0.090	–0.091	–0.088	–0.124
	(3.47)	(3.38)	(3.30)	(3.24)	(1.38)	(1.89)
Province (Ontario)						
Prairie provinces	0.178	0.448	0.182	0.470	—[b]	—[b]
	(1.83)	(4.41)	(1.86)	(4.60)		
British Columbia	–0.179	–0.165	–0.187	–0.151	—[b]	—[b]
	(2.18)	(1.90)	(2.27)	(1.74)		
Quebec	0.585	0.799	—[b]	—[b]	—[b]	—[b]
	(6.58)	(8.63)				
Atlantic provinces	0.598	1.285	—[b]	—[b]	—[b]	—[b]
	(0.59)	(1.26)				
Lives in CMA	–0.326	–1.154	–0.342	–1.051	–1.035	–2.795
	(2.58)	(8.98)	(2.65)	(8.00)	(1.01)	(2.74)
Married	0.128	–0.093	0.142	–0.050	–0.036	–0.303
	(1.55)	(1.09)	(1.60)	(0.53)	(0.16)	(1.29)
Sample Size	32,168		26,484		5,483	
Chi-Square	11,041		9,815.6		1,356.1	
Pseudo-R^2	0.2016		0.2160		0.1507	

Notes: L1 = speaks neither English nor French; L2 = speaks an official language but usually speaks a nonofficial language at home; L3 = speaks an official language and usually speaks an official language at home. Asymptotic *t*-statistics in parentheses.
a Excludes Atlantic provinces; see text for explanation.
b Variable not relevant.

Source: 1991 Census of Canada, Public Use Microdata File (Individuals).

variables for birthplace (Table 1.3). In the third, the birthplace variables are replaced with variables for minority language concentration, linguistic distance, geographic distance (miles) between the country of origin and Canada, refugee status, and former British, French, or American colony (Table 1.5).

It is apparent from Table 1.2 that age at migration, educational attainment, and duration of residence are significant determinants of the measure of language practice. The coefficients in the equation for the log-odds of L3 to L1 are considerably larger in absolute value than in the equation for the log-odds of L2 to L1. This indicates that the shift to an official language increases with duration in Canada (the largest increases come in the early years) and with educational attainment.[12] Immigration at an older age, however, is associated with a lower probability of being in the third language state (speaks an official language and usually speaks an official language at home) than in the second (speaks an official language but usually speaks a nonofficial language at home), and a lower probability of being in the second language state than of not speaking either of the two official languages.

Region of residence also influences immigrants' language outcomes. Those living in Quebec and the prairie provinces use English or French more extensively than do immigrants in Ontario, all else being equal. Although the reasons for these regional effects are not clear, these variables mean that province-specific influences (fixed effects) are held constant. Residents of a Census Metropolitan Area (CMA) show a lower use of an official language than do immigrants living outside these major cities, possibly because of the limited concentration of foreign-language speakers in rural areas. Marital status, however, is not a significant factor in the model.[13]

These broad patterns carry over in Table 1.2 to the separate analyses conducted for English Canada (second column) and Quebec (third column). The only notable difference in the results for the two distinct language regions of Canada is that the duration-of-residence effects produce a pronounced nonlinear pattern in English Canada but are approximately linear in Quebec.

Table 1.3 augments the basic model of Table 1.2 with 14 dichotomous variables for birthplace. Immigrants from Italy are the reference group. Inclusion of the birthplace dummy variables in the model has little effect on the magnitudes or levels of statistical significance of any of the variables in the original set of explanatory variables. In this analysis, 20 of the 28 birthplace coefficients for Canada (total sample) are statistically significant at conventional levels; as a group they are highly significant. They reveal that some birthplace groups, such as immigrants from China, have higher probabilities of inability to speak an official language, while other birthplace groups, such as Vietnamese, are more likely to continue speaking a nonofficial language at home. From the size of the estimated coefficients (compared, for example, with the coefficient on age at migration), and from the statistical significance of the coefficients, it is apparent that birthplace matters in attempts to account for the distribution of language skills in Canada. This point is

Table 1.3 Estimates of logit model of language practice, 25- to 64-year-old male immigrants from non-English-speaking countries, including country of birth: 1991 Census of Canada

	Total Sample		*English Canada*[a]		*Quebec*	
	Log(L2/L1)	*Log(L3/L1)*	*Log(L2/L1)*	*Log(L3/L1)*	*Log(L2/L1)*	*Log(L3/L1)*
Constant	0.940	–0.039#	0.746	–0.256#	3.707	3.628
Age at Migration	–0.053	–0.102	–0.051	–0.103	–0.071	–0.108
Educational Attainment	0.210	0.344	0.213	0.341	0.197	0.349
Period of Residence (PER)	0.141	0.234	0.143	0.243	0.146	0.219
$PER^2/100$	–0.124	–0.120	–0.130	–0.125	–0.125†	–0.148
Province (Ontario)						
Prairie provinces	0.115#	0.432	0.101#	0.441	—[b]	—[b]
British Columbia	–0.130#	0.052#	–0.172†	0.028#	—[b]	—[b]
Quebec	0.499	0.558	—[b]	—[b]	—[b]	—[b]
Atlantic provinces	0.164#	0.338#	—[b]	—[b]	—[b]	—[b]
Lives in CMA	–0.252†	–0.847	–0.265	–0.754	–1.292#	–2.895
Married	0.260	0.083#	0.277	0.131#	0.140#	–0.121#
Birthplace (Italy)						
Germany[c]	12.915#	14.135#	13.035#	14.187#	11.955#	13.244#
Portugal	–0.526	–0.404	–0.518	–0.472	–0.277#	–0.001#
Poland	–0.006#	–0.098#	0.125#	0.018#	–0.796†	–1.116
USSR	0.776	0.761	0.845	0.698	12.815#	13.738#
Other Europe	0.636	1.530	1.002	1.812	–0.515†	0.499†
Middle East	0.914	1.179	0.885	1.172	1.043	1.048
South Asia	0.862	0.821	0.911	0.826	1.792	1.988
Hong Kong	1.029	–0.157#	1.159	–0.046#	–0.105#	–1.543†
China	–0.952	–2.381	–0.802	–2.322	–1.965	–3.069
Philippines	3.206	3.192	3.278	3.254	14.433#	14.208#
Vietnam	–0.119#	–1.571	–0.073#	–1.539	0.240#	–1.350
Other Asia	0.352	0.092#	0.568	0.377†	–1.448#	–1.494

Africa	2.959	4.333	2.863	4.108	13.471#	15.027#
C. and S. America	0.204#	0.554	0.201#	0.359†	0.098#	0.583#
Sample Size	32,168		26,484		5,483	
Chi-Square	14,599		12,694		2,048.3	
Pseudo-R^2	0.2666		0.2794		0.2276	

Notes: L1 = speaks neither English nor French; L2 = speaks an official language but usually speaks a nonofficial language at home; L3 = speaks an official language and usually speaks an official language at home. The logit coefficients are statistically significant at .05, two-tailed test, unless indicated otherwise. The full regression equations with *t*-statistics are available on request.

a Excludes Atlantic provinces; see text for explanation.

b Variable not relevant.

c There are no respondents from Germany in language category L1. The large estimated coefficients in the logit model bound the predicted probability to zero.

Not significant at .10; †Significant at .10, but not at .05.

Source: 1991 Census of Canada, Public Use Microdata File (Individuals).

illustrated further by the predicted distributions across the language groups presented in Table 1.4.

The predicted distributions presented in Table 1.4 are for two duration-of-residence groups: those who have lived in Canada for five years (left-hand side) and those who have lived in Canada for 15 years (right-hand side). The data for the more recent arrivals (predicted for five years' residence in Canada) show considerable variation across birthplace groups in the percentage unable to speak an official language. This figure ranges from around 30% for immigrants from China to 19% for Portuguese immigrants to negligible proportions among immigrants from Germany, the Philippines, and Africa. Among those immigrants who can conduct a conversation in an official language, we also find considerable variation in the propensity to speak a nonofficial language at home. This practice is more frequent among immigrants from Italy, Poland, and the Asian countries. The figures presented in the right-hand columns of Table 1.4 (predicted for 15 years' residence in Canada) illustrate the rapid growth in fluency with duration of residence, as well as the shift, in most cases, away from the use of the language of the country of origin. Immigrants from Vietnam and China are notable exceptions to this pattern, but very few of the Vietnamese have lived long in Canada.

Table 1.4 Predicted distributions across language categories by birthplace and duration of residence: 1991 Census of Canada[a] (percentages)

	After 5 Years in Canada			*After 15 Years in Canada*		
Birthplace	*L1*	*L2*	*L3*	*L1*	*L2*	*L3*
Italy	12.38	73.83	13.79	3.43	65.32	31.25
Germany	0.00	61.24	38.76	0.00	38.16	61.84
Portugal	18.97	66.91	14.12	5.45	61.38	33.17
Poland	12.59	74.68	12.73	3.55	67.16	29.30
USSR	6.12	79.28	14.60	1.62	66.86	31.52
Other Europe	5.74	64.71	29.55	1.26	45.51	53.22
Middle East	5.13	76.29	18.58	1.28	60.80	37.92
South Asia	5.66	79.99	14.35	1.50	67.51	31.00
Hong Kong	5.37	89.53	5.11	1.61	85.84	12.54
China	29.36	67.61	3.03	10.87	79.96	9.16
Philippines	0.57	83.95	15.48	0.14	67.83	32.02
Vietnam	15.32	81.13	3.55	5.05	85.38	9.56
Other Asia	9.34	79.25	11.41	2.63	71.14	26.23
Africa	0.50	57.25	42.25	0.09	34.58	65.33
C. and S. America	9.75	71.33	18.92	2.49	58.07	39.44

Notes: L1 = speaks neither English nor French; L2 = speaks an official language but usually speaks a nonofficial language at home; L3 = speaks an official language and usually speaks an official language at home.

a Predictions computed at means of all variables other than duration of residence. Row totals for each immigration period may not add to 100.0 because of rounding.

Source: Estimates in Table 1.3.

The information presented in Table 1.4 shows clearly that language practice and skill vary appreciably across the birthplace groups. The fundamental reasons for this variation are the subject for the remainder of this section. Hence Table 1.5 presents a model in which the birthplace variables have been omitted, and in which the estimating equation includes measures for minority language concentration, linguistic distance, geographical distance of the country of origin from Canada, refugee origin, and former colony of Britain, France, or the United States. Because of a one-to-one correspondence between the birthplace dichotomous variables and the measure of geographic distance, both cannot be included in the model at the same time. Moreover, the refugee and colony variables and (for some respondents) the minority language and linguistic distance measures were computed with information on country of birth (see appendix).[14]

The five new variables carry the expected signs, and all are highly significant. The performances of the models presented in Tables 1.3 and 1.5 can be compared only informally because the models are formally nonnested; yet we constructed the nonoverlapping variables using, in part, the same (birthplace) information. Compare the chi-square test statistic and the pseudo-R^2 in Tables 1.2, 1.3, and 1.5. The addition of the five variables with behavioral interpretations (Table 1.5) substantially increases the fit over the equations without these variables (Table 1.2). The improvement in fit is even greater if these five variables are replaced by the set of birthplace dichotomous variables (Table 1.3). In the analysis for Canada as a whole, the pseudo-R^2 is 0.202 when no birthplace variables are present (Table 1.2); it increases to 0.245 when the five substantive variables are added to the equation (Table 1.5), but only to 0.267 when they are replaced by the birthplace dichotomous variables (Table 1.3). The five behavioral variables account for about two-thirds of the explanatory power attributable to the birthplace dichotomous variables. Thus the models containing the birthplace variables provide a better "fit"; this is most important for prediction, but is clearly inferior for testing hypotheses and for understanding the underlying behavioral phenomena.

The greater the geographic distance between country of origin and Canada (and hence the greater the favorable selectivity and the less likely a return migration, all else being equal), the more likely that an immigrant will be able to conduct a conversation in an official language, and that he will use an official language at home.

An increase in the proportion of the area's population (CMA or balance of province) that can speak the immigrant's origin language is associated with an increase in the probability that the immigrant will not be able to speak an official language. It is also associated with an increase in the likelihood that an immigrant who can speak an official language will speak a nonofficial language at home. Thus immigrants show less ability in the official languages and less use of those languages, the larger the proportion of the population in the area where they live who speak their origin language.

Table 1.5 Estimates of logit model of language practice, 25- to 64-year-old male immigrants from non-English-speaking countries, extended model: 1991 Census of Canada

	Total Sample		*English Canada*[a]		*Quebec*	
	Log(L2/L1)	*Log(L3/L1)*	*Log(L2/L1)*	*Log(L3/L1)*	*Log(L2/L1)*	*Log(L3/L1)*
Constant	1.354	2.546	0.994	1.968	5.265	8.156
Age at Migration	–0.050	–0.093	–0.047	–0.094	–0.064	–0.096
Educational Attainment	0.234	0.388	0.240	0.389	0.201	0.365
Period of Residence (PER)	0.123	0.211	0.124	0.221	0.110	0.184
$PER^2/100$	–0.072	–0.064	–0.070	–0.066	–0.070#	–0.087#
Province (Ontario)						
Prairie provinces	0.128#	0.488	0.105#	0.496	—[b]	—[b]
British Columbia	0.062#	0.337	0.049	0.344	—[b]	—[b]
Quebec	0.530	0.675	—[b]	—[b]	—[b]	—[b]
Atlantic provinces	0.493#	0.877#	—[b]	—[b]	—[b]	—[b]
Lives in CMA	–0.115#	–0.690	–0.112	–0.581	–1.031#	–2.418
Married	0.186	–0.046#	0.216	0.008#	–0.001#	–0.260#
Minority Language Concentration	–0.110	–0.206	–0.124	–0.205	–0.055#	–0.271
Linguistic Distance	–1.690	–7.009	–1.102	–6.460	–4.655	–10.492
Miles Origin Country From Canada/1,000[c]	0.091	0.219	0.066#	0.231	0.126#	0.114#
Refugee	–0.944	–2.256	–0.938	–2.133	–1.143	–3.380
Colony	0.948	0.754	0.960	0.592	1.552	2.402
Sample Size	32,168		26,484		5,483	
Chi-Square	13,432		11,801		1,861	
Pseudo-R^2	0.2453		0.2598		0.2068	

Notes: L1 = speaks neither English nor French; L2 = speaks an official language but usually speaks a nonofficial language at home; L3 = speaks an official language and usually speaks an official language at home. The logit coefficients are statistically significant at .05, two-tailed test, unless indicated otherwise. The full regression equations with *t*-statistics are available on request.

a Excludes Atlantic provinces; see text for explanation.

b Variable not relevant.

c Variable defined with reference to Montreal in the equations for Quebec.

\# Not significant at .10

Source: 1991 Census of Canada, Public Use Microdata File (Individuals).

Exactly the same pattern of effects is associated with the linguistic distance measure. That is, where an immigrant's mother tongue is linguistically distant from English, and hence where it is expected that the immigrant will find it more difficult to learn English, it is less likely that the immigrant will be able to conduct a conversation in an official language; if he is able to do so, it is also less likely that he will speak an official language at home.

This specification of the measure of linguistic distance in Table 1.5 does not take account of the dual-language nature of Canada. Recall that the linguistic scores used in constructing the linguistic distance variable are defined with reference to English. We find a similarity in the signs and the statistical significance of the results when we obtain separate estimates for English Canada and for Quebec. However, the multinomial logit equations for the total sample with interaction terms between Quebec and the linguistic distance variable suggest that greater linguistic distance of the origin language from English has a greater adverse effect on official language practice in Quebec than in English Canada.[15] This finding would be consistent with the hypothesis that for immigrants in Quebec with a large linguistic distance from English (e.g., Asian and Middle Eastern languages), the reduced exposure to English reduces the shift to English or French.[16]

For the sample of immigrants living in Quebec, we also estimated a model in which a dichotomous variable for the Romance-language countries replaced the linguistic distance measure. The rationale is that French is closer (linguistically) to other Romance languages than to non-Romance languages. We defined this variable as unity for immigrants from Italy, Portugal, and Central and South America, who do not speak English as a mother tongue. With controls for refugee and colony variables, immigrants from Romance-language countries are more likely to be in categories L1 and L2 than in L3.[17] That is, they are more likely to retain their origin language in the home even if they can speak an official language. The findings based on the Romance-language variable, however, are markedly inferior to those presented for Quebec in Table 1.5. This suggests that the information on interlanguage variation in the linguistic distance variable, defined with reference to English, is relevant to language acquisition in this province. This situation may arise because the continuous linguistic distance variable defined with reference to English is superior to the dichotomous Romance-language variable even in Quebec, perhaps in part because English and French are quite close linguistically and because English is used extensively in Quebec.

Given the limitations on the information on country of birth in the Census of Canada, we define the refugee variable as unity for immigrants from Vietnam and from the (now) former Soviet Union. All other variables being the same, refugees are less likely to speak an official language; if they can speak an official language, they are less likely to report that they usually speak it at home (Table 1.5).

Immigrants from former colonies of Britain, France, or the United States, on the other hand, are more likely to speak an official language. Among immigrants who can speak an official language, however, those from a former colony are more likely than other immigrants to speak their origin language in Canada overall and in English Canada. In Quebec, although the colony effect on speaking an official language is larger than in English Canada, a colonial origin apparently has no effect on whether a speaker of an official language usually speaks one of these languages at home.

The patterns of language practice associated with minority language concentration, linguistic distance, geographic distance, refugee status, and former colony can be illustrated through the computation of predicted distributions similar to those presented in Table 1.4. We compute these predictions for two duration-of-residence categories: immigrants who have lived in Canada for five years and for 15 years. Predicted language probabilities are reported in Tables 1.6 through 1.8 for the five variables for Canada as a whole.

Table 1.6 reveals that residence in an area where the origin language is spoken more intensively is associated both with a higher probability of being unable to speak an official language (higher L1) and with retention of the mother tongue in the home even when the immigrant speaks an official language (higher ratio of L2 to L3). The effects of minority language concentration on retention of the mother tongue are smaller among those who have lived in Canada for a longer period. Consider the effect, evaluated at the mean for the other variables, at a minority language concentration of 2% (close to the mean of 2.2%): for example, Arabic speakers in Montreal. After five years in

Table 1.6 Predicted distributions across language categories by minority language concentration, by duration of residence[a]

		After 5 Years in Canada			*After 15 Years in Canada*		
Minority Lang. Concentration	*Example*[b]	*L1*	*L2*	*L3*	*L1*	*L2*	*L3*
0.0	–	6.25	71.89	21.86	1.65	56.25	42.09
1.0	Punjabi in Toronto	7.07	72.81	20.12	1.92	58.38	39.71
2.0	Arabic in Montreal	7.97	73.55	18.47	2.22	60.43	37.36
3.0	Portuguese in Toronto	8.97	74.11	16.92	2.55	62.39	35.06
4.0	Chinese in Sudbury	10.07	74.48	15.46	2.94	64.25	32.82
5.0	German in Kitchner	11.26	74.65	14.08	3.37	65.99	30.64
6.0	Italian in Windsor	12.57	74.63	12.80	3.85	67.61	28.54
7.0	Italian in Toronto	13.99	74.41	11.60	4.39	69.10	26.51
8.0	—	15.54	73.98	10.48	5.00	70.43	24.56
9.0	Chinese in Vancouver	17.20	73.36	9.45	5.68	71.62	22.70

Notes: L1 = speaks neither English nor French; L2 = speaks an official language but usually speaks a nonofficial language at home; L3 = speaks an official language and usually speaks an official language at home.

a Predictions computed at means of all variables other than duration of residence. Row totals for each immigration period may not add to 100.0 because of rounding.

b Examples of minority language concentration values that are close to designated value. There is no close value for 8.0. Mean value is 2.2.

Source: Estimates in Table 1.5.

Canada, 8% of adult immigrant men cannot speak English or French, and only 18% usually speak one or the other at home. The corresponding figures are 6% and 22% respectively for those living in areas where virtually nobody (0.0%) speaks their origin language. By 15 years in Canada, at a 2% concentration ratio, only 2% do not speak an official language, whereas 37% usually speak an official language at home. In contrast, for a group with a high concentration of a minority language—say, 6% (the concentration ratio for Italian speakers in Windsor)—at 15 years in Canada nearly 4% cannot speak English or French, and only 28% usually speak an official language at home. Thus Table 1.6 demonstrates that the proportion not usually using an official language at home (L1 and L2) rises and the proportion usually speaking an official language at home (L3) declines with an increase in the minority language concentration ratio.

Table 1.7 shows that the geographic distance between the country of origin and Canada exerts a substantial effect on fluency rates, particularly among the more recent arrivals. The inability to speak English or French decreases with distance, whereas the extent of usually speaking one of the official languages at home increases. With increases in duration of residence in Canada, proficiency increases for all distances from the origin. After five years in Canada, the predicted percentage who cannot speak English or French declines with distance, from 10.1% at 3,000 miles (approximately the distance from Bogotá to Toronto) to 6.6% at 7,000 miles (approximately the distance from Ho Chi Minh City to Vancouver). The percentage who usually speak English or French at home increases respectively from 14.4% to 22.5% for these distances. At 15 years' duration, the percentage usually speaking English or French at home increases from 30.9% to 43.1% with the increase in distance from 3,000 to 7,000 miles.

Table 1.7 Predicted distributions across language categories by miles between origin and Canada, by duration of residence[a]

		After 5 Years in Canada			*After 15 Years in Canada*		
Miles	*Example of Origin*[b]	*L1*	*L2*	*L3*	*L1*	*L2*	*L3*
2,000	Guadeloupe, WI	11.13	76.15	12.72	3.38	68.47	28.14
3,000	Bogota	10.09	75.56	14.35	2.98	66.11	30.90
4,000	Warsaw	9.11	74.74	16.14	2.62	63.58	33.80
5,000	Seoul	8.20	73.69	18.10	2.29	60.89	36.82
6,000	Jerusalem	7.36	72.41	20.23	2.00	58.07	39.94
7,000	Ho Chi Min	6.58	70.89	22.53	1.73	55.14	43.13
8,000	Bombay	5.86	69.14	24.99	1.50	52.13	46.38

Notes

a For notes to this table, see Table 1.6.

b Mileage approximate (within 10%) from designated city to nearest gateway city (Vancouver or Toronto) in Canada.

Source: Estimates in Table 1.5.

Cohorts with a longer duration of stay in Canada will contain fewer immigrants who expect to return to their country of origin. In part this is the case because many of those who had a high expectation of returning did return, while others, who stayed, reduced their expectation of returning before retirement. A decline in the average expectation of returning would be accompanied by greater investment in Canada-specific skills, and a greater investment in such skills would lower the propensity for return migration. This decline, with duration, in expectations of return migration will result in a smaller spread of predicted values across the "miles" measure in Table 1.7 at 15 years' duration of residence in Canada than at five years' duration.

Predicted distributions across language categories by linguistic score are reported in the top panel in Table 1.8. The three language scores listed are 1.38, 2.00, and 2.75, the respective scores for Chinese, Polish, and Dutch. It is apparent that language score exerts a major effect on fluency rates and retention of mother tongue. After five years in Canada, 13.1% of those with a language score of 1.38 (the value for Chinese speakers) but only 5.5% of those with a language score of 2.75 (the value for Dutch speakers) cannot speak English or French, while only 6.1% of the former and 32.2% of the latter usually speak one of the official languages at home. By 15 years in Canada, we predict that only 15.0% of Chinese speakers usually speak an official language at home and that 80.6% can speak an official language but usually speak their origin language at home. For the Dutch-origin speakers these predicted percentages are quite different: 55.3% and 43.4% respectively.

Table 1.8 Predicted distributions across language categories by linguistic score, refugee status, and birthplace a former colony by duration of residence[a]

		After 5 Years in Canada			*After 15 Years in Canada*		
		L1	*L2*	*L3*	*L1*	*L2*	*L3*
Linguistic Score[b]	Language						
1.38	Chinese	13.08	80.79	6.13	4.41	80.55	15.04
2.00	Polish	8.14	73.47	18.40	2.26	60.46	37.27
2.75	Dutch	5.48	62.30	32.22	1.29	43.42	55.28
Refugee Origin Status							
Refugee[c]		20.37	74.40	5.24	7.31	79.00	13.68
Other		8.22	73.71	18.07	2.30	60.93	36.77
Colonial Origin Status							
Colony[d]		4.33	78.93	16.74	1.21	64.91	33.88
Other		8.22	73.71	18.07	2.30	60.93	36.77

Notes

a For notes to this table, see Table 1.6.

b Language scores where Chinese is the lowest and Dutch is the highest for the languages identified in the data.

c Unity if from Vietnam or the former USSR.

d Unity if from South Asia, Vietnam, Africa, Hong Kong, or the Philippines.

Source: Estimates in Table 1.5.

These data tell a simple and compelling story. The reason why some immigrants have not learned an official language of Canada, or usually speak their origin language rather than an official language at home even if they report that they can speak the latter, is that it is much more difficult for them than for other immigrants to learn French or English. It follows that attainment of a given level of fluency will require more investment if a person's mother tongue is (for example) Chinese than if it is Polish; among the groups studied here, the least investment will be required if it is Dutch.

The predicted language practice for refugees and others is reported in the middle panel of Table 1.8. Refugees who have lived in Canada only a few years are far less likely to be able to speak an official language: 20% cannot do so, compared with 8% for others. They are far less likely to usually speak an official language at home: 5%, compared with 18% for others. Although official language skills and use increase with duration, even after 15 years we find a large gap based on refugee origin.

On the other hand, immigrants from former British, French, or American colonies (Table 1.8, bottom panel) are more proficient in an official language than others, but are more likely to usually speak their origin language at home. This holds at both five years and 15 years in Canada.

Vietnamese immigrants are an interesting group: they are classified here both as refugees and as immigrants from a former colony. As refugees they would have a lower level of proficiency, but origin in a former colony would enhance their proficiency in English or French. Evaluated at the means of other variables, in comparison with immigrants who are neither refugees nor from a former colony, the Vietnamese are less likely to use English or French at home, even those who can speak one of the official languages.[18]

Summary and conclusion

Language proficiency is an important aspect of immigrants' adjustment. Knowing what language or languages an immigrant can speak, those that the immigrant usually speaks at home, and the level of proficiency in the destination language provides important information on the immigrant's socioeconomic status and on the extent of integration into the social, political, cultural, and economic life of the majority population in the host country.

We have been concerned here with the determinants of destination-language proficiency among immigrants who came from a background with languages other than the primary or dominant languages of destination. In addition to the immigrant's own characteristics, we emphasize the characteristics of the origin as determinants of destination-language proficiency. We make the empirical application to immigrants in Canada, as reported in the 1991 Census.

The analyses show that age at migration (an efficiency variable), educational attainment (reflecting efficiency, economic, and wealth effects), and duration of residence in Canada (an exposure variable) are significant determinants of language practice. Immigration at an older age is associated with a lower probability of knowing an official language and usually, if such a language is known, of speaking it at home. Shifting to an official language increases with duration in Canada and with educational attainment. Language practice also varies considerably across birthplace groups.

Our analysis of the effect of birthplace is extended by substituting five variables that we constructed, in part, using information on birthplace and mother tongue. These variables are the geographic distance between the country of origin and Canada, the linguistic distance between the immigrant's mother tongue and English or French, the proportion of individuals living in the same region as the immigrant who speak his mother tongue, whether the person is a refugee, and whether the origin is a former British, French, or American colony. Unlike dichotomous variables for country of birth, these variables are subject to behavioral interpretations, and they permit greater understanding of the factors affecting language practice among immigrants.

A greater geographic distance implies a greater positive selectivity in migration and a lower probability of return migration (efficiency and economic effects). The results show that the greater the geographic distance between the country of origin and Canada, the more likely that an immigrant will be able to conduct a conversation in an official language, and the more likely that he will use an official language at home.

A greater linguistic distance between the mother tongue and the official languages implies a higher cost of acquiring English-/French-language proficiency (efficiency effect). Empirically a greater linguistic distance is associated with a lesser use of the official languages. This holds whether the analysis uses the linguistic distance measure developed for English or the Romance-language dichotomous variable used for Quebec.

An increase in the proportion of the population in the region where an immigrant lives that uses the immigrant's mother tongue implies greater ease of avoiding use of the official languages (exposure effect). Empirically it is associated with an increase in the probability that the immigrant will not be able to speak an official language. It is also associated with an increase in the chances that an immigrant who can speak an official language will usually speak the origin language at home.

We find refugees to be less likely than economic migrants to speak one of the official languages; even those who report that they can do so are less likely to do so, as a rule, at home. This may be the case because of refugees' less favorable selectivity, less planning for the move, or a lesser commitment to the destination (efficiency effects).

Immigrants from a former British, French, or American colony are more likely to have been exposed to English or French in their country of origin (exposure effect). We find them more likely to be able to speak an official language, but those who can speak an official language also are more likely, as a rule, to use their origin language at home.

Our analysis has implications for public policy. The foreign-born in Canada, or in any destination, will be more proficient in the destination language if the immigration policy focuses on younger immigrants, with higher levels of schooling, who are not refugees, and who come from a geographically more distant place, with exposure to the destination language in the country of origin. The adjustment is easier (less costly) if the languages of origin and destination are linguistically closer. The adjustment is slower, however, if the immigrants are concentrated in a small number of language groups or if they segregate themselves by language group than if they are linguistically diverse.

Some of these characteristics can be incorporated explicitly into an immigration policy. These include age at migration, level of schooling, motive for migrating (e.g., economic), and diversity of origins. All of these characteristics inform about immigrants' likely difficulty (cost) of acquiring destination-language skills.

The analysis also suggests additional avenues for research. This study has been limited to men, but a comparable analysis for women would be fruitful, particularly if it incorporated labor-supply and family-composition effects on language practice. Additional research is also needed on the mechanism by which linguistic concentrations retard destination-language proficiency. Do linguistic concentrations arise from labor market activities, household consumption activities, selective migration, or some combination of the three? If linguistic concentrations retard destination-language proficiency, do they enhance retention of the origin language? If so, which effect on language capital is greater? To what extent, if any, are labor market outcomes (e.g., earnings) influenced by living in a linguistic concentration area, independent of the indirect effects on the labor market outcomes of these concentrations through the person's own language skills? (See Chiswick and Miller 2002.) Moreover, is there a relation between the parents' degree of linguistic assimilation and the language proficiency, school performance, and earnings of their Canadian-born children? Finally, the measure of language proficiency used in this study was quite simple. Would other measures, particularly specific information on the degree of proficiency or test-based measures, yield similar results?

Appendix. Definitions of variables

Definition of population

This consists of foreign-born men from non-English-speaking countries, age 25 to 64. Nonpermanent residents (i.e., persons on a student authorization, employment authorization, or a minister's permit, or a refugee claimant) are excluded from the analysis because the 1991 Census Public Use Microdata File (PUMF) does not contain information on the year of entry into Canada for this group. Also excluded from the analysis are a small number of persons for whom data were not available on questions used in the construction of variables, and those living in the Yukon and Northwest Territories. Other sample exclusions are noted in the definitions. For further information on the data set, see Statistics Canada (1994).

Language practice (LANG)

LANG is a trichotomous variable. The first category (L1) comprises individuals who cannot conduct a conversation in English or French. The second category (L2) comparises individuals who can conduct a conversation in English or French, but usually speak a nonofficial language at home. The third category (L3) comprises those who can conduct a conversation in English or French and usually use an official language at home.

Years of education (EDUC)

This variable records the total years of full-time education. It is constructed from the Census information on total years of schooling for respondents who do not possess a university qualification. For individuals who possess a university qualification, we add the following years of full-time equivalent schooling to the years of secondary schooling: diploma below bachelor level (2.4 years); bachelor's degree (three years for those reporting three or fewer years of university, four years for all others); diploma above bachelor level (four years for those reporting four or fewer years of university, five years for all others); degree in medicine, dentistry, etc. (seven years); master's degree (six years); earned doctorate (eight years).

Years since migration (YSM)

The census information on year at arrival is presented in single years for some arrival cohorts, in small intervals for some cohorts in the non-Atlantic provinces, and in large intervals for the Atlantic provinces. We formed a continuous measure from this information by assigning midpoints to all arrival intervals and subtracting this value from 1991. A quadratic specification is used.

Birthplace (BIRTH)

The following countries or regions of birth are identified in the census file for immigrants living outside the Atlantic provinces (listed in order of numerical importance): United Kingdom, other Europe, Central and South America and Caribbean, Italy, United States, southern Asia (e.g., Bangladesh, India, Pakistan), Federal Republic of Germany, Africa, Poland, Middle East and western Asia (e.g., Turkey, Iran, Lebanon, Saudi Arabia), Portugal, other Eastern and southeast Asia (e.g., Japan, South Korea, Singapore, Thailand), People's Republic of China, Hong Kong, Philippines, Vietnam, USSR, and other. For immigrants resident in the Atlantic provinces, the only birthplace categories distinguished are United States, United Kingdom, other Europe, and other. Immigrants from the United Kingdom, the United States, and Central America, the Caribbean and Bermuda, and South America whose mother tongue is English are excluded from the analysis, given that study of language fluency focuses most appropriately on immigrants from non-English-speaking backgrounds. In addition, we exclude the small number of immigrants from the residual "other" birthplace category because a direct-line distance cannot be assigned to this group in the construction of the MILES variable (see below). Immigrants from Italy are used as the benchmark group.

Minority language concentration (CONC)

We assign each respondent a measure equal to the percentage of the population age 18 to 64 in the region where he lives that reports the same mother tongue as the respondent. We define the region using information on residence in a Census Metropolitan Area and province of residence. We use the CMA as well as the balance of the province (for those not living in a CMA). The nonofficial language groups of German, Netherlandic (e.g., Dutch), Italian, Spanish, Portuguese, Polish, Ukrainian, Greek, Chinese, Austro-Asiatic (e.g., Khmer, Vietnamese), Arabic, Punjabi, and other Indo-Iranian (e.g., Bengali, Hindi, Kurdish) are identified on the Census Public use Microdata File (Individuals). Those reporting another language (other than English or French) are assigned the value 0 on the assumption that the incidence of that language is trivial. Those reporting only English or French are assigned the mean value of the CONC variable for those from the same country-of-birth group.

Marital status (MARRIED)

This is a dichotomous variable that is set equal to 1 for individuals who are married (including common-law partners) and equal to 0 for all other marital states.

Location

We use two location variables in the study. The first records province of residence. This information was grouped as follows: Atlantic provinces (Newfoundland, Nova Scotia, New Brunswick, Prince Edward Island), Quebec, Ontario, prairie provinces (Manitoba, Saskatchewan, Alberta), and British Columbia. The second locality variable records the size of the place of residence. Individuals living in CMAs (defined as having a population of at least 100,000 based on the 1986 census) are distinguished from other individuals.

Age

Age is available in single years.

Linguistic distance (LD)

We construct this variable from a measure of the difficulty, for English-speaking Americans, of learning a foreign language. It is based on a set of language scores (LSs) measuring achievements in speaking proficiency by English-speaking Americans at the U.S. Department of State, School of Language Studies, reported by Hart-Gonzalez and Lindemann (1993). For the same number of weeks of instruction, a lower score (LS) represents less language facility and, we assume, greater linguistic distance between English and the foreign language in question. For example, Italian is scored 2.5 (in a range from 1 to 3); Arabic is scored 1.5. This methodology assumes symmetry across languages: that is, if English-speaking Americans find it difficult to learn a language, native speakers of that language find it equally difficult to learn English.

The scores reported by Hart-Gonzalez and Lindemann (1993) are matched to the mother tongue codes used in the PUMF as follows:

Table 1A.1

Mother Tongue	*Hart-Gonzalez and Lindemann Language*	*Linguistic Score*
English and/or French	—	See text
Aboriginal Languages	—	n.a.
German	German	2.25
Netherlandic	Dutch	2.75
Italian	Italian	2.50
Spanish	Spanish	2.25
Portuguese	Portuguese	2.50
Polish	Polish	2.00
Ukrainian	Russian	2.25
Greek	Greek	1.75
Chinese	Mandarin (1.5), Cantonese (1.25)	1.375
Austro-Asiatic Languages	Cambodian	2.00
Arabic	Arabic	1.50
Punjabi	Hindi	1.75
Other Indo-Iranian	Hindi	1.75
Other	—	See text

Note: n.a. = not available. The small number of foreign-born persons in this category is deleted from the sample.

In the construction of this variable, foreign-born persons who report English or French as the mother tongue, and persons in the "other" home language category, are assigned a value of the linguistic score on the basis of their country-of-birth group. The following values are used for the birthplace categories: Germany (German language, score of 2.25); Italy (Italian, 2.50); Portugal (Portuguese, 2.50); Poland (Polish, 2.00); USSR (Russian, 2.25), Hong Kong (Chinese, 1.375), China (Chinese, 1.375), Philippines (Tagalog, 2.00); Vietnam (Vietnamese, 2.00); other Europe (mean score of "other Europe" countries computed for prime-age male immigrants in the 1990 U.S. Census, 2.23); Middle East and western Asia (mean score of countries of Middle East and western Asia computed for prime-age male immigrants in the 1990 U.S. Census, 1.89); southern Asia (mean score of countries of southern Asia computed for prime-age male immigrants in the 1990 U.S. Census, 1.91); other Eastern and southeast Asia (mean score for these regions computed for prime-age male immigrants in the 1990 U.S. Census, 1.36); Africa (mean score of countries of Africa computed for prime-age male immigrants in the U.S. Census, 2.11); and Central America, Caribbean and Bermuda, and South America (mean score for these regions computed for prime-age male immigrants in the 1990 U.S. Census, 2.25).

The method of computing scores for the broad birthplace regions assumes that the birthplace distributions and language backgrounds of immigrants in the United States are the same as in Canada. This assumption may not be strictly valid, but it is preferable to the alternatives: excluding this sizable group from the study or assigning the mean of the language score for immigrants in Canada for whom valid scores could be computed.

The variable in the regression equations is linguistic distance, which is 1 divided by the linguistic score: that is, LD = 1/LS.

In the analyses pooled across all regions of Canada, we also use a shift variable for Quebec in conjunction with the linguistic distance measure. The specification is $\beta_0\text{LD} + \beta_1\text{QUEBEC} \times \text{LD}$.

In addition, a variable for having a Romance-language background (ROMANCE) is included in some specifications for Quebec.

Romance-language background (ROMANCE)

We define this variable as being born in a Romance-language country identified in the Census: Portugal, Italy, and Central and South America (excluding persons of English mother tongue).

Direct-line distances (MILES)

We define this as the miles between the major city in the immigrant's country of origin (or, for broad regions, a selected country within that region) and either Vancouver or Toronto, whichever distance is the smaller. For analyses limited to Quebec, the variable records the distance to Montreal. The distances are taken from data in Fitzpatrick and Modlin's (1986) *Direct Line Distances, International Edition*.

Refugee (REFUGEE)

This is a dichotomous variable equal to 1 for those born in Vietnam or the USSR, and equal to 0 for all other birthplaces.

Colony (COLONY)

This is a dichotomous variable equal to 1 for those born in a colony of Britain, the United States, or France: that is, South Asia, Vietnam, Africa, Hong Kong, and the Philippines. This variable is set equal to 0 for all other birthplaces.

Notes

1 We appreciate the comments received from the participants at the Conference on Immigration With an International Comparative Perspective, Vancouver, January 1999; the Applications Workshop, Department of Economics, University of Chicago, May 1999; the annual meeting of the European Society for Population Economics, Turin, June 1999; and the RIIM Seminar, Simon Fraser University, July 2000.

Partial financial support for this project was provided through a grant from the Embassy of Canada, Washington, DC, from the George J. Stigler Center for the Study of the Economy and the State, Graduate School of Business, University of Chicago, and from the Center for Excellence in Vancouver, Research on Immigration and Integration in the Metropolis (RIIM), Simon Fraser University. Opinions expressed are those of the authors and should not be attributed to the sponsors.

2 There is a large literature on the effects of destination-language skills on labor market outcomes in the major immigrant-receiving countries, including Australia, Canada, Germany, Israel, the Netherlands, the United States, and the United Kingdom. For studies of Canada, see, for example, Abbott and Beach (1987), Charette and Meng (1998), Chiswick and Miller (1988, 1992, 2003), Grenier (1987), and Shapiro and Stelcner (1997).

3 Other studies of the determinants of language skills among immigrants in Canada include Richmond and Kalback (1980) and De Vries and Valee (1980), who use cross-tabulations, as well as Breton (1978a, 1978b), Veltman (1983), Chiswick and Miller (1994a), and Charette and Meng (1998). For Canadian government reports on official languages, see Canadian Heritage (1999).

4 For an analysis of current Canadian immigration policy, see Green and Green (1995).

5 Studies using data from a variety of countries demonstrate the large positive effect of destination-language skills on immigrants' earnings. See, for example, research for the United States (Chiswick 1991; Chiswick and Miller 1995, 2002; McManus, Gould, and Welch 1983; Tainer 1988), for Australia (Chiswick and Miller 1995), for Canada (Chiswick and Miller 1988, 1992, 2003), for Israel (Berman, Lang, and Siniver 2000; Chiswick 1998), for the Netherlands (Kee and van Ophem 1996), for Germany (Dustmann 1994), and for the United Kingdom (Shields and Wheatley-Price 1999). We know of no empirical research on the effect of destination-language skills on the prices that immigrants pay as consumers.

6 A *New York Times* article (Hedges 2000) reported on children's role as translators

for their immigrant parents in the United States, and on the negative effect on their parents' acquisition of English-language skills.

7 The finding, among immigrants in Israel, that schooling has a positive effect on proficiency in Hebrew (Chiswick 1998) suggests that exposure to the destination language during schooling before immigration is not the primary mechanism explaining the relation between schooling and destination-language skills in English-speaking countries. See also Beenstock (1996).

8 For an analysis of the language questions in the Census of Canada and recommendations for improvements, see Chiswick and Miller (1998b).

9 Gender differences in immigrants' language proficiency are a primary focus of Stevens (1986) and Chiswick and Miller (1994b).

10 French-speaking birthplaces (e.g., France and Belgium) are not identified separately in the Canadian Census. Few immigrants to Canada come from French-speaking countries.

11 The minority language concentration index contains substantial positive skewness. It cannot be entered in a logarithmic transformation because of the zero values. Tests regarding more complex specifications yield essentially the same results as those obtained from the simple specification used in this analysis.

12 The quadratic effect of duration on language practice follows from a human capital investment model. The lower wage (opportunity cost of time) of a new immigrant, a longer period in which to receive benefits, and an incentive to make profitable investments sooner rather than later all encourage immigrants to make their investments in language skills as soon as they arrive. Returns eventually grow smaller, the more investments are made in a given unit of time; therefore these investments are not made instantaneously on arrival. Also, particularly in regard to language capital, much learning is done by doing. Hence the greatest investments are made in the early period, and the intensiveness of investments diminishes over time. This translates into an increase in proficiency, but at a rate that decreases with duration of residence.

13 This may be a consequence of the discrepancies between the data available in the Individuals file (see appendix) and the ideal variable (see Chiswick and Miller 1992) that distinguishes between immigrants married to an official-language speaker and those who are married to a person with whom they share a mother tongue.

14 For approximately 32% of the sample, we constructed the linguistic distance measure using information on birthplace. For about 12% of the sample, we constructed the minority language concentration measure using information on birthplace.

15 The coefficient on LD in the model with the interaction term for log (L2/L1) is −1.328 (t = 3.91); for log (L3/L1) it is −6.887 (t = 18.66). The coefficients on the interaction term between Quebec and LD respectively are log (L2/L1) −2.959 (t = 4.05) and log (L3/L1) −1.337 (t = 1.74).

16 Chiswick and Miller (1994a) show that immigrants in Canada, whether in English Canada or in Quebec, tend to acquire proficiency in English rather than in French; in Quebec, however, immigrants from Romance-language countries are more likely to acquire French proficiency or to become official-language bilinguals.

17 The coefficient (and t-ratio) on the Romance-language variable in the equation for log (L2/L1) is 0.033 (t = 0.11). In the equation for log (L3/L1) it is −1.400 (t = 4.36). Therefore log (L3/L2) = −1.433.

18 Predicted values of language categories by duration in Canada, computed from Table 1.6:

	5 Years			*15 Years*		
	L1	*L2*	*L3*	*L1*	*L2*	*L3*
Vietnamese	11.29	83.63	5.09	3.81	83.65	12.53
REFUGEE = 0 and COLONY = 0	9.60	70.94	19.46	2.66	58.10	39.24

References

Abbott, M.G. and C.M. Beach. 1987. "Immigrant Earnings Differentials and Cohort Effects in Canada." Discussion Paper 705, Institute for Economic Research, Queen's University, Kingston, ONT.

Becker, G.S. and B.R. Chiswick. 1966. "Education and the Distribution of Earnings." *American Economic Review* 56:358–69.

Beenstock, M. 1996. "The Acquisition of Language Skills by Immigrants: The Case of Hebrew in Israel." *International Migration* 4:3–30.

Berman, E., K. Lang, and E. Siniver. 2000. "Language-Skill Complementarity: Returns to Immigrant Language Acquisition." Department of Economics, Boston University. Photocopied document.

Breton, A. 1978a. *Bilingualism: An Economic Approach*. Montreal: C.D. Howe Research Institute.

—— 1978b. "Nationalism and Language Policies." *Canadian Journal of Economics* 11:656–68.

Canadian Heritage. 1999. *New Canadian Perspectives: Official Languages Research Publications, Publications Catalogue*. Ottawa: Canadian Heritage.

Charette, M.F. and R. Meng. 1998. "The Determinants of Literacy and Numeracy, and the Effect of Literacy and Numeracy on Labour Market Outcomes." *Canadian Journal of Economics* 31:495–517.

Chastain, K. 1975. "Affective and Ability Factors in Second-Language Learning." *Language Learning* 25(1):153–62.

Chiswick, B.R. 1991. "Speaking, Reading and Earnings Among Low-Skilled Immigrants." *Journal of Labor Economics* 9:149–70.

—— 1998. "Hebrew Language Usage: Determinants and Effects on Earnings Among Immigrants in Israel." *Journal of Population Economics* 11:253–71.

—— 1999. "Are Immigrants Favorably Self-Selected?" *American Economic Review* 89:181–85.

Chiswick, B.R. and P.W. Miller. 1988. "Earnings in Canada: The Roles of Immigrant Generation, French Ethnicity, and Language." Pp. 183–228 in *Research in Population Economics*, Vol. 6, edited by T.P. Schultz. Greenwich, CT: JAI.

—— 1992. "Language in the Labor Market: The Immigrant Experience in Canada and the United States." Pp. 229–96 in *Immigration, Language and Ethnic Issues: Canada and the United States*, edited by B.R. Chiswick. Washington, DC: American Enterprise Institute.

—— 1994a. "Language Choice Among Immigrants in a Multi-Lingual Destination." *Journal of Population Economics* 7:119–31.

—— 1994b. "Language and Labor Supply: The Role of Gender Among Immigrants in Australia." *Research on Economic Inequality* 17:153–89.

—— 1995. "The Endogeneity Between Language and Earnings: International Analyses." *Journal of Labor Economics* 13:246–88.

—— 1996. "Ethnic Networks and Language Proficiency Among Immigrants." *Journal of Population Economics* 9:16–35.

—— 1998a. "English Language Fluency Among Immigrants in the United States." *Research in Labor Economics* 17:151–200.

—— 1998b. "Census Language Questions in North America." *Journal of Economic and Social Measurement* 25:73–95.

—— 2002. "Immigrant Earnings: Language Skills, Linguistic Concentrations and the Business Cycle." *Journal of Population Economics* 15:31–57.

—— 2003. "The Complementarity of Language and Other Human Capital: Immigrant Earnings in Canada." *Economics of Education Review* 22:469–80.

Crystal, D. 1987. *The Cambridge Encyclopedia of Language*. Cambridge, UK: Cambridge University Press.

De Vries, J. and F.G. Vallee. 1980. *Language Use in Canada*. Ottawa: Statistics Canada.

Dustmann, C. 1994. "Speaking Fluency, Writing Fluency and Earnings of Migrants." *Journal of Population Economics* 7:133–56.

Espenshade, T.J. and H. Fu. 1997. "An Analysis of English Language Proficiency Among U.S. Immigrants." *American Sociological Review* 62:288–305.

Fitzpatrick, G.L. and M.J. Modlin. 1986. *Direct-Line Distances*. International ed. Metuchen, NJ: Scarecrow Press.

Gardiner, R.C. 1990. "Attitudes, Motivation, and Personality as Predictors of Success in Foreign Language Learning." Pp. 179–221 in *Language Aptitude Reconsidered*, edited by T.S. Parry and C.W. Stansfield. Englewood Cliffs, NJ: Prentice-Hall.

Green, A.G. and D.A. Green. 1995. "Canadian Immigration Policy: The Effectiveness of the Point System and Other Instruments." *Canadian Journal of Economics* 28:1006–41.

Grenier, G. 1987. "Earnings by Language Group in Quebec in 1980 and Emigration From Quebec Between 1976 and 1981." *Canadian Journal of Economics* 24:774–91.

Grenier, G. and F. Vaillancourt. 1983. "An Economic Perspective on Learning a Second Language." *Journal of Multilingual and Multicultural Development* 4:471–83.

Grimes, J.E. and B.F. Grimes. 1993. *Ethnologue: Languages of the World*. 13th ed. Dallas: Summer Institute of Linguistics.

Guindon, R. and P. Poulin. 1998. *Francophiles in Canada: A Community of Interests*. Ottawa: Canadian Heritage.

Hart-Gonzalez, L. and S. Lindemann. 1993. "Expected Achievement in Speaking Proficiency, 1993." School of Language Studies, Foreign Services Institute, Department of State, Washington, DC. Mimeographed document.

Hedges, C. 2000. "Translating America for Parents and Family." *New York Times*, June 19, p. A19.

Kee, P. and H. von Ophem. 1996. "Immigrant Wages in the Netherlands: The Role of Dutch Language Proficiency." Pp. 62–84 in *Migrants in the European Labor Market*, edited by S. Gazioglu. Aderdeen: J-Net.

Long, M.H. 1990. “Maturational Constraints on Language Development.” *Studies in Second Language Acquisition* 12:251–85.

McManus, W., W. Gould, and F. Welch. 1983. “Earnings of Hispanic Men: The Role of English Language Proficiency.” *Journal of Labor Economics* 1:101–30.

Newport, E.L. 1990. “Maturational Constraints on Language Learning.” *Cognitive Science* 14(1):57–77.

Richmond, A.H. and W.E. Kalback. 1980. *Factors in the Adjustment of Immigrants and Their Descendants*. Ottawa: Statistics Canada.

Service, E. and F.I.M. Clark. 1993. “Differences Between Young and Older Adults in Learning a Foreign Language.” *Journal of Memory and Language* 32:608–23.

Shapiro, D.M. and M. Stelcner. 1997. “Language and Earnings in Quebec: Trends Over Twenty Years, 1970–1990.” *Canadian Public Policy/Analyse de Politiques* 23(2):115–40.

Shields, M.A. and S. Wheatley-Price. 1999. “The English Language Fluency and Occupational Success of Ethnic Minority Immigrant Men Living in English Metropolitan Areas.” Presented at the CEPR Conference on Marginal Labour Markets in Metropolitan Areas, October 10–12, Dublin.

Statistics Canada. 1994. *User Documentation for Public Use Microdata File on Individuals, 1991 Census of Canada*. Ottawa: Statistics Canada.

Stevens, G. 1986. “Sex Differences in Language Shift in the United States.” *Sociology and Social Research* 71(1):31–36.

—— 1994. “The English Language Proficiency of Immigrants in the U.S.” Pp. 63–86 in *Immigration and Ethnicity: The Integration of America’s Newest Immigrants*, edited by J. Passel and B. Edmonston. Washington, DC: Urban Institute.

Svanes, B. 1987. “Motivation and Cultural Distance in Second-Language Acquisition.” *Language Learning* 37:341–59.

Tainer, E. 1988. “English Language Proficiency and Earnings Among Foreign-Born Men.” *Journal of Human Resources* 23(1):108–22.

Veltman, C.J. 1983. *Language Shift in the United States*. New York: Mouton.

2 Language skill definition

A study of legalized aliens[1]

The development of English language skills is an important part of the economic, social and political adjustment of immigrants in the United States. Immigrants differ systematically in the rate at which they acquire English language skills, and this affects rates of economic success, social integration, and political involvement. It has been shown in a number of studies that English language skills are greater among the better educated, those who migrate while young, those with a longer duration of residence in the United States, those who married after migration, and those who reside outside areas of concentration of immigrants with whom they share a mother tongue (see Chiswick and Miller, 1992, 1995, 1998, 1999). These studies also show that English language skills vary directly with the physical distance of the country of origin from the United States, vary inversely with the expected propensity for return migration, and vary inversely with the linguistic distance between English and the immigrant's mother tongue. Among women, English proficiency varies inversely with the age and number of their children living at home.

The research findings noted above are useful in a range of policy contexts. It has been argued, for example, that countries that select immigrants via a point system that emphasizes destination language fluency, educational attainment and age at migration, such as Canada and Australia, will have higher rates of destination language fluency among their immigrants, *ceteris paribus*, than countries, such as the United States and Israel, that do not select their immigrants on such a basis. It has also been argued that as immigrants' English language skills improve markedly with the passage of time there is less need for a more proactive language policy by governments.

The studies to date have been based largely on self-reported measures of language skills that have a general focus. For example, in the 1990 U.S. Census respondents who spoke a language other than English at home were asked to report their English-speaking proficiency as "Very Well," "Well," "Not Well," or "Not at All." The extent to which the responses to this type of question carry over to particular situations is not known. For example, a person who self-rated his English skills favorably in response to a general, context-free question (such as in the census) may report differently when asked about language proficiency in a given situation, such as the ability to speak to a doctor or a sales clerk. Moreover, the language skills that may matter for economic well-being may be the specialized language of the workplace, and the determinants of such language skills may differ from the factors that affect general language usage. Indeed, given the ordering of the language questions in the census, respondents may interpret the fluency question as referring to fluency in speaking English at home. A study of language proficiency at work may provide a more relevant focus for the analysis of labor market outcomes.

This study uses data from the Legalized Population Survey (LPS) to analyze a range of measures of the language skills of "legalized aliens," that is, aliens who received amnesty under Section 245A of the 1986 Immigration Reform and Control Act (IRCA). Included are a general measure of speaking proficiency, self-assessed speaking and reading proficiency in particular situations, perceptions of whether language skills are impeding job opportunities, and speaking and reading skills in the workplace. These data are in the form of an initial survey (1989) and a follow-up survey (1992) of a subset of the initial group of respondents.[2]

In the next section the data on language skills in the LPS are presented in the form of cross-tabulations. It is shown that the information on language skills is internally consistent. Next, a model of language acquisition is outlined and estimates of this model are obtained. Several measures of language proficiency are used in the study of the 1992 data in order to test the robustness of the model with respect to the definition of the dependent variable. The results from the application of the model in 1989 and 1992 are compared, and the comparison is used to highlight salient features of the development of language skills over the three-year period. Models of the perceived English reading and speaking proficiency in the workplace are also estimated. The econometric analysis is based on ordinary least squares with a correction for heteroscedasticity in the residuals and ordered probit analysis. Finally, the study is summarized and conclusions are drawn.

Descriptive analysis of language skills

The analyses reported in this chapter are based on the Legalized Population Survey Public Use Sample, made available by the Immigration and Naturalization Service (INS). This data set contains information on 6,193 aliens who attained temporary legal status under IRCA Section 245(A). To qualify under the provisions of this Act, aliens must have lived in the United States "continuously" in an illegal status since before January 1, 1982, and had to apply to the INS for the temporary legal status which after one year would result in permanent legal (resident alien) status. Approximately 1.8 million persons qualified for temporary legal status under these provisions. Most applications were filed between May 1987 and May 1988. The sample drawn from this population was collected using a two-stage stratified cluster design. (In the first stage, 40 legalization offices were selected, in the second stage, subsamples of applicants within sampled legalization offices were selected.) Only individuals 18 years or older were interviewed. Interviews were conducted in the first half of 1989. The sample contains data on demographic characteristics, language proficiency, immigration details (number of times entered the United States, year of entry, reasons for staying in the United States, reasons for leaving the United States, country or region of citizenship), state of residence in the United States, employment prior to entering and in the United States, family composition, health, use of social services, education, and income, among other variables (Westat, 1992).

The respondents in the LPS (1989) from non-English speaking countries tend to be young (about 34 years of age) and not well educated (on average about 7 years of schooling). They are primarily of Mexican origin (about 70%) with a large contingent from Central and South America (18%). Thus, for about 90 percent Spanish was their mother tongue. On average they first came to the United States about twelve years before they were interviewed in 1989.

Approximately two-thirds (4,012 individuals) of the original sample were reinterviewed in 1992 based on random sampling with some sample attrition. The second panel from the LPS covers a range of aspects, many of which build upon the information collected in the first panel in 1989. For example, information on labor market activity, education undertaken since applying for temporary legal residence, and language proficiency was collected. The information on language use in the second panel of the LPS is more comprehensive than that contained in the first panel, which was analyzed in Chiswick and Miller (1999). For example, information was collected in 1992 but not in 1989 on English fluency using a question similar to that used in the 1990 U.S. Census where English-speaking skills are categorized as "Very Well," "Well," "Not Well," or "Not at All." Data were also collected in a form similar to that used in the first panel on the respondent's ability to read and speak English in specific circumstances (*e.g.*, whether the respondent could speak in English to a doctor). In addition, qualitative information was collected on whether the respondent's English-speaking skills were believed by the respondent to limit job opportunities, the respondent's participation in English language classes, and the respondent's use of English in the workplace. These data provide the basis for the detailed analyses presented in this study.

The way the responses to the various English-language usage questions are treated, however, is conditioned by the "skip" patterns in the questionnaire. For example, in LPS2 respondents were first asked whether they spoke a language other than English at home. If respondents spoke only English at home, English fluency was assumed and no further questions on language skills were asked. Respondents who spoke a language other than English, however, were asked whether their English skills had limited their job opportunities (compared to persons born in the United States) and how well they speak English. As previously noted, the latter question is similar to that in the U.S. Census, with responses being coded as Very Well, Well, Not Well, or Not at All.

Individuals reporting that they spoke English Very Well in LPS2 were not asked further information on their language skills. Those who reported their proficiency in English as Well, Not Well, or Not at All were asked six questions relating to their language fluency in specific circumstances.[3] These were:

- If you have to read in English, can you read and understand a newspaper written in English?
- If you have to read in English, can you read and understand a magazine written in English?
- If you have to read in English, can you read and understand a recipe written in English?
- If you have to speak in English, can you speak in English with a sales clerk?
- If you have to speak in English, can you speak in English with a doctor, nurse, or teacher?
- If you have to speak in English, can you speak in English on the telephone?

To accommodate the skip patterns in the LPS2 data, for the purposes of this study all individuals who speak only English at home (4% of the sample) are held to speak English Very Well and to be able to read and speak English in the specific circumstances listed. All individuals who self-report their English-speaking proficiency as Very Well (10% of the sample) are held to be able also to read and speak English in the specific circumstances listed. In all other cases (86% of the sample) the assessment of English proficiency in the six specific situations listed is determined strictly by the data.

Respondents who were working were asked how much they communicate in English in the workplace. Individuals who did not use English at work "all the time" were asked whether they could read (where required) and understand work-related manuals and written instructions in English and (where required) communicate with their supervisor or transact business in English. In this instance, individuals who used English at work all the time are assumed to be able to read work-related materials and to be able to communicate in English in the workplace.

A number of issues are addressed in this descriptive overview. First, are the responses to the general census-type question on language skills and the responses to the specific questions consistent? The answer to this question is important as a number of researchers have reservations about the accuracy of self-reported language fluency (*see, e.g.*, Charette and Meng, 1994).[4] Using alternative measures of the type reported here provides a way of assessing the reliability of this information. Second, do the skill levels of English speaking and reading in the workplace diverge from those in general usage? Third, how has language proficiency changed between the two "panels" of data (1989 and 1992)? Both the development of language skills and the atrophy of skills are considered.

Table 2.1 presents data on the distribution of the sample of legalized aliens across the language proficiency categories used in the 1990 U.S. Census. Recall that respondents are providing information on a general question on English-speaking ability. These data reveal that 14 percent of the sample self-reported their language-speaking proficiency as Very Well (including the 4% who spoke only English), 22 percent have categorized themselves as being able to speak English Well, 42 percent of the sample have self-classified their English-speaking ability as Not Well, while 22 percent reported that they were not able to speak English at all. Elsewhere, Chiswick and Miller (1998) construct a dichotomous variable from the census data, where the categories Very Well and Well are distinguished from the categories Not Well and Not at All. Chiswick and Miller (1998) describe the first group as being proficient in English. Using this definition, around 36 percent of the sample of legalized aliens are fluent in spoken English.

Are the responses to this general question on English-speaking ability consistent with the responses to the specific questions on language ability? Table 2.2 lists information on English-speaking proficiency cross-classified by whether the legalized aliens could read or speak English in specific circumstances. Three patterns are evident in these data. First, a negligible percentage (between 1.2 and 3.4%) of individuals who said that they could not read or speak English in the particular situations mentioned in Table 2.2 are classified as fluent under the more general definition. Second, around 5 percent of those who indicated that they could cope in English in the six scenarios listed are classified as not being able to speak English at all under the more general question. Third, approximately 40 percent of the group who indicated that they could cope in English in the scenarios listed indicated that their general ability in English was Not Well. This might be expected,

Table 2.1 Distribution of legalized aliens across English proficiency categories, 1992[a]

English Language Proficiency	*Percent Distribution*	*Sample Size*[b]
Very Well[c]	14.0	563
Well	21.7	872
Not Well	42.0	1683
Not at All	22.3	893
Total	100.0	4013

Notes

a Percentages derived from weighted data, computed using only valid responses.

b Sample size refers to the number of respondents and not the weighted sample size which is approximately 234 times greater.

c Includes the 4 percent who speak only English at home.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice.

Table 2.2 Distribution of legalized aliens across English proficiency categories, by responses to English reading and speaking ability in specific situations, 1992[a] (percentage)

	English Speaking Proficiency					
	Very Well	*Well*	*Not Well*	*Not at All*	*Total*	*Sample Size*[b]
If you have to read in English, can you read and understand:						
1. A newspaper written in English?						
Yes	23.6	34.0	36.5	5.9	100.0	2388
No	[c]	3.4	50.1	46.5	100.0	1615
2. A magazine written in English?						
Yes	23.8	34.4	36.4	5.4	100.0	2366
No	[c]	3.4	50.0	46.6	100.0	1638
3. Recipes written in English?						
Yes	20.7	30.6	39.4	9.2	100.0	2715
No	[c]	3.0	47.4	49.6	100.0	1284
If you have to speak English, can you speak in English:						
1. With a sales clerk?						
Yes	20.0	30.3	42.9	6.8	100.0	2821
No	[c]	1.2	39.6	59.2	100.0	1180
2. With a doctor, nurse or teacher?						
Yes	20.9	31.1	41.4	6.5	100.0	2700
No	[c]	2.2	42.9	54.9	100.0	1303
3. On the telephone?						
Yes	21.2	32.0	41.3	5.5	100.0	2658
No	[c]	1.5	43.0	55.5	100.0	1338

Notes

a Percentages derived from weighted data, computed using only valid responses to each question.

b Sample size refers to the number of respondents and not weighted sample size, which is approximately 234 times greater.

c Cell is zero by construction of the survey instrument.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice.

however, on the basis of a difference between command of English under a given situation (being able to talk with a sales clerk) and in general. The Table 2.2 data, therefore, are indicative of a high degree of consistency in the responses to the various questions on language use in the LPS.

Respondents were also asked to report how much they speak English at work. Possible responses (with percentage distributions in parentheses) were: "All the Time" (30.8%), "Most of the Time" (14.6%), "One-half of the

Time" (15.3%), "Very Little" (24.1%), or "Not at All" (15.1%). These data indicate a solid core of individuals using English at work and also a considerable proportion reporting very little or no use of English at work. Among those who did not use English all the time, about two-thirds reported that they could read manuals written in English, with equal proportions of the balance of the relevant population indicating they could not read manuals written in English or were not required to be able to do this. The percentage distribution across response categories to the question on the ability to communicate with supervisors/transact business in English is similar to that for the question on the ability to read manuals written in English.

Table 2.3 panel A lists information on the percentages responding in the affirmative to the six specific situational questions in 1989 and 1992, together with the net changes over this period. Panels B and C in Table 2.3 report the cross-tabulation of proficiency in 1989 and 1992 for two of these situations, reading a newspaper and speaking to a sales clerk. The data show that there was a net improvement of around three percentage points in English-reading proficiency and between five and six percentage points in English-speaking proficiency. Given the absence of comparable "panel" data sets, assessment of the improvement in English proficiency is difficult.

These net improvements are, of course, the result of improvements of some individuals and deteriorations for others. Panels B and C in Table 2.3 indicated that for both reading a newspaper and speaking with a sales clerk (and in the four other situations not reported in the Table), the net changes in proficiency mask larger gross changes. Improvements in proficiency from No to Yes were reported by around 11 percent of the respondents for reading skills and about 13 percent for speaking skills. Declines in situational proficiency (from Yes to No) occurred for about 9 percent of the sample for reading skills and 7 percent for speaking skills. Expressed differently, between 26 and 36 percent of the group self-reporting that they could not read or speak English in the situations described in 1989 reported they could do so three years later (Table 2.4). Between 11 and 15 percent of individuals reporting that they could read or speak English in specific situations in 1989, however, indicated that they could not do so three years later. The characteristics of this group will be analyzed in further detail below.

Of particular interest is whether there is any evidence that atrophy provides an explanation for the changes between "proficient" and "deficient" over the three-year period. That is, just as labor market skills may atrophy during periods of absence from labor market activities, competency in reading and speaking English may also deteriorate when these skills are not practiced. On the other hand, it should be noted that the modal proficiency categories in 1992 of those whose self-reported status changes from proficient to deficient is Not Well. Hence, it is possible that what is being recorded in the data is simply a slight change in the degree of dominant language proficiency or a slight change in the benchmark by which they assess their own skills with a greater exposure to the labour market and life in the United States. Indeed, if

Table 2.3 Summary information on responses to language questions in the Legalized Population Survey, 1989 and 1992[a]

Panel A: Responses to Six Questions

Questions	*Percent responding "Yes"* 1989	1992	Change
Can you read and understand a newspaper written in English?	57.0	59.7	2.7
Can you read and understand a magazine written in English?	56.5	59.1	2.6
Can you read and understand instructions or recipes written in English?	64.7	67.9	3.2
Can you speak to a sales clerk in English?	65.1	70.5	5.4
Can you speak to a doctor, nurse or teacher in English?	62.6	67.5	4.9
Can you speak on the telephone in English?	60.1	66.5	6.4

Panel B: Responses to able to read and understand a newspaper written in English (percentage)

1989	*1992*		
	Yes	*No*	*Total*
Yes	48.5	8.5	57.0
No	11.2	31.8	43.0
Total	59.7	40.3	100.0

Panel C: Responses to able to speak to a sales clerk in English (percentage)

1989	*1992*		
	Yes	*No*	*Total*
Yes	57.9	7.2	65.1
No	12.6	22.3	34.9
Total	70.5	29.5	100.0

Notes

a The form but not the substance of the survey questions differ from that used in the table. Percentages derived from weighted data, computed using only valid responses to each question and are restricted to individuals represented in the second panel of data.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice.

legal status results in a greater sense of security and hence greater involvement with the general society rather than an ethnic linguistic enclave, there may be a tendency towards a downward assessment of language skills for the same level of objective proficiency.

Table 2.4 Gross changes in language proficiency categories between 1989 and 1992

Question[a]	*Percent of Deficient Changing to Proficient*	*Percent of Proficient Changing to Deficient*[b]
Can you read and understand a newspaper written in English?	26.0	14.9
Can you read and understand a magazine written in English?	25.0	14.6
Can you read and understand instructions or recipes written in English?	34.5	13.7
Can you speak to a sales clerk in English?	36.1	11.1
Can you speak to a doctor, nurse or teacher in English?	33.0	11.8
Can you speak on the telephone in English?	33.2	11.5

Notes

a The form but not the substance of the survey questions differ from that used in the table.

b Percentages derived from weighted data, computed using only valid responses to each question and are restricted to individuals represented in the second panel of data.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice.

The language model

The models of language fluency estimated in this section are based on the model proposed by Chiswick and Miller (1992, 1995, 1998, 1999). This model is structured around three main sets of conceptual variables; economic incentives, exposure, and efficiency. The empirical measures of these conceptual variables are discussed in detail in Chiswick and Miller (1992, 1995, 1998).

The schema of the language model may be described as:

LANG = f (economic incentives, exposure, efficiency),

where LANG is a measure of the alien's fluency in the dominant language. The empirical counterpart of this conceptual equation used in the study of the LPS is a modification of the model and is outlined in Chiswick and Miller (1998), namely:

LANG = f [duration in destination (+), marital status (?), age at migration (–), education (+), language concentration (–), relatives in U.S. (–), children (?), location (?), country of origin (?)],

where the expected partial effects (positive or negative) are indicated in parentheses and a question mark (?) indicates that the model does not offer an unambiguous sign.

The length of time that an alien has resided in the United States provides a measure of the alien's exposure to English. Many studies of immigrant adjustment have shown that as length of residence in the United States increases, immigrants adjust to the specific conditions of U.S. society. Part of the adjustment process appears to involve the learning of English. This is presumably due to the interactions that come about as part of everyday life in the country of destination, as well as specific investments that are made in language training, such as English as a second language (ESL) programs. It is expected that a similar process is relevant for illegal aliens and, therefore, that language skills will improve with duration of residence in the United States, even when the legalized aliens were in an illegal status (Chiswick, 1991). To capture the nonlinear effect, that is, that duration has its largest impact on language skills in the first few years and its marginal effect diminishes over time, duration of residence is entered in the model in quadratic form.

Age and educational attainment are expected to impact the individual's efficiency in learning English. It is well established in the linguistics literature (see, *e.g.*, Long, 1990) that the very young have a superior ability to acquire language skills. The greater efficiency in language acquisition of the better educated may arise because they have a greater mastery of their mother tongue and are more efficient in learning new concepts and new terminology. Or it may be that those with greater ability have a lower cost and greater benefits from investments in both schooling and language capital. Furthermore, those with schooling in the destination would be expected to be more fluent in the destination language if fluency is a prerequisite for school enrollment and if the destination schooling itself would enhance fluency (see Espenshade and Fu, 1997; for an analysis of postmigration schooling in the United States, see Chiswick and Sullivan, 1995; Khan, 1997).

The minority language concentration variable is included in the estimating equation to capture an ability to avoid exposure to English per unit of time in the destination. It has been measured in several previous studies by the extent to which the individual's mother tongue is spoken in the area in which the respondent lives, whether by immigrants or natives. It is hypothesized that the greater the extent of this phenomenon, the easier it is to avoid using English and hence the poorer the English language skills.

Marital status is likely to impact language skill in a number of ways. If the person was married prior to entry into the United States, the person would typically share a mother tongue with the spouse.[5] In this situation, opportunities for conversations in that mother tongue within the home substitute for conversations in English, and thus both reduce the need to learn English and limit the learning-by-doing that may otherwise take place. If the individual married after arrival in the United States, however, it is more likely that the marriage was to a person not monolingual in the individual's origin language. This would tend to enhance the individual's proficiency in English.[6]

The preceding discussion has focused on the key determinants of English language fluency. A host of other variables can be included in the analysis but are not included here as they are not germane to the general discussion, and some have been considered elsewhere (*e.g.*, for the effect of children on parental language fluency, see Chiswick and Miller, 1992, 1995, 1998, for an analysis using census data; Chiswick and Miller, 1999, for an analysis using the LPS).

Estimates of the language model

Preliminary estimations, 1989

Table 2.5 contains the means and standard deviations of the variables for the initial sample (full sample) interviewed in 1989 and the subsample interviewed in 1992. It is readily apparent that the two samples are quite similar. Hence, even though only approximately two-thirds of the initial sample was in the second panel, the individuals included in the second survey appear to be broadly representative of the initial set of respondents.

Another way of assessing whether the respondents included in the follow-up survey are representative of all respondents in the initial sample is to estimate models of speaking and reading fluency for the two groups using only the data collected in 1989.[7] Table 2.6 for males and Table 2.7 for females present these estimates using all respondents and only those respondents who were interviewed in the 1992 follow-up survey. The dependent variable in the equation for English-speaking proficiency is unity for those who can converse with a doctor, nurse, or teacher, while the dependent variable in the equation for English-reading proficiency is unity for those who can read a newspaper written in English. The two sets of estimates for speaking skills and reading skills are almost identical.[8]

The comparisons contained in Tables 2.5, 2.6, and 2.7 indicate that the data for 1992 provide a solid basis for study of the determinants of language skills. As a result, a number of equations are estimated using the 1992 data with different definitions of the dependent variable. These show that the main findings from the analysis are not sensitive to choice of the measure of language proficiency. Then, the results from analysis of the data in 1992 are compared to the results for 1989. The changes in mean levels of language fluency (reading skills, speaking skills) between the three years are decomposed into a component that is due to changes in duration over the three years and to changes in other characteristics of the sample between the two surveys, as well as a component attributable to changes in the way that characteristics affect language skills. Finally, the determinants of workplace reading and speaking proficiency among those in the labor market are analyzed.

Table 2.5 Means and standard deviations, selected variables, Legalized Population Survey, 1989[a]

Variable	*Males*		*Females*	
	Full Sample	*Follow-Up Sample*	*Full Sample*	*Follow-Up Sample*
Age	33.49 (8.94)	33.48 (8.73)	34.21 (9.33)	34.44 (9.44)
Educational Attainment	8.00 (4.34)	7.99 (4.29)	7.55 (4.13)	7.34 (4.14)
Years since Migration	12.01 (5.09)	11.88 (4.82)	11.40 (3.99)	11.56 (4.06)
Minority Language Concentration	16.02 (8.16)	16.07 (7.99)	16.74 (7.75)	17.06 (7.50)
Resident of Southern State	0.208 (0.406)	0.187 (0.390)	0.198 (0.399)	0.182 (0.386)
Married	0.531 (0.499)	0.569 (0.495)	0.606 (0.489)	0.615 (0.487)
Born in Mexico	0.715 (0.451)	0.718 (0.450)	0.694 (0.461)	0.704 (0.457)
Born in Central/So. America	0.169 (0.375)	0.171 (0.377)	0.206 (0.405)	0.209 (0.407)
English Speaking Proficiency	0.667 (0.471)	0.671 (0.470)	0.557 (0.497)	0.549 (0.498)
English Reading Proficiency	0.601 (0.490)	0.604 (0.489)	0.517 (0.500)	0.507 (0.500)
Sample Size[b]	3,183	1,939	2,425	1,595

Notes

a Sample restricted to aliens from non-English speaking countries; standard deviations in parentheses.

b Weighted data with sample size scaled to actual number of observations.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice.

Language use at home, 1992

Tables 2.8 and 2.9 list, for males and females, respectively, estimates of models using the 1992 survey data of English-speaking skills, where the analyses differ by the definition of the dependent variable for the same set of explanatory variables. In the specifications in columns (i) and (ii), the dependent variable equals unity for those who are able to use English in a specific situation: to converse with a doctor, nurse, or teacher (column (i)), or to be able to read a newspaper written in English (column (ii)). In column (iii), the focus is on a measure of English-speaking proficiency: those who

Table 2.6 Regression estimates of English speaking and reading proficiency, males, Legalized Population Survey, 1989[a]

Variable	*All Respondents*		*Respondents Followed Up*	
	Speaking[b]	*Reading*[b]	*Speaking*[b]	*Reading*[b]
Constant	0.442 (7.66)	0.174 (3.10)	0.431 (5.72)	0.143 (1.93)
Age at Migration	–0.009 (9.53)	–0.007 (7.32)	–0.009 (7.32)	–0.006 (5.18)
Educational Attainment	0.037 (19.92)	0.054 (31.16)	0.036 (15.00)	0.054 (23.95)
Years since Migration (YSM)	0.035 (6.90)	0.027 (5.69)	0.033 (4.99)	0.026 (4.04)
$YSM^2/100$	–0.064 (4.86)	–0.052 (4.25)	–0.059 (3.35)	–0.050 (3.00)
Married	0.086 (5.43)	0.033 (2.09)	0.082 (3.96)	0.034 (1.67)
South	–0.020 (1.03)	0.023 (1.23)	0.009 (0.37)	0.067 (2.73)
Minority Language Concentration	–0.006 (6.99)	–0.002 (2.04)	–0.006 (4.55)	–0.001 (0.71)
$\bar{R}^2$	0.1864	0.2556	0.1637	0.2446
Sample Size	3,183	3,183	1,939	1,939

Notes

a Sample restricted to aliens from non-English speaking countries; figures in parentheses are White’s (1980) heteroscedasticity-consistent *t* statistics.

b Dependent Variables: Speaking = the ability to speak in English with a doctor, nurse or teacher; Reading = the ability to read and understand a newspaper written in English.

c Weighted data with sample size scaled to actual number of observations.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice.

speak only English at home or speak English Very Well or Well, in contrast to those who speak a language other than English at home and speak English either Not Well or Not at All. In column (iv), the measure of English proficiency differentiates individuals who report that (compared to individuals born in the United States) their English-speaking skills did not limit their job opportunities from those who felt that it did limit job opportunities.

The mean value of the dependent variable differs appreciably across the alternative measures of proficiency in English. Thus, while only 38 percent of the male sample stated that they either spoke only English at home or, where other than English was spoken, they spoke English either Very Well or Well, as many as 74 percent of males reported they were able to converse with a doctor, teacher, or nurse—a difference of around 35 percentage points.

Table 2.7 Regression estimates of English speaking and reading proficiency, females, Legalized Population Survey[a], 1989

Variable	*All Respondents*		*Respondents Followed Up*	
	Speaking[b]	*Reading*[b]	*Speaking*[b]	*Reading*[b]
Constant	–0.036 (0.51)	–0.133 (1.89)	–0.046 (0.52)	–0.200 (2.29)
Age at Migration	–0.006 (5.88)	–0.006 (6.28)	–0.006 (5.07)	–0.006 (4.90)
Educational Attainment	0.050 (23.36)	0.054 (25.97)	0.050 (18.85)	0.056 (22.46)
Years since Migration (YSM)	0.063 (8.92)	0.059 (8.22)	0.067 (7.72)	0.061 (7.03)
$YSM^2/100$	–0.145 (6.68)	–0.133 (6.13)	–0.149 (5.74)	–0.132 (5.18)
Married	–0.007 (0.37)	0.003 (0.15)	0.001 (0.01)	0.013 (0.59)
South	0.049 (2.21)	0.129 (5.98)	0.046 (1.65)	0.151 (5.59)
Minority Language Concentration	–0.006 (5.72)	–0.004 (3.34)	–0.007 (4.91)	–0.003 (2.06)
$\bar{R}^2$	0.2568	0.2732	0.2592	0.2853
Sample Size[c]	2,425	2,425	1,595	1,595

Notes

a Sample restricted to aliens from non-English speaking countries; figures in parentheses are White's (1980) heteroscedasticity-consistent *t* statistics.

b Dependent Variables: Speaking = the ability to speak in English with a doctor, nurse or teacher; Reading = the ability to read and understand a newspaper written in English.

c Weighted data with sample size scaled to actual number of observations.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice.

Among females, the mean values of the different measures of English proficiency differ by up to 30 percentage points.[9]

Despite the differences in means, the main feature of the results in Tables 2.8 and 2.9 is the broad consistency of the estimated effects across the various definitions of the dependent variable. For example, among males, the partial effect of age at migration on language proficiency varies between –0.007 and –0.010. It is highly significant in each case.

There is a noteworthy difference only in the case of the minority language concentration variable for the equations estimated for males. It is not statistically significant in the one equation that examines reading skills. As argued in Chiswick and Miller (1998), the detrimental effect on English skills of residence in an area with a relatively large representation of individuals

Table 2.8 Regression estimates of English speaking and reading proficiency, males, Legalized Population Survey, 1992[a]

Variable	*Speaking Ability*[b] *(i)*	*Reading Ability*[b] *(ii)*	*Speaking Proficiency*[b] *(iii)*	*Not Limit Job Opportunities*[b] *(iv)*
Constant	0.452 (5.25)	0.163 (1.80)	0.062 (0.70)	0.148 (1.55)
Age at Migration	–0.010 (8.37)	–0.007 (5.42)	–0.009 (8.53)	–0.006 (4.74)
Educational Attainment	0.033 (14.82)	0.054 (24.28)	0.050 (21.89)	0.024 (9.18)
Years since Migration (YSM)	0.039 (5.62)	0.023 (3.35)	0.039 (5.44)	0.032 (4.19)
$YSM^2/100$	–0.075 (4.83)	–0.036 (2.33)	–0.062 (3.80)	–0.049 (2.95)
Married	0.063 (3.11)	0.005 (0.23)	–0.011 (0.56)	–0.010 (0.41)
South	–0.017 (0.70)	–0.009 (0.37)	0.026 (1.14)	–0.004 (0.14)
Minority Language Concentration	–0.003 (2.66)	0.001 (0.84)	–0.011 (8.91)	–0.006 (4.45)
Mean of Dependent Dependent Variable	0.738	0.633	0.377	0.375
$\bar{R}^2$	0.1775	0.2539	0.3351	0.0895
Sample Size[c]	1,939	1,939	1,939	1,939

Notes

a Sample restricted to aliens from non-English speaking countries; figures in parentheses are White's (1980) heteroscedasticity-consistent *t* statistics.

b Dependent Variables: Speaking Ability = the ability to speak in English with a doctor, nurse or teacher; Reading Ability = the ability to read and understand a newspaper written in English; Speaking Proficiency indicates the individual speaks only English at home or, where a language other than English is spoken, English is spoken either Very Well or Well; Not Limit Job Opportunities indicates that the respondent did not perceive his language skills as limiting job opportunities in the United States.

c Weighted data with sample size scaled to actual number of observations.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice.

speaking the same non-English second language as the respondent comes about because speaking requires the active participation of a second person. Reading does not require this, and hence it is expected that this neighborhood factor would be far less important in the case of reading skills than for speaking skills.

Table 2.9 Regression estimates of English speaking and reading proficiency, females, Legalized Population Survey, 1992[a]

Variable	*Speaking Ability*[b] *(i)*	*Reading Ability*[b] *(ii)*	*Speaking Proficiency*[b] *(iii)*	*Not Limit Job Opportunities*[b] *(iv)*
Constant	0.110	–0.166	–0.027	–0.083
	(1.02)	(1.55)	(0.27)	(0.76)
Age at Migration	–0.009	–0.006	–0.008	–0.002
	(8.24)	(5.36)	(7.80)	(1.71)
Educational Attainment	0.049	0.058	0.045	0.018
	(20.17)	(24.75)	(18.63)	(6.42)
Years Since Migration (YSM)	0.056	0.053	0.046	0.047
	(5.88)	(5.46)	(5.13)	(5.09)
$YSM^2/100$	–0.108	–0.106	–0.079	–0.088
	(4.50)	(4.24)	(3.51)	(4.13)
Married	–0.002	–0.002	–0.025	0.010
	(0.08)	(0.07)	(1.27)	(0.44)
South	–0.001	0.101	0.014	–0.032
	(0.06)	(3.71)	(0.57)	(1.18)
Minority Language Concentration	–0.007	–0.003	–0.012	–0.009
	(5.62)	(2.52)	(8.75)	(5.86)
Mean of Dependent Variable	0.581	0.539	0.285	0.297
$\bar{R}^2$	0.2886	0.3059	0.3399	0.0845
Sample Size[c]	1,595	1,595	1,595	1,595

Notes

a Sample restricted to aliens from non-English speaking countries; figures in parentheses are White's (1980) heteroscedasticity-consistent *t* statistics.

b Dependent Variables: Speaking Ability = the ability to speak in English with a doctor, nurse or teacher; Reading Ability = the ability to read and understand a newspaper written in English; Speaking Proficiency indicates the individual speaks only English at home or, where a language other than English is spoken, English is spoken either Very Well or Well; Not Limit Job Opportunities indicates that the respondent did not perceive his language skills as limiting job opportunities in the United States.

c Weighted data with sample size scaled to actual number of observations.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice.

Comparing language skills in 1989 and 1992

Tables 2.10 and 2.11 present for males and females, respectively, models of English speaking and reading skills for 1989 and 1992. Table 2.10 reveals that the partial effects estimated for males for the two time periods are remarkably similar: the one difference is that the minority language concentration variable has a reduced impact on speaking proficiency in 1992. This may arise because the adverse partial effect of residing in a linguistic enclave diminishes

Table 2.10 Regression estimates of English speaking and reading proficiency, males, Legalized Population Survey, 1989 and 1992[a]

	Speaking[b]		*Reading*[b]	
Variable	*1989*	*1992*	*1989*	*1992*
Constant	0.431 (5.72)	0.452 (5.25)	0.143 (1.93)	0.163 (1.80)
Age at Migration	–0.009 (7.32)	–0.010 (8.37)	–0.006 (5.18)	–0.007 (5.42)
Educational Attainment	0.036 (15.00)	0.033 (14.82)	0.054 (23.95)	0.054 (24.28)
Years since Migration (YSM)	0.033 (4.99)	0.039 (5.62)	0.026 (4.04)	0.023 (3.35)
$YSM^2/100$	–0.059 (3.35)	–0.075 (4.83)	–0.050 (3.00)	–0.036 (2.33)
Married	–0.082 (3.96)	0.063 (3.11)	0.034 (1.67)	0.005 (0.23)
South	0.009 (0.37)	–0.017 (0.70)	0.067 (2.73)	–0.009 (0.37)
Minority Language Concentration	–0.006 (4.55)	–0.003 (2.66)	–0.001 (0.71)	0.001 (0.84)
Mean of Dependent Variable	0.671	0.738	0.604	0.633
$\bar{R}^2$	0.1637	0.1775	0.2446	0.2539
Sample Size[c]	1,939	1,939	1,939	1,939

Notes

a Sample restricted to aliens from non-English speaking countries; figures in parentheses are White's (1980) heteroscedasticity-consistent *t* statistics.

b Dependent Variables: Speaking = the ability to speak in English with a doctor, nurse or teacher; Reading = the ability to read and understand a newspaper written in English.

c Weighted data with sample size scaled to actual number of observations.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice.

with years of residence (see Chiswick and Miller, 1995) and, by definition, the duration of residence of each member of the sample increased by three years.

The estimates of the models of English speaking and reading skills for females (Table 2.11) are also quite stable over the panel.

Tables 2.10 and 2.11 also include information on the mean level of the measure of proficiency in each year. Proficiency in speaking and in reading English increased over the three-year interval, although the percentage point increase was larger for males than for females. Thus, English speaking skills among males improved by 6.7 percentage points over the three-year period, while males' English reading skills improved, on average, by 2.9 percentage

Table 2.11 Regression estimates of English speaking and reading proficiency, females, Legalized Population Survey, 1989 and 1992[a]

Variable	*Speaking[b]*		*Reading[b]*	
	1989	*1992*	*1989*	*1992*
Constant	–0.046 (0.52)	0.110 (1.02)	–0.200 (2.29)	–0.166 (1.55)
Age at Migration	–0.006 (5.07)	–0.009 (8.24)	–0.006 (4.90)	–0.006 (5.36)
Educational Attainment	0.050 (18.85)	0.049 (20.17)	0.056 (22.46)	0.058 (24.75)
Years since Migration (YSM)	0.067 (7.72)	0.056 (5.88)	0.061 (7.03)	0.053 (5.46)
$YSM^2/100$	–0.149 (5.74)	–0.108 (4.50)	–0.132 (5.18)	–0.001 (4.24)
Married	0.001 (0.01)	–0.002 (0.08)	0.013 (0.59)	–0.002 (0.07)
South	0.046 (1.65)	–0.001 (0.05)	0.151 (5.59)	0.101 (3.71)
Minority Language Concentration	–0.007 (4.91)	–0.007 (5.62)	–0.003 (2.06)	–0.003 (2.52)
Mean of Dependent Variable	0.549	0.581	0.507	0.539
$\bar{R}^2$	0.2592	0.2886	0.2853	0.305
Sample Size[c]	1,595	1,595	1,595	1,595

Notes

a Sample restricted to aliens from non-English speaking countries; figures in parentheses are White's (1980) heteroscedasticity-consistent *t* statistics.

b Dependent Variables: Speaking = the ability to speak in English with a doctor, nurse or teacher; Reading = the the ability to read and understand a newspaper written in English.

c Weighted data with sample size scaled to actual number of observations.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice.

points. For females, the comparable improvement in mean speaking skills was 3.2 percentage points and that for reading skills also 3.2 percentage points.

The changes in language skills over the three-year period can be analyzed using a Blinder (1973) decomposition. The Blinder technique, when applied to a variable that has been analyzed using least squares regressions at two points in time, allows the change over time in the mean value of the dependent variable to be apportioned into two components. The first component is attributable to changes over time in the mean values of explanatory variables in the least squares regression ("due to changes in characteristics"). The second component is due to changes in the relationships between the dependent and explanatory variables between the two points of comparison ("due to changes in coefficients"). In the current context, the first component can also be thought of as the part of the change over time that is predicted by the model, while the second component is the "unexplained" change.

The decomposition in Table 2.12 shows that two-thirds of the change in English speaking skills among males and all the change in English reading skills among males would have been predicted by the model. The principal determinant is duration of residence. When the model in Table 2.10 for males is used to predict the change in skills associated with three extra years of residence (increasing both age and duration by three years), an improvement is obtained in speaking skills of 2.5 percentage points and in reading skills of 2.0 percentage points.

Among females the picture is less clear. Both speaking and reading skills actually improved by around 3 percentage points, yet they were predicted to improve by about 7 percentage points. Hence there is an unexplained differential between prediction and actual outcome of 4 percentage points. The reason for the prediction of such strong growth in language proficiency is the powerful effect of duration of residence among females. For females, the partial effect of duration of residence on English speaking skills is 2.2 percentage points per year, evaluated at fifteen years of residence. This point estimate is almost 50 percent greater than for males. The reasons for the gender differential are unclear; however, the differences in effects by gender are not statistically significant.

It is interesting that the duration of residence effects for males and females are much more similar in the estimations based on the census-type English proficiency questions. This question has a more stringent criterion for proficiency than the measures of fluency considered in Tables 2.10 and 2.11 (see the mean values in Tables 2.8 and 2.9).

Further insights into the effect of duration of residence on English proficiency among males and females can be gained by estimating an ordered

Table 2.12 Decomposition of percentage point changes in proficiency in English between 1989 and 1992

Component	*Males*		*Females*	
	Speaking	*Reading*	*Speaking*	*Reading*
Improvement in Proficiency	0.066	0.029	0.032	0.032
Due to Changes in Circumstances	0.042	0.030	0.071	0.072
(i) Increase in age and duration	0.025	0.020	0.060	0.060
(ii) Other variables	0.017	0.010	0.011	0.012
Unexplained	0.024	–0.001	–0.039	–0.040

Source: Estimates in Tables 2.10 and 2.11 and sample means.

probability model of language proficiency which allows the dependent variable to take any of the four categories provided in the data, namely that the respondent speaks English Not at All, Not Well, Well, or Very Well. This is a more flexible approach to modeling that allows language shift to be analyzed more closely (Espenshade and Fu, 1997, used an ordered logit model on their four-category English-speaking variable). The results from this model can be used to predict the probability of being in each of the four language categories for individuals defined by given sets of characteristics.[10] Table 2.13 lists predictions for legalized aliens who have the mean levels of characteristics (education, age, etc.) other than duration of residence. The duration of residence of the particular set of individuals is given in the left-hand column.

It is apparent that while females have a consistently lower level of proficiency than males in this population, the broad patterns of the positive effects of duration of residence on the language skills exist for both males and females. The improvements in language skills with duration of residence occur across the board.

A final issue concerning changes in language skills between 1989 and 1992 is the considerable number of respondents who reported a deterioration in their language proficiency over the three-year period (see Table 2.4). Analysis of this skill deterioration using the language attainment model shows that

Table 2.13 Predicted membership of English speaking proficiency categories by duration of residence

Duration of Residence		*English Speaking Proficiency*		
	Not at All	*Not Well*	*Well*	*Very Well*
Males				
0	0.47	0.47	0.05	0.00
5	0.28	0.58	0.13	0.01
10	0.16	0.59	0.22	0.03
15	0.09	0.54	0.31	0.06
20	0.05	0.47	0.37	0.10
25	0.04	0.42	0.40	0.14
30	0.03	0.40	0.42	0.15
Females				
0	0.70	0.29	0.01	0.00
5	0.51	0.45	0.04	0.00
10	0.34	0.56	0.10	0.01
15	0.22	0.60	0.16	0.02
20	0.15	0.59	0.23	0.03
25	0.10	0.56	0.29	0.05
30	0.08	0.53	0.32	0.07

Source: Authors' calculations from estimates of an ordered probit model with the census-style measure of English proficiency (see Table 2.1) as the dependent variable and the regressors from Table 2.10 as explanatory variables.

loss of speaking skills occurs among older, less well-educated males and females, while loss of reading proficiency is concentrated among the less well-educated. Thus, it is the more fluent members of groups with lower than average language skills in 1989 that were more likely to show a deterioration in these skills. This suggests that reporting error in 1989 and a regression to the mean may be partially responsible for the "loss" of language skills. An alternative explanation may be that with legal status came greater contact with the general society as distinct from the ethnic enclave, and with it a downward reassessment of self-evaluated language skills. Finally, one cannot rule out a true atrophy of skills among some respondents. Atrophy would be expected to be greater among those living and working in a linguistic enclave, which would be more likely among those with lower levels of education.

Language use at work

As noted above, the LPS follow-up survey contains information on legalized aliens' ability to read and speak English in the workplace. Table 2.14 presents models of English proficiency based on the workplace English information. These estimates show that, for both males and females, English proficiency at work is negatively associated with age at migration and is positively associated with educational attainment and years since migration. However, neither workplace reading proficiency nor workplace speaking proficiency is significantly affected by the minority language concentration variable. That is, the specialized vocabulary of work is not affected by the residential neighborhood language composition. This result has intuitive appeal and adds to the confidence one can place in the model of language proficiency.

Summary and conclusion

Language attainment is an important aspect of the immigrant adjustment process in all Western countries. It has been modeled extensively in recent years (see Beenstock, 1994; Chiswick, 1991; Chiswick and Miller 1992, 1995, 1998, 1999; Dustmann, 1994). The typical approach in these studies is to create a single index of destination language proficiency and to study the distribution of this variable using probability models. It has been shown for the United States that English language skills are greater among those who have more schooling, migrate while young, married after migration, and reside outside areas of concentration of immigrants with their mother tongue. English language skills have also been shown to vary directly with the physical distance of the country of origin from the United States and inversely with the expected propensity for return migration and with the linguistic distance between English and the immigrant's mother tongue. This process of language attainment appears to be remarkably robust across countries and for different time periods (see, in particular, Chiswick and Miller, 1995).

Table 2.14 Regression estimates of English speaking and reading proficiency at work, Legalized Population Survey, 1992[a]

Component	*Males*		*Females*	
	Speaking[b]	*Reading*[b]	*Speaking*[b]	*Reading*[b]
Constant	0.662 (8.64)	0.631 (8.14)	0.136 (0.75)	0.295 (1.55)
Age at Migration	–0.004 (3.38)	–0.005 (3.63)	–0.008 (5.13)	–0.007 (4.98)
Educational Attainment	0.017 (8.73)	0.022 (10.50)	0.033 (10.41)	0.037 (11.34)
Years since Migration (YSM)	0.023 (3.41)	0.020 (2.88)	0.076 (3.68)	0.048 (2.21)
$YSM^2/100$	–0.051 (2.88)	–0.047 (2.64)	–0.187 (3.07)	–0.103 (1.63)
Married	0.029 (1.64)	0.063 (3.26)	0.015 (0.61)	0.014 (0.56)
South	–0.020 (1.02)	–0.023 (1.09)	–0.013 (0.46)	0.001 (0.02)
Minority Language Concentration	–0.001 (0.09)	–0.001 (0.86)	–0.001 (0.23)	0.001 (0.10)
Mean of Dependent Variable	0.901	0.878	0.819	0.806
$\bar{R}^2$	0.0853	0.1193	0.1941	0.2035
Sample Size[c]	1,492	1,492	839	839

Notes

a Sample restricted to aliens from non-English speaking countries; figures in parentheses are White's (1980) heteroscedasticity-consistent t statistics.

b Dependent Variables: Speaking = the ability to communicate with job supervisor/transact business in English; Reading = the ability to read and understand work-related manuals and written instructions in English.

c Weighted data with sample size scaled to actual number of observations.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice.

This chapter investigates whether the findings from the study of language proficiency among immigrants are robust with respect to variations in the definition of proficiency. Using data from a pair of surveys of "legalized aliens," that is, illegal aliens granted amnesty under the 1986 IRCA, measures of language proficiency based on the alien's ability to speak or read English in specific situations, measures constructed from the alien's self-assessment of their English speaking skills, measures constructed from the alien's perceptions of whether their language skills have limited their job opportunities in the United States, and measures of speaking and reading proficiency at work are used in a model of language attainment. The results from previous research carry over remarkably well to all alternative definitions. Some differences in findings are, however, reported. In particular, it appears that residence among others who speak the same non-English second language as the alien adversely affects English-speaking skills, but it does not affect reading skills among males or English-language speaking or reading proficiency in the workplace.

Estimates of models of language proficiency for the same individuals at two time periods, 1989 and 1992, are also derived in this study. Viewed as two cross-sections, the results are stable over time. Moreover, on average, language proficiency increased over the period in a manner consistent with the estimated model. The study of the change in language skills between 1989 and 1992, however, reveals improvements in skills among many but the atrophy of skills among some aliens. The individuals whose English skills deteriorated tend to be those who in 1989 were the more fluent members among the relatively old and less well-educated. It is unclear whether the deterioration in English language skills is a true atrophy effect (perhaps) from lack of use of English language skills due to living in a linguistic enclave, or whether it is due to reporting errors and a regression to the mean. These findings suggest that English language skill retention needs to be considered alongside skill development in future research.

Notes

1 The research for this chapter was funded, in part, by the Bureau of International Labor Affairs (ILAB), U.S. Department of Labor. Comments received from Shirley Smith of ILAB were most helpful.

2 The initial survey in 1989 (LPS1) was sponsored by the Immigration and Naturalization Service, while the follow up in 1992 (LPS2) was sponsored by the Bureau of International Labor Affairs, U.S. Department of Labor. The fieldwork for LPS1 was conducted between February and June 1989 and for LPS2 between April and August 1992. LPS1 sampled aliens who applied for temporary legal status (during the one year starting May 5, 1987) for those who had been in a continuous illegal status since January 1, 1982. LPS2 was a subsample of the individuals in LPS1 who had received permanent resident alien status. Those in temporary legal status could apply for permanent status during a one-year period that started 19 months after they were granted temporary status. To receive permanent resident alien status, they had to demonstrate that they had "a minimal understanding of ordinary English and a knowledge of the history and government of the United States, or were in the process of securing the training needed to acquire that knowledge" (Immigration and Naturalization Service, 1992:4). This requirement could be satisfied by passing a simple test or, as was generally done, attending 40 hours of the 60 hour course approved by INS that was offered by community groups and other organizations (Department of Health and Human Service, 1991:3–7). In effect, the English language requirement was trivial.

Although by matching the respondents in LPS1 and LPS2 a panel data set is created, the data are not ideal for treatment as a longitudinal data set because of changes in the format of certain key questions. For example, post-high school education was reported by years of schooling attended in the 1989 survey and by "qualifications" measured by degrees received (e.g., "Bachelor's Degree," "Master's Degree") in the 1992 survey.

3 All respondents in the 1989 survey were asked these six questions but the 1989 survey did not ask the four-category English language proficiency question.

4 Recent research in Canada has shown that proxy indicators such as those used in this study match up reasonably well with information from direct literacy assessment surveys. This research, however, suggests a role for combinations of proxy indicators (see Neice and Adsett, 1994).

5 Since marriage to a U.S. citizen would result in the nearly automatic award of legal resident alien status, pre-1987 spouses in the LPS were presumably not U.S. citizens.

6 Data available in the LPS preclude identifying the birthplace of the spouse.

7 These equations are estimated using ordinary least squares. When the dependent variable is a dichotomous variable, the results of least squares regression are to be interpreted as a linear probability model. The problem of heteroscedastic residuals that usually arises when there is a dichotomous dependent variable is minimized using White's (1980) heteroscedasticity correction technique.

8 There is, however, one noteworthy difference in the results for reading among males—the minority language concentration variable, which is at the margin of significance in the estimations derived for the full sample, is insignificant in the equation estimated on the subsample of individuals in the second panel. It is possible that this is a form of selection bias. An additional and possibly related difference is observed in other specifications. When equations that include a variable recording whether the alien had entered the United States more than once prior to the application for temporary legal status are estimated, the coefficient on this variable is negative and statistically significant for both males and females when all respondents are considered. But when only those respondents represented in the follow-up survey are considered, the variable is negative and statistically significant for females but insignificant for males. Multiple entries suggest less total time in the United States when years since migration is held constant. It also suggests a less permanent attachment to the United States and hence a smaller investment in human capital specific to the United States. English language skills may also atrophy during periods of absence from the United States.

9 It will be apparent from the data in Table 2.1 that it is not possible to alter the definition of "proficient" under the census-style language skills question to generate a mean level of fluency similar to the mean response to the question on English speaking ability in a specific situation.

10 The ordered probability model analysis is available upon request. The exogenous variables are the same as those used in the linear probability models reported in Table 2.10.

References

Beenstock, M. 1994 "Learning Hebrew and Finding a Job: Econometric Analysis of Immigrant Absorption in Israel." Paper presented at the Conference on the Economics of International Migration: Econometric Evidence. Konstanz, Germany. February.

Blinder, A. 1973 "Wage Discrimination: Reduced Form and Structural Estimation," *Journal of Human Resources*, 8(4):436–455.

Charette, M. and R. Meng 1994 "Explaining Language Proficiency: Objective versus Self-Assessed Measures of Literacy," *Economics Letters*, 44:313–321.

Chiswick, B. R. 1991 "Speaking, Reading and Earnings among Low-Skilled Immigrants," *Journal of Labor Economics*, 9(2):149–170.

Chiswick, B. R. and P. W. Miller 1999 "Language Skills and Earnings among Legalized Aliens," *Journal of Population Economics*, 12.

—— 1998 "English Language Fluency among Immigrants in the United States," *Research in Labor Economics*, 17.

—— 1995 "The Endogeneity Between Language and Earnings: International Analyses," *Journal of Labor Economics*, 13(2):246–288.

—— 1992 "Language in the Labor Market: The Immigrant Experience in Canada and the United States." In *Immigration, Language and Ethnic Issues: Canada and the United States*. Ed. B. R. Chiswick. Washington, DC: American Enterprise Institute. Pp. 229–296.

Chiswick, B. R. and T. A. Sullivan 1995 "The New Immigrants." In *State of the Union: America in the 1990s*, Vol. 2: *Social Trends*. Ed. R. Farley. New York: Russell Sage Foundation. Pp. 211–270.

Department of Health and Human Services, U.S. 1991 "State Legalization Impact Assistance Grant Program," *Administration for Children and Families, FY 1989 Report to Congress*. Washington, DC: Department of Health and Human Services. December.

Dustmann, C. 1994 "Speaking Fluency, Writing Fluency and Earnings of Migrants," *Journal of Population Economics*, 7(2):226–236.

Espenshade, T. J. and H. Fu 1997 "An Analysis of English Language Proficiency among U.S. Immigrants," *American Sociological Review*, 62:288–305. April.

Immigration and Naturalization Service, U.S. 1992 *Immigration Reform and Control Act: Report on the Legalized Alien Population* Washington, DC: U.S. Department of Justice. March.

—— 1991 *Provisional Legalization Application Statistics*. Washington, DC: Statistics Division, Office of Strategic Planning, U.S. Department of Justice. December 1.

—— n.d. *Documentation for the Legalized Population Survey Public Use Tape*. Washington, DC: Statistics Division, Office of Strategic Planning, United States Department of Justice.

Khan, A. H. 1997 "Post-Migration Investments in Schooling among Immigrants in the United States," *Quarterly Review of Economics and Finance*, 37:285–313.

Long, M. H. 1990 "Maturational Constraints on Language Development," *Studies in Second Language Acquisition*, 12(3):251–285.

Neice, D. and M. Adsett 1994 "Direct Versus Proxy Measures of Adult Functional Literacy: A Preliminary Re-Examination." Ottawa, Canada: The Department of the Secretary of State. Xerox.

Westat 1992 *Legalized Population Follow-Up Survey*. Rockville, Maryland.

White, H. 1980 "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity," *Econometrica*, 48:817–838.

3 Language choice among immigrants in a multi-lingual destination

I. Introduction

The development of language skills among immigrants is important for their economic adjustment. Language is an important dimension of the skill levels of immigrants relevant for the labor market, and hence influences both their economic attainment and their impact on the economy. Moreover, language plays a key role in the social adjustment of immigrants and in the social and political cohesion both within and among groups. As a result, the linguistic skills of immigrants also has important political implications.[1]

For immigrants to mono-lingual destinations, decisions regarding language acquisition are straightforward. If they lack fluency in the dominant language of the destination they need to decide, implicitly or explicitly, the extent to which they will invest in dominant language skills. Other immigrant receiving countries, however, are multi-lingual. In some countries two (or more) segments of the population speak different languages (as in Canada, Belgium and Switzerland). In others bilingualism in the destination language and an "international" language are required for certain occupations. For example, in the Netherlands or Israel, knowledge of only Dutch or Hebrew would limit educational or employment opportunities in many occupations.

During the past decade two separate, important literatures in the economics of language acquisition have developed. One analyzes the determinants of the acquisitions of dominant language skills among immigrants. This research has been conducted for several countries, including the United States, Canada and Australia. The emphasis has been on the demographic, environmental and economic variables that influence the extent to which immigrants become fluent and/or literate in the dominant language.[2] The second literature focuses on the native-born population in bi-lingual or multi-lingual societies and is concerned with the determinants of bilingualism in the population. The research is necessarily of particular interest for countries with two or more dominant or official languages.[3] Among individuals in dual-language countries the incentives for bilingualism are also determined by demographic, environmental and economic factors. In these situations political factors, such as nationalism, ethnic identity and concerns for both minority rights and for social cohesion, among other issues, can help shape the environmental and economic incentives.[4] What is missing from the literature, however, is the intersection of these two issues—the determinants of dominant language choice among immigrants in multi-lingual countries.

This chapter develops a model in Section II to study language choice decisions among immigrants in a multi-lingual destination. Section III tests the model using data for Canada, where English and French are the two dominant and official destination languages. The final section is a summary and conclusion that develops broader implications of the analysis for understanding the economics of language and immigrant adjustment.

II. The model

Previous research on dominant language acquisition among immigrants provides the framework for this analysis. It has been shown for several major immigrant receiving countries (see Chiswick and Miller, 1991 [1995], 1992) that destination language skills are systematically related to exposure, efficiency and economic variables. Language skills increase with exposure, that is, with the use of the destination language in the country of origin, duration in the destination, whether the spouse speaks the destination language and the extent to which the origin language is spoken in the area where the immigrant lives, among other variables. Language skills also increase with the efficiency of language acquisition, where efficiency appears to be greater the younger the age at migration and the higher the level of education. Moreover, those who anticipate greater economic returns from dominant language fluency (as measured by earnings) are more likely to become fluent. Those with a lower probability of return migration could expect a greater economic return from destination language fluency.

These same sets of structural variables, that is, exposure, efficiency and economic factors, enter a model in which there is more than one relevant destination language. For example, suppose there are two distinct linguistic communities, E and F. Immigrants not fluent in either E or F will have an incentive to acquire fluency in one or the other language, or perhaps both languages.[5] Language exposure prior to arrival may not be neutral between E and F. One of the languages, for example, may have been a common second language in the origin (e.g., English in India) or for exogenous reasons the immigrant may have settled in a region in which either E or F predominates. Linguistic distance may also be relevant. If the immigrant's mother tongue is from the same language family as, say F, but is linguistically further from E, the cost of acquiring fluency would be less for F than for E.

Economic incentives also become relevant. One of the languages may predominate as the language in which economic activity takes place. Alternatively, let us postulate that in all branches of the local economy E and F are equally useful, but that E is more of an "international" language than is F; that is, there is a broader and deeper market outside the local economy for transactions and/or information in E than in F. Then an immigrant who may otherwise be indifferent between investing in either E or F would have a greater incentive to invest in E.

These incentives may vary by level of education. If those with less schooling confine their activities to local labor and product markets, but those with more schooling operate in broader labor or product markets, the economic returns from acquiring the more "internationally" useful language would rise with the level of schooling. As a result, fluency in the international language, E, rather than in the second language, F, (or in areas where F predominates as a local language, bilingualism in E and F) would increase with the level of schooling.

Education may also be associated with an efficiency effect. Those with higher levels of schooling may be more efficient in acquiring second- and third-language fluency because of greater innate ability that resulted in the higher level of schooling, because of greater knowledge of the structure of language from their more advanced schooling in their mother tongue, or because of exposure to foreign languages in higher levels of schooling. Thus, it is anticipated that fluency in the international dominant language would rise with the level of schooling.

Canada provides a unique opportunity for testing the implications of the model. It is a dual-language country in which English and French are the two official languages. French is the majority and predominant language in one province, Quebec, while English is the majority and dominant language elsewhere in Canada. The Quebec economy is smaller that the English-speaking economy. The population of Quebec is about one-quarter of Canada, but only about 3% of the combined Canada/United States population.[6] To the extent that the scope of the labor market and other economic activity rises with one's level of education, the more highly educated immigrants would have a greater incentive to learn English than French, or if French speakers already, they would have a greater incentive to become English-French bilingual speakers, particularly if they live in Quebec.

Although the location choice of immigrants is not of central interest in this chapter, immigrants are distributed across provinces in Canada in a manner influenced by linguistic considerations.

As one would expect, immigrants with a lower cost of learning French are more likely to settle in Quebec. According to the 1981 Census of Canada, of the adult male immigrants in Quebec 11% are from France and 26% are from Italy and Portugal, Romance language countries, while only 15% are from the United States, United Kingdom and Ireland. In contrast, in "English Canada" less than 1% are from France, 16% from Italy and Portugal, and 33% are from the United States, United Kingdom and Ireland. This tendency is now influenced by immigration law. The 1978 Immigration Act grants provinces the right to participate in immigration decisions and Quebec awards more points for fluency in French under its selection criteria than in the federal government's immigration program. However, only the most recent of immigrants in the 1981 Canadian Census could have been influenced by this policy.

III. Empirical analysis

This section tests the hypotheses regarding the determinants of language choice among immigrants in a multi-lingual destination. The data are from the 1981 Census of Canada and the empirical technique employed is multinominal logit analysis as there are several discrete language categories.

A. The data

The empirical analysis is based on the one-in-fifty individual file from the 1981 Census of Canada. The analysis is limited to adult (age 25 to 64 years) foreign-born men. The variables used in the analysis are described and their means and standard deviations are reported in Appendix 3A. The dependent variable is dominant language fluency which may take one of four mutually exclusive fluency values: English only (ENGL), French only (FREN), both English and French (BOTH) and neither (NONE). Fluency is determined by the self-reported response to the question on whether the respondent could carry on a conversation in the official languages. There are no data on the degree of fluency or on literacy in the official languages. Although it might be desirable to have more "objective" measures of language fluency rather than self-assessment, this has become standard practice in the censuses and surveys in the major immigrant receiving countries. There is no particular reason for believing that self-assessment reporting errors vary systematically with the variables included in the analysis. Furthermore, it is not obvious that a test-based measure of language fluency (even if it existed) would be superior to a self-assessment question in a survey as a measure of the language skills relevant in the labor market.

English is the only official language spoken by over 90% of the immigrants in each of the provinces, except Quebec (14%) and New Brunswick (80%). Including English-French bilinguals among the English speakers raises the proportion to 93% or more in each of the provinces except Quebec (44%). In Quebec, however, 54% of the immigrants speak French only (in contrast to much less than 1% in the other provinces) and this increases to 85% for Quebec if English-French bilinguals are included. Even including English-French bilinguals, less than 5% of the immigrants speak French in the other provinces, except for the Atlantic provinces.[7]

When the analysis is computed for all of Canada, that is, Quebec and English Canada combined, the Quebec dichotomous variable is very highly significant—it dominates the equation—but creates problems in the interpretation of some of the other variables.

Immigrants in Quebec are significantly less likely to be monolingual English speakers compared to each of the other three categories. The empirical analysis is therefore performed separately for Quebec and English Canada.

In principle, it would be desirable to make the choice of location in Canada endogenous, but there are problems with this approach. First, in the conventional sample selection problem the primary equation of interest has a continuous dependent variable, and thus may be estimated by least squares. In the current study, however, the dependent variable in the language equation has four distinct categories and a discrete choice model is appropriate. The analysis of sample selectivity must accommodate this characteristic of the dependent variable. Second, while identification may be achieved in sample selection correction models through either variable exclusions or functional form, the former tends to result in arbitrary restrictions and the latter is associated with imprecise two-step estimates (Nelson, 1984). Limited experimentation with both approaches to achieving identification has merit in cases like the present where the data set is not rich in identifying instruments.

Against this background, two procedures were followed. First, the multinomial logit model (see below) was re-estimated using the generalized residuals test suggested by Vella (1992, 1993).[8] Second, the dependent variable was re-defined to reflect an underlying ordering and an ordered probit model estimated that accounted for sample selectivity (Greene, 1992).[9] In each case a simple locational choice model was estimated where location (Quebec/non-Quebec) was postulated to be a function of birthplace. However, when birthplace variables were included in the second stage equation, the sample selection correction term was so highly correlated with the exogenous regressors in the primary (language choice) equation that the minimum of the loglikelihood function could not be found. When identification was achieved by excluding the birthplace variables from the language choice equation, the selection correction factors were highly significant. But this exclusion restriction is untenable in a language equation where place of birth is of obvious importance to language attainment. Accordingly, given the poor performance and deficiencies of the alternative estimators, we present only estimates that do not take account of sample selection.[10]

The independent variables include duration in Canada, marital status, location, birthplace and a measure of minority group language concentration as measures of exposure to the destination languages, and age and education as efficiency and labor market variables. Country of origin also serves as a proxy for the probability of return migration. The discussion of the links between these variables and the theoretical model of language acquisition can be found in Chiswick and Miller (1991 [1995], 1992).

Appendix 3A reports the definitions of these variables and Table 3.3 reports their means and standard deviations. A striking feature of the data is the virtual absence of any difference in the mean values of many of the explanatory variables between immigrants in Quebec and English Canada. In particular, there is virtually no difference in the average level of schooling (around 11.7 years), average age (around 21.6 years), percent married (about 83%), percent married overseas (about 27%), and duration in Canada (18.3 years in Quebec, 19.7 years elsewhere), although immigrants in Quebec are more likely to live in a metropolitan area (92% compared to 72%). However, as noted above, the distribution of immigrants by country of origin difers sharply between the two major regions. Immigrants from France and other Romance language countries are far more likely to settle in Quebec.

The minority language concentration measure is hypothesized to capture the intensity of exposure per unit of time in an environment in which more people converse in the immigrant's mother tongue. In the study of language choice in multi-lingual Canada, however, a methodological issue arises as to whether French should be treated as a minority language in English Canada, and, likewise, English treated as a minority language in Quebec. The analyses reported below treat only the non-official languages as minority languages. Estimating the equation without the minority language concentration variable, or defining it where French is treated as a minority language in English Canada and English as a minority language in Quebec, has no material impact on the estimated impacts of other variables.

B. The analysis

Table 3.1 reports the multinominal logit equations for language choice by adult foreign-born men separately for Quebec and English Canada. The French-only speakers (FREN), English-French bilingual speakers (BOTH), and those who cannot converse in either language (NONE) are compared to the benchmark English-only speakers (ENGL). Table 3.2 uses the coefficients in the multinominal logit equations to develop predicted probabilities for being in each language category. The predicted values for each variable are evaluated at the mean value of all of the other variables, except for duration of residence which is evaluated for recent arrivals (duration of 1 year).

The analysis indicates that an older age at migration is associated with less fluency in the dominant languages. This appears as a greater proportion reporting fluency in neither language and in Quebec a smaller proportion bilingual. Immigrants arriving in Canada at a later age have an increased probability of speaking only French rather than only English.

On the other hand, a longer duration in Canada is associated with increasing fluency in the dominant languages. While English language skills are widespread among immigrants in English Canada, they do increase with duration. In Quebec, English fluency rises sharply with duration largely as a consequence of a transformation of French-only speakers into English-French bilinguals. Both French-only and English-only speakers in Quebec decline in number with duration, where the decline is greater among the former.

Education also plays an important role. In English Canada higher levels of schooling are primarily associated with a rise in bilingualism, with a decline

Table 3.1 Multinomial logit estimates of language fluency among adult foreign-born men, English Canada and Quebec, 1981

	English Canada			*Quebec*		
	$In\left[\frac{FREN}{ENGL}\right]$	$In\left[\frac{BOTH}{ENGL}\right]$	$In\left[\frac{NONE}{ENGL}\right]$	$In\left[\frac{FREN}{ENGL}\right]$	$In\left[\frac{BOTH}{ENGL}\right]$	$In\left[\frac{NONE}{ENGL}\right]$
Constant	−7.477 (4.43)	−5.714 (26.16)	−6.190 (5.76)	1.570 (2.90)	0.224 (0.56)	−23.828 (0.14)
Education	−0.033 (0.55)	0.236 (27.37)	−0.278 (14.78)	−0.116 (5.71)	0.118 (8.38)	−0.214 (4.58)
Age	0.063 (2.97)	0.006 (1.77)	0.072 (11.59)	0.020 (2.24)	−0.039 (6.25)	0.082 (4.60)
Years since migration (YSM)	−0.289 (3.19)	0.013 (1.16)	−0.239 (9.68)	−0.174 (6.40)	0.025 (1.28)	−0.270 (4.30)
YSM squared/100	0.304 (1.02)	−0.037 (1.41)	0.192 (2.72)	0.364 (5.74)	0.095 (2.21)	0.356 (2.07)
Married	−1.360 (2.73)	−0.417 (6.08)	0.110 (0.61)	0.249 (1.33)	−0.055 (0.43)	0.449 (0.90)
Married overseas	0.081 (0.16)	−0.146 (1.96)	0.143 (1.20)	−0.298 (1.74)	−0.122 (0.94)	0.138 (0.46)
CMA	0.344 (0.63)	0.167 (2.50)	0.043 (0.26)	−2.641 (9.47)	−1.383 (5.68)	9.635 (0.07)
Province						
NFL, Nova Scotia, PEI	−9.301 (0.05)	0.322 (1.92)	−10.614 (0.07)	(a)	(a)	(a)
New Brunswick	−8.616 (0.03)	1.093 (4.23)	0.660 (0.84)	(a)	(a)	(a)
Prairie Provinces	0.516 (1.02)	−0.230 (2.80)	−0.255 (1.41)	(a)	(a)	(a)
British Columbia	−0.196 (0.33)	−0.169 (2.18)	−0.422 (2.40)	(a)	(a)	(a)

Minority language	–0.096	–0.175	0.236	–0.007	–0.066	0.396
concentration	(0.63)	(5.19)	(8.06)	(0.07)	(0.77)	(2.52)
Birthplace						
Western Europe-France	6.869	3.208	5.106	6.772	4.618	4.492
	(6.05)	(19.10)	(3.57)	(10.68)	(7.80)	(0.03)
Other Western Europe	–6.793	0.035	2.477	–0.113	0.466	1.068
	(0.09)	(0.37)	(2.01)	(0.22)	(2.34)	(0.01)
Eastern-Europe	–7.351	–0.053	4.585	–0.383	0.540	11.440
	(0.08)	(0.47)	(4.44)	(0.79)	(3.03)	(0.10)
Southern Europe-Romance	4.281	1.136	5.452	4.375	3.408	13.773
lang countries	(3.80)	(11.12)	(5.38)	(11.68)	(11.92)	(0.12)
Other Southern Europe	2.222	–0.365	4.662	–1.644	0.527	12.886
	(1.55)	(2.08)	(4.54)	(2.15)	(2.72)	(0.12)
Chinese Asia	2.407	–1.500	5.634	0.408	–0.281	13.304
	(2.02)	(6.03)	(5.56)	(0.70)	(0.89)	(0.12)
Other Asia	1.525	–0.467	3.849	0.960	0.934	–0.175
	(1.31)	(4.11)	(3.76)	(2.82)	(5.47)	(0.01)
South and Central America	–8.551	0.615	5.448	5.311	3.623	13.913
	(0.04)	(2.96)	(5.25)	(10.73)	(8.24)	(0.13)
Africa	1.880	0.925	2.490	3.401	2.816	1.596
	(1.32)	(6.93)	(2.00)	(9.29)	(11.17)	(0.01)
Other	2.292	0.612	4.249	2.969	2.051	12.406
	(1.61)	(4.49)	(3.94)	(6.92)	(6.52)	(0.11)
Sample size	20352	20352	20352	3389	3389	3389
Chi Square		4051.2			2211.4	

Note: Asymptotic *t* statistics in parentheses. (a) = variable not entered, ENG = English, NENF = Neither English nor French.
Benchmark for region of residence is Ontario, for country of birth it is the English-speaking countries (US, UK, Ireland and British West Indies).

Source: 1981 Census of Canada, Public Use Sample, Individual File, 1/50 Sample of the Foreign born.

Table 3.2 Predicted distributions across language categories by various characteristics for Quebec and English Canada

Characteristic	*English Canada*				*Quebec*			
	English only	*French only*	*Both*	*Neither language*	*English only*	*French only*	*Both*	*Neither language*
Educational Attainment (years)								
11	91.40	0.10	4.35	3.15	31.26	41.75	26.99	0.00
12	92.04	0.10	5.49	2.38	31.64	37.62	30.73	0.00
13	91.23	0.09	6.89	1.79	31.73	33.60	34.67	0.00
14	89.97	0.09	8.61	1.33	31.53	29.72	38.74	0.00
15	88.23	0.08	10.69	0.99	31.04	26.05	42.90	0.00
16	85.99	0.08	13.20	0.73	30.29	22.63	47.08	0.00
Age (years)								
25	94.50	0.03	4.72	0.75	26.74	23.23	50.03	0.00
35	93.47	0.06	4.95	1.52	30.09	31.88	38.03	0.00
45	91.68	0.11	5.15	3.06	31.79	41.06	27.14	0.00
55	88.46	0.20	5.27	6.07	31.73	49.97	18.30	0.01
65	82.74	0.35	5.23	11.67	30.21	58.01	11.77	0.01
Length of residence in Canada (years)								
1	92.20	0.10	5.11	2.60	31.55	38.90	29.54	0.00
5	93.49	0.03	5.41	1.06	36.66	24.58	38.76	0.00
10	93.97	0.01	5.65	0.37	37.69	13.89	48.42	0.00
15	94.09	0.00	5.76	0.14	34.72	8.44	56.84	0.00
20	94.17	0.00	5.77	0.06	29.58	5.69	64.73	0.00
25	94.29	0.00	5.68	0.03	23.52	4.29	72.19	0.00
Minority language concentration (%)								
1	92.29	0.09	4.72	2.90	31.87	39.17	28.95	0.00
2	92.29	0.08	3.96	3.67	32.56	39.76	27.68	0.00
3	91.98	0.08	3.32	4.63	33.23	40.31	26.45	0.01
4	91.35	0.07	2.77	5.82	33.89	40.85	25.25	0.01
5	90.35	0.06	2.30	7.29	34.54	41.36	24.09	0.02
Birthplace								
English	95.38	0.14	4.34	0.13	59.24	22.77	17.99	0.00
France	26.61	37.32	29.98	6.09	0.27	91.35	8.38	0.00
Other W Europe	94.01	0.00	4.43	1.55	54.73	18.79	26.48	0.00
E Europe	84.80	0.00	3.66	11.53	55.43	14.52	28.88	1.17
Romance lang.	63.67	6.71	9.03	20.59	2.44	74.62	22.41	0.53
Other South Europe	83.90	1.13	2.65	12.32	59.58	4.43	30.65	5.35
Chinese Asia	70.70	1.14	0.72	27.44	51.44	29.75	11.79	7.01
Other Asia	90.88	0.61	2.60	5.92	36.02	36.15	27.83	0.00
S & C America	71.11	0.00	5.99	22.90	1.10	86.05	12.57	2.80
Africa	87.63	0.84	10.07	1.47	5.68	65.50	28.83	0.00
Other	83.64	1.21	7.03	8.13	9.17	68.66	21.66	0.51

Note: Probabilities predicted for immigrants in their first year of residence in Canada and the mean values of all other characteristics.

Source: Estimates presented in Table 3.1.

Table 3.3 Means and standard deviations of variables

	Total Canada		*English Canada*		*Quebec*	
	Mean	*S.D.*	*Mean*	*S.D.*	*Mean*	*S.D.*
Education	11.689	3.851	11.707	3.776	11.581	4.273
	42.645	10.542	42.741	10.549	42.070	10.483
Years since migration	19.452	10.626	19.651	10.696	18.261	10.118
Married	0.827	0.378	0.830	0.376	0.807	0.395
Married overseas	0.272	0.445	0.276	0.447	0.250	0.433
Metropolitan area	0.744	0.436	0.715	0.451	0.919	0.273
Province						
Ontario	0.538	0.499	0.628	0.483		
NFL, Nova Scotia, PEI	0.016	0.126	0.019	0.135	(a)	
New Brunswick	0.005	0.073	0.006	0.079	(a)	
Quebec	0.143	0.350	(a)		1.000	0.000
Prairie provinces	0.139	0.346	0.162	0.369	(a)	
British Columbia	0.159	0.365	0.185	0.388	(a)	
Country of birth						
English-speaking	0.305	0.460	0.331	0.471	0.150	0.357
France	0.024	0.154	0.010	0.010	0.111	0.314
Other Western Europe	0.129	0.335	0.142	0.349	0.050	0.218
Eastern Europe	0.094	0.292	0.098	0.297	0.071	0.257
Romance countries	0.171	0.376	0.157	0.364	0.255	0.436
Other Southern Europe	0.061	0.240	0.058	0.234	0.081	0.274
Chinese-Asia	0.048	0.213	0.052	0.223	0.020	0.141
Other Asia	0.086	0.280	0.084	0.278	0.095	0.294
South & Central America	0.020	0.141	0.014	0.116	0.060	0.237
Africa	0.030	0.170	0.022	0.146	0.077	0.266
Other	0.032	0.175	0.032	0.176	0.030	0.170
Minority language	0.540	1.395	0.523	1.413	0.643	1.279
Sample size	23741		20352		3382	

Source: 1981 Census of Canada, Public Use Sample, Individual File, 1/50 Sample of the Foreign Born.

in the English only and neither language groups. The education effect is particularly dramatic in Quebec with bilingualism increasing and French monolingualism decreasing with higher levels of schooling. That is, fluency in English rises in both regions with schooling level.

Immigrants from France, and to a lesser extent from Italy and Portugal, are much more likely to speak French but not English if they live in Quebec. In English Canada nearly all of the immigrants from other parts of Europe, Asia and Central and South America speak English when they learn one of the official languages. Among immigrants from the same countries living in Quebec, however, many speak only French (particularly if they are from Latin America) or are English-French bilingual speakers.

There are separate provincial effects in English Canada. Residence in the

Eastern provinces rather than Ontario is associated with a higher level of bilingualism, relative to speaking only English. In the provinces west of Ontario, however, the proportion speaking only English is significantly larger, at the expense of the bilingual and neither language categories.

Metropolitan area residence matters in Quebec, where speaking only English is more common in the large metropolitan areas. Elsewhere in Canada living in a metropolitan area is associated with a greater incidence of English-French bilingualism.

Marital status has no effect on language choice among immigrants in Quebec. In English Canada, on the other hand, those who are married are more likely to be fluent in English, but not in French.

IV. Summary and conclusions

The analysis indicates that immigrants entering a multi-lingual country select the dominant language they learn seemingly in accordance with the model of language choice. This is done, in part, by the choice of region (Quebec or English Canada), but also by their language choice within the region in which they live.

Immigrants tend to gravitate, both by region of residence and the dominant language selected, to the language closer to their mother tongue (lower cost of obtaining dominant language fluency), to the language that predominates in their region of residence (due to exposure and economic benefits), and to the language with the broader labor market or economy (English). Thus, in English Canada immigrants from countries where a Romance language is not spoken have a very high propensity to be English-only speakers when they learn an official language. At the other extreme, French-only speakers or English-French bilinguals are much more prevalent in Quebec, especially among those coming from French-speaking or other Romance language countries. English-French bilingualism rises in importance the higher the level of schooling in both regions. French-only speakers in Quebec are particularly prevalent among those from France or Romance language countries with less skill (i.e., less schooling), who are more recent arrivals, and immigrated at an older age.

In conclusion, learning English is the dominant feature of the immigrant experience, not just in English Canada but also in Quebec, For Quebec to be successful in promoting French language fluency among the immigrants it selects it should focus on those with pre-existing French or other Romance language skills. A preference for French but not English or bilingual speakers among the immigrants would suggest focusing on those with lesser skills. Yet, French only speakers will decrease and English-French bilingualism will increase the longer the immigrants are in Canada.

Appendix 3A: The variables

Name	*Description*
Language spoken:	Ability to carry on a conversation in one of the official languages.
English only	
French only	
English/French bilingual	
Neither language	
Explanatory variables	
Education	Years of schooling completed.
Age	
Duration in Canada	Years since immigration, converted to a continuous variable.
Country of Birth[a]	Dichotomous variables for birthplace.
English-speaking	United Kingdom, Ireland, United States. British West Indies
France/Belgium	France, Belgium
Western Europe	W. Germany, Netherlands, Luxenbourg, Austria
Eastern Europe	Hungary, Poland, USSR, Czechoslovakia
Other Romance lang.	Italy, Portugal
Other South. Europe	Greece, Yugoslavia
Chinese Asia	Born in Asia of Chinese ancestry
Other Asia	
South/Central America	Excludes English speaking (British) West Indies
Africa	
Other	Oceania, other countries and country not identified.
Marital status	Unity if married, spouse present.
Married overseas	Unity if age at first marriage is less than age at arrival in Canada, and married to first spouse (year of first marriage same as the spouse).
Metropolitan area	Unity if live in a Census Metropolitan Area (a place having a population of 100000 or more).
Minority group concentration	Percentage of the Canadian population age 18 to 64 in the respondent's region (defined by province and CMA) that speaks the same non-dominant language at home. The non-dominant languages identified are Chinese, German, Italian, Ukrainian, Greek, Netherlands, Polish and Portuguese. These constitute 66% of non-dominant language responses. Only non-official languages are treated as minority languages.

Note

a Because of the broad coding of country of birth, information on mother tongue and ethnic origin are used to develop more precision (e.g., to identify English-origin immigrants from the Caribbean among the Central and South Americans, to allocate some in "other Europe" to specific countries, and to distinguish Chinese from Other Asians).

Notes

1. See Beaujot (1979) and Vaillancourt (1992) for discussions of the economic and political aspects of language in Canada.
2. See, for example, Chiswick (1991), Chiswick and Miller (1991 [1995], 1992), Grin (1990), and Veltman (1983, 1988), and the references therein. This literature also includes analyses of the impact of dominant language fluency on the economic adjustment or economic status of immigrants. These analyses show higher earnings among those more fluent in an official language. Moreover, among immigrants in Canada, those who are English-French bilinguals have higher earnings than immigrants fluent in only one of the official languages (Abbott and Beach, 1987; Chiswick and Miller, 1988).
3. For Canada, for example, see Grenier (1987) and Robinson (1988). This literature also examines the economic returns to bilingualism.
4. See, for example, the discussion of these issues for Canada and the United States in Beaujot (1979), Breton (1978) and Vaillancourt (1992).
5. There may be diminishing returns from multiple language acquisition. That is, the economic gains from acquiring fluency in a second dominant language may be smaller than the gains from the first language, even if the gains are independent of whether E or F is the first language learned.
6. The United States and Canada are each other's largest single trading partners, a pattern that is likely to persist with the implementation of the North American Free Trade Agreement.
7. The proportion is 13% for New Brunswick and 7% for Newfoundland, Nova Scotia and Prince Edward Island.
8. See Vella (1991) for an application to the multinomial logit model.
9. A referee suggested that an alternative dependent variable be defined as bilingual, first/primary language of area of residence only, second language of area of residence only, neither English nor French. This variable reflects the sequence of language acquisition. When the analysis is conducted separately for Quebec and English Canada, however, the variable is identical to that used in the equations reported in the text.
10. This is consistent with the growing skepticism in the profession over the merit of sample correction techniques.

References

Abbott MG, Beach CM (1987) Immigrant earnings differentials and cohort effects in Canada. Institute for Economic Research, Queen's University, Discussion Paper No. 705

Beaujot RP (1979) A demographic view on Canadian language policy. Can Publ Policy 5:16–29

Breton A (1978) Nationalism and language policies. Can J Econ 11:656–668

Chiswick BR (1991) Speaking, reading, and earnings among low-skilled immigrants. J Labor Econ 9:149–170

Chiswick BR, Miller PW (1988) Earnings in Canada: the roles of immigrant generation, French ethnicity, and language. Res Popul Econ 6:183–224

Chiswick BR, Miller PW (1991 [1995]) The endogeneity between language and earnings. Mimeo, Department of Economics, University of Illinois at Chicago, republished as, The endogeneity between language and earnings: International analyses. J Labor Econ 13(2):246–288

Chiswick BR, Miller PW (1992) Language in the immigrant labor market. In:

Chiswick BR (ed) Immigration, language and ethnic issues: Canada and the United States. American Enterprise Institute, Washington DC, pp 229–296

Greene WH (1992) Limdep, Version 6.0: User's Manual and Reference Guide. Castle Hill: Econometric Software Australia

Grenier G (1987) Earnings by language group in Quebec in 1980 and emigration from Quebec between 1976 and 1981. Can J Econ 20:774–791

Grin F (1990) A theory of language choice. Center for Research and Development in Economics, University of Montreal, xerox

Nelson FD (1984) Efficiency of the two-step estimator for models with endogenous sample selection. J Econometrics 24:181–196

Robinson C (1988) Language choice: the distribution of language skills and earnings in a dual-language economy. Res Labor Econ 9:53–90

Vaillancourt F (1992) Language and public policy in Canada and the United States: an economic perspective. In: Chiswick BR (ed) Immigration, language and ethnic issues: Canada and the United States. American Enterprise Institute, Washington DC, pp 179–228

Vella F (1991) Gender roles, occupational choice and gender wage differentials. Working Papers in Economics and Econometrics, The Australian National University, No. 235

Vella F (1992) Simple tests for sample selection bias in censored and discrete choice models. J Appl Economet 7:413–421

Vella F (1993) A simple estimator for simultaneous models with censored endogenous regressors. Int Econ Rev 34:441–457

Veltman C (1983) Language shifts in the United States. Mouton, Berlin

Veltman C (1988) Modelling the language process of Hispanic immigrants. Int Migrat Rev 22:545–562

4 Language in the immigrant labor market

Spoken language skill, the ability to communicate verbally, is the most basic form of human capital. It is the first type of human capital to be acquired among children, and usually the last to be lost by the aged. Spoken language skills are acquired primarily in the home as infants learn to imitate older children and adults. The initial learning by imitating is quickly followed by a learning by doing.

Spoken language skills are so basic that they are usually taken for granted. Yet it is clear that the ability to communicate verbally through a common language must have substantial economic value. Economic transactions can take place without verbal communication, but the cost of these transactions is sharply increased, and their frequency sharply decreased, when this communication cannot occur.[1]

International immigrants are perhaps the group most acutely sensitive to the importance of language capital. Moving to a country where a different language is spoken results in a depreciation of the value of this catalyst for economic and social interaction. The decision to migrate, the choice of destination, and the success of adjustment in the destination all depend, in part, on language skills.

Two key questions are addressed in this study. First, what are the determinants of the extent to which immigrants not fluent in the dominant language acquire dominant-language skills? The adjustment process may vary systematically with the immigrant's economic, human-capital, and demographic characteristics. Second, what are the impacts of dominant-language fluency on labor market outcomes, such as earnings? Economic theory predicts that, ceteris paribus, those less fluent in the dominant language of the destination will have lower earnings. The more difficult issue is determining the extent to which earnings are lower.

These research questions are important for several reasons. First, their answers will provide a better understanding of how labor markets operate and of the earnings determination process. Such understanding, of course, is important for public policy regarding efficiency, income distribution, and poverty. Another reason is that the answers shed light on the economic and noneconomic incentives for, and determinants of, an important aspect of skill formation. Third, an understanding of these issues will provide better insights regarding public policy toward the maintenance of minority languages. As is shown by Vaillancourt, public policies of the United States and Canada differ regarding minority languages and regarding the dominant languages—English in the United States and, English and French in Canada. Finally, language skills can have explicit and implicit roles in immigration policy and naturalization policy. The role of language in immigration policy differs in the two countries, as is shown by Reimers and Troper. Fluency in one or the other of the two official languages of Canada is explicitly considered in the Canadian immigration system. Proposals for explicitly incorporating English-language skills in U.S. immigration policy were under debate in the development of the 1990 Immigration Amendments, and most assuredly will emerge again in the future.

The following section of this chapter addresses the issue of the determinants of dominant-language fluency among immigrants. After developing the theoretical framework, empirical analyses are performed for adult male immigrants in the United States, using microdata from the 1980 Census of Population, and in Canada, using microdata from the 1981 Census of Canada. The section closes with a comparative analysis of the U.S.–Canadian findings.

The next section is an analysis of the role of dominant-language fluency in determining earnings in the two countries. The interaction of language fluency with other determinants of earnings is also analyzed. The theoretical discussion is followed by earnings analyses for adult male immigrants in the two countries, again using microdata from the 1980 and 1981 censuses.

The last section is a summary and conclusion, with implications for public policy. The appendixes include the language questions used in the 1980 U.S. and 1981 Canadian Censuses, a detailed discussion of the variables used in the statistical analysis, and tables of the means and standard deviations of these variables.

The determinants of dominant-language proficiency

This section presents comparative analyses of the determinants of dominant-language proficiency in both Canada and the United States. These analyses are based on census data for each country: the 1981 Census of Canada and the 1980 Census of Population, for the United States. These sources include questions on fluency in speaking the dominant languages: English in the United States, English or French in Canada. Data are not available in these sources for reading or writing skills in the dominant language. Furthermore, although there are data in the censuses on nondominant languages spoken in the home, there is no information on the degree of proficiency in speaking, reading, or writing nondominant languages.

Special emphasis is placed on defining variables that are broadly comparable across the two data sets, and this emphasis has a bearing on the manner in which the investigations are conducted. While the 1981 Census of Canada contains three questions on language, only one can be used as an indicator of fluency in a dominant language: the so-called official language question on the census permits respondents to be distinguished on the basis of their ability to speak one or both of the official languages of Canada well enough to carry on a conversation.[2]

It is not possible to construct for Canada a measure of dominant-language fluency with finer gradations from these data. In the U.S. Census, however, individuals who spoke a language other than English in the home were asked to report their level of proficiency in English. Responses were coded into four categories: very well, well, not well, and not at all. To facilitate the Canadian—U.S. comparisons, individuals who spoke only English in the home and those who spoke a language other than English in the home but spoke English very well or well are distinguished from those less fluent in English.

Previous research suggests that factors such as mother tongue, educational attainment, country of origin, duration of residence, age at arrival, and region of residence are important determinants of dominant-language proficiency.[3] To this list it seems reasonable to add variables—indicating the presence of young children in the family; whether the individual has been in the armed forces; measuring a minority group concentration; and indicating marital status and country of marriage. The relationships expected between these factors and dominant-language fluency are described below.

Country of birth may affect proficiency in the dominant language through an exposure factor. Countries differ in the extent to which particular languages are used as the dominant language, as a second language, or as the language of commerce. For example, individuals born in a predominately English-speaking country presumably know, or at least have been exposed to, the English language. Thus, immigrants from Britain, Canada, the United States, Australia, New Zealand, the British West Indies, and Ireland, for example, are arguably proficient in the English relevant for the United States and Canada—in spite of differences in accents and minor differences in terminology, idioms, and spelling.[4] For many North Europeans (such as the Dutch and Scandinavians) English is a second language, and hence English-language fluency is presumably both greater at arrival and easier to acquire than it is for others. Similarly, because of colonial experiences, among immigrants in Canada fluency in the French language may be greater or easier to acquire for those born in Vietnam or Lebanon than among those born in Korea or Turkey.

Immigrants arriving as youths are likely to gain greater fluency in the destination language than older immigrants with the same number of years in the destination. Children have a facility for acquiring new spoken-language skills that diminishes sharply as they become adults—that is, their production function for acquiring dominant-language skills is more efficient. In addition, youths gain a more intensive exposure to the dominant language through schools than adults gain at home or in the labor market, and thus youths rapidly acquire fluency in the primary language of the destination country.[5] Moreover, the benefits accruing to language skills will tend to be greater among youths, due to the complementarity between dominant-language skills and other human capital (such as schooling) and their longer payoff period. Hence, from the human capital perspective, one also would expect that immigrants arriving as youths are more likely to undertake the investments necessary to become proficient in the dominant language.

Dominant-language proficiency should vary directly with years since migration. The longer the individual has been in the host country, the more likely it is that he would have been exposed to the dominant language and hence would have acquired some language skills or would have improved existing skills. This adjustment factor has been emphasized in studies of the economic adjustment of immigrants that focus on the determinants of earnings and occupational status. The present study attempts to model the process explicitly. There may also be important interactions between levels of schooling and years since migration. Chiswick, for example, argues that in a population that initially has a very low level of fluency, the impact of education on immigrants' dominant-language fluency should increase with duration of residence.[6] His empirical evidence was consistent with this proposition.

Incentives to invest in dominant-language skills also vary with the expectation of remaining in the destination. Other factors being equal, the higher the probability of return migration in the near term, the weaker the incentives for investment in destination-specific skills, including dominant-language skills. Therefore, the greater the expectation is of return migration, the poorer the fluency in the dominant language. While data are not available on the probability of return migration for specific individuals, it is known that return migration rates are much higher for some groups than for others. In the U.S. context it is very high for Mexican immigrants but very low for another group of Hispanics—the Cubans. In the Canadian context, it is much higher for Italian and Greek immigrants than for the Vietnamese.

Although the exact causal process is open to debate, it is generally assumed that there will be a positive relationship between educational attainment and proficiency in the dominant language for immigrants from countries in which the dominant-destination language is not the primary language. For these immigrants the positive effect on fluency of preimmigration schooling may reflect the curriculum of the school attended, with second-language skills being learned only in the more advanced grades. More generally, it is likely that there is a complementarity between schooling and dominant-language proficiency. That is, those with more schooling would be more proficient in acquiring other forms of human capital, including language capital. Also, the language of instruction is generally the dominant language, and language capital perhaps more than other capital is enhanced by exposure and usage. Causality, however, may also go in the opposite direction because of the complementarity of forms of human capital. Those with greater dominant-language skills, other factors being equal, may have a greater productivity from additional schooling in the destination. The positive relationship between educational attainment and dominant-language proficiency could also be the outcome of a third process. For example, those with higher levels of ability may both acquire more schooling and be more capable of mastering other skills, such as a second language.

It has been suggested above that an individual's incentive to acquire dominant-language skills will be inversely related to the extent to which his native tongue is used in his present environment. DeVries and Vallee report that the language composition of the individual's environment is important to understanding the distribution of bilingualism in Canada.[7] Similarly, Chiswick suggests that for inhabitants of the Los Angeles area, the presence of a sizable Spanish-speaking, Mexican-origin enclave community may reduce the incentive to acquire English-language skills, as compared with other immigrants.[8] Some insights into this issue can be gained by adding statistical controls to the estimating equation for the fraction of the regional population that has the same origin-language as the individual concerned; for example, Spanish is relevant for many countries of origin. This minority-language concentration measure is similar in motivation to Veltman's battery of dummy variables for region of residence, constructed with reference to concentrations of minority-language groups within geographic units in the United States.[9]

The presence of children in the household may affect the language-acquisition process. First, children are expected to learn the dominant language more readily than their parents, in part because children have superior language-acquisition skills and in part because they are placed in circumstances that facilitate this—in school and in association with native-born children who speak the dominant language. Such skills may then be passed on to the parents within the home environment, so that parents learn the dominant language from the children.[10] Alternatively, where young children are present, parents may attempt to facilitate the assimilation of the young by learning and then speaking only the dominant language in the household. Finally, having children in the household, rather than leaving them in the origin or being childless, may reflect a stronger permanent attachment to the destination. Thus, in households where there are or have been young children, the older members of the household are expected to be more proficient in the dominant language. This effect can be expected to be larger if there is more than one child because of the language interaction between or among the children and the longer period and greater intensity of parental exposure.

Marital status could also affect dominant-language fluency, although the most important influence may come from the timing of the marriage. If the individual was married prior to migration, it is more likely that the spouse is of the same language group. It is hypothesized that this circumstance would weaken the incentive to become proficient in the language of the host country. On the other hand, marriage after migration is more likely to be with a dominant-language speaker. This suggests a differential effect of being married, depending on whether it is pre- or post-migration.

Finally, veteran status is expected to be associated with an increase in the probability that the individual is proficient in the dominant language, mainly through the remedial courses and dominant-language exposure that the armed forces offer for individuals deficient in language skills.

The next subsection contains the analysis of the determinants of English-language proficiency in the United States. Following this, dominant-language (English or French) fluency in Canada is studied. The section concludes with a comparison of the major findings from the analyses of dominant-language fluency in the two countries.

Dominant language proficiency in the United States

The study of dominant-language fluency in the United States is based on the 1980 U.S. Census of Population Public Use Microdata Sample C. All foreign born twenty-five to sixty-four-year-old males employed in 1979 in this 1/100 random sample of the population are included in the analysis. Further details on the data are presented in Appendix 4B. This appendix also contains descriptive statistics for the variables included in the estimating equation. Fully 80 percent of those in the sample are proficient in the English language, although this figure varies appreciably across birthplace regions. English-language fluency is almost universal (99.2 percent) among immigrants from English-speaking countries, while for immigrants from non-English-speaking countries the fluency rate is 76.6 percent. The mean age of the sample is forty-one years and the average immigrant has been in the United States for sixteen years and has twelve years of education. The distribution of the population across birthplace groups reveals that 17 percent of the sample are from Mexico, 28 percent from Europe, 6 percent from Canada, 9 percent from Asia (South Asia, Vietnam, or Other Asia), and 10 percent from South and Central America.

Table 4.1 presents results from Ordinary Least Squares (OLS) estimation of equations with the language-proficiency measure *GOODENG* as the dependent variable.[11] *GOODENG* equals unity when the person speaks only English in the home or, if another language is also spoken in the home, when English is spoken either very well or well. *GOODENG* equals zero for those whose English-speaking skills are not well or nil. These results are for the total adult, male, foreign-born work force. The estimates in column *a* of Table 4.1 are for a simple specification of the language model that includes neither the minority-language concentration measure that is one of the features of this study nor interaction terms between variables. This specification permits some comparisons with earlier research. Column *b* includes the minority-language concentration variable. Column *c* adds two interaction terms to the estimating equation.

Table 4.1 Regression estimates of English-language fluency among adult foreign-born men, United States, 1980

	a	*b*	*c*
Constant	0.549 (45.20)	0.568 (47.13)	0.514 (27.31)
Education	0.029 (60.75)	0.027 (58.08)	0.040 (53.20)
Age	–0.004 (16.00)	–0.004 (16.33)	–0.007 (15.66)
Years since migration (YSM)	0.019 (33.72)	0.020 (34.51)	0.026 (30.83)
YSM squared/100	–0.028 (22.98)	–0.028 (23.93)	–0.035 (25.39)
Married	0.012 (2.30)	0.012 (2.38)	0.011 (2.04)
Married overseas	–0.035 (6.86)	–0.035 (6.89)	–0.028 (5.43)
Child <6 years only	0.001 (0.20)	0.004 (0.60)	0.007 (1.07)
Child 6–17 years only	–0.003 (0.62)	–0.001 (0.13)	0.007 (1.50)
Children <6 & 6–17 years	0.003 (0.45)	0.009 (1.35)	0.018 (2.78)
Veteran	0.013 (3.22)	0.010 (2.45)	0.023 (5.86)
Rural location	–0.013 (2.20)	–0.018 (3.06)	–0.018 (3.12)
South	0.003 (0.66)	0.018 (3.89)	0.017 (3.80)
Minority-language concentration	n.e.	–0.014 (20.59)	–0.014 (20.84)
Birthplace			
Europe	–0.099 (25.65)	–0.092 (24.07)	–0.089 (24.02)
Vietnam	–0.131 (5.59)	–0.135 (5.76)	–0.150 (6.44)
Philippines	–0.018 (2.54)	–0.008 (1.19)	–0.026 (3.64)
China	–0.212 (21.20)	–0.207 (20.65)	–0.211 (21.38)
South Asia	–0.046 (7.02)	–0.047 (7.28)	–0.079 (11.93)
Other Asia	–0.156 (16.70)	–0.156 (16.71)	–0.169 (18.38)
Mexico	–0.314 (44.35)	–0.151 (14.25)	–0.142 (13.41)

Cuba	–0.282	–0.191	–0.181
	(27.19)	(16.81)	(16.04)
Other America	–0.162	–0.083	–0.080
	(22.38)	(10.76)	(10.42)
Africa	–0.028	–0.028	–0.050
	(3.53)	(3.55)	(6.57)
Middle East	–0.052	–0.053	–0.053
	(5.13)	(5.35)	(5.38)
Not reported	–0.105	–0.061	–0.061
	(12.50)	(7.80)	(7.91)
Age * YSM/100	n.e.	n.e.	0.015
			(9.52)
Education * YSM/100	n.e.	n.e.	–0.082
			(23.86)
Sample size	32,255	32,255	32,255
Adj R^2	.3540	.3660	.3813

n.e. = variable not entered.

Notes: The dependent variable is GOODENG. The *t* statistics in parentheses were derived using the White, "Heteroskedasticity-consistent Covariance Matrix Estimator."

Source: 1980 Census of Population, Public Use Sample, 1/100 sample of the foreign born.

The estimates in the linear probability model presented in column *a* have a number of distinguishing features. Each additional year of education increases the probability of being proficient in English by 2.9 percentage points.[12] This partial effect is quite large. It implies, for example, that a person who attended college for three years would have a predicted probability of being proficient in English 14.5 percentage points higher than that of a comparable individual who left school following the completion of the tenth grade.

A higher fraction of the young than of the old have an adequate command of the English language, other things being the same.[13] As these effects emerge when controlling for duration of residence in the United States, the age variable can be interpreted as a measure of age at migration. From this perspective, the results suggest that immigrants arriving as youths are more likely to become proficient in English.[14] This feature of the language-proficiency model is also evident when the age variable is replaced by an age-at-arrival measure.

Language skills increase with years since migration, but at a decreasing rate, until thirty-four years of residence.[15] This finding is consistent with the explanation often advanced in studies of earnings determination where the curvilinear relationship between earnings and duration of residence is often associated with adjustment factors such as the acquisition of language skills. It is worth noting that the Anglicization process reflected in these data continues twenty years longer than suggested by Veltman's analysis (fifteen years). This difference could reflect the different focus (all immigrants versus

Veltman's Spanish mother-tongue immigrants) or the different statistical approaches (multivariate versus Veltman's bivariate analysis). The first of these explanations is investigated below.

Birthplace is also seen to matter to the explanation of the distribution of language skills. The ranking in order of skills is: English-speaking origin (the benchmark), the Philippines, Africa, South Asia, the Middle East, Europe, Not Reported, Vietnam, Other Asia, Other America, China, Cuba, and Mexico. The relatively low ranking of the Chinese and Spanish-speaking groups is consistent with findings reported by Veltman.[16] The Asian countries fall into two groups. English proficiency is greater among immigrants from the Philippines and South Asia, both of which are multilingual areas in which a legacy of the American and British administrations is the acceptance of English as a lingua franca. The much poorer English proficiency of those from China, Vietnam, and Other Asia (primarily Korea and Japan) may be attributable to the greater linguistic distance between their native languages and English.

This argument, of course, cannot be used for the Spanish-speaking Mexican and Cuban immigrants. The similarity in the ranking, ceteris paribus, of Cuban and Mexican immigrants is somewhat surprising. Cuban immigrants have a lower probability of return migration and hence would be expected to have a greater propensity to invest in U.S.-specific human capital, including language capital. The greater incentive to invest that derives from this source, however, may be offset by the fact that refugees are less likely to be favorably selected for migration. For these immigrants, the compelling factors may be the adverse effects on English-language acquisition caused by many of the Mexicans' view of themselves as temporary migrants, by the Cubans being refugees, and, as is shown below, by both the Mexicans and the Cubans being more likely to live in large minority-language enclaves.

The results in Table 4.1 indicate that individuals who were married in the United States are more likely to be proficient in English than are those who never married, other things being the same.[17] Marriage prior to migration, however, reduces the probability of being proficient in English below the level of those who married after migration, and even below that of the unmarried. This result can be viewed as a simple extension of the language-group enclave argument; those married prior to migration are more likely to have a spouse fluent in the same immigrant language and to speak this language at home.

Veteran status is a statistically significant determinant of English-language proficiency, and it has the expected positive sign. Thus, individuals who have been in the U.S. armed forces are more likely to be proficient in English, ceteris paribus.[18] Grenier and Vaillancourt report a similar finding.[19]

The presence of young children in the household affects the level of language proficiency, but not in Table 4.1, columns *a* or *b*. The variable for the presence of children under six and from six to seventeen is stastically significant in Table 4.1, column *c*, but the variables indicating the presence of one or more children under six or between ages six and seventeen are not. That is, only the variable that unambiguously indicates the presence of at least two children is statistically significant. Perhaps it is the linguistic interaction between or among children that enhances parental fluency.[20]

There is also an interaction on English-language fluency between the effect of children in the household and duration of residence. Using a specification similar to the one in Table 4.1, column *a* and a one-in-fifty sample, the equation was recomputed for those who immigrated after 1965. The dichotomous variable for the presence of one or more children in the household has a statistically significant partial effect (coefficient, 0.0093, t-ratio = 1.98) on the English-language fluency variable.[21] Thus it appears that the presence of children has a larger effect on enhancing parental English-language fluency among the more recent immigrants.

Rural residence is associated with a marginally significant lower level of fluency, while southern residence is associated with greater reported fluency. The latter effect, however, is significant only when the minority-language concentration variable is held constant. By implication, fluency is least in the rural non-South and greatest in the urban South, other things being the same.

Column *b* in Table 4.1 adds the minority-language concentration measuring to the estimating equation. This variable is defined as the percentage of the population in the state speaking the same foreign language as the respondent for the twenty numerically most populous languages. If the respondent speaks only English at home or a language not in the top twenty, the variable is defined to be zero. There is a very strong effect of minority-language concentration. Immigrants living in states that have a relatively high representation of their language group are less likely to be fluent in English, other things being the same. Comparing, for example, a state where 1 percent of the population speaks Spanish at home with a state where 10 percent speak Spanish at home, Spanish-background workers in the second state would have a rate of English-language fluency 13 percentage points lower than similar workers in the first state.[22]

The inclusion in the estimating equation of the minority-language concentration variable has a marked impact on the Mexican, Cuban, and Other-America birthplace dummy variables. The partial effect of being born in Mexico or Other America falls by one-half, and that of being born in Cuba by one-third. As is discussed below, this enclave effect also operates within individual birthplace regions, and therefore the minority-language concentration measure does not appear to be acting simply as a surrogate for birthplace.

Table 4.1, column *c* adds several interaction terms to the basic estimating equation. The coefficient on the interaction term between age and years since migration is positive, and this reinforces a finding discussed earlier. That is, immigrants who arrive at an older age have less fluency initially but a more rapid improvement.

The second interaction term included in the Table 4.1, column *c* specification is between education and years since migration. The negative coefficient here reveals that the positive partial effect of educational attainment on English-language fluency diminishes with duration of residence in the United States. The implication is that it takes a longer duration in the United States for those with less schooling to acquire the same level of English-language proficiency. This finding contrasts with the finding reported by Chiswick for a study of low-skilled illegal aliens in the United States for a short period of time.[23] Chiswick held constant fluency at immigration, however, a variable not available in the census and positively correlated with level of education. The finding here is consistent with some other analyses of earnings determination. Another study by Chiswick, for example, reported that the partial effect of education on earnings among immigrants from English-speaking countries declines the longer they have been in the United States.[24] In other words, the complex pattern of effects that education appears to have on earnings may originate from the adjustment process associated with the learning of the dominant language.

Table 4.2 develops the analysis of the minority-language concentration effect by incorporating interaction terms into the model between the minority-language variable and education, age, and years since migration. The inclusion of these interaction terms does not affect the coefficients on other variables in any material way. The estimates listed in Table 4.2 show that the language-concentration effect varies significantly with education, age, and years since migration. The adverse effect on English-language skills of living in an ethnic-language enclave is greater for those with less skill—that is, less schooling, or for more recent arrivals who immigrated at an older age. These are the immigrants with the lowest language facility, ceteris paribus.

Further insights into the determination of English-language proficiency among immigrants can be gained by disaggregating the analysis by birthplace region. Of particular interest are the analyses, summarized in Table 4.3, for the major birthplace groups, especially the Spanish-speaking groups that have attracted the most attention in previous research.

Educational attainment exercises a strong positive influence on language fluency in each birthplace region. The effect is greatest for immigrants from Vietnam, China, Cuba, and Mexico, and lowest for immigrants from Africa, South Asia, the Middle East, the Philippines, and Europe. The former groups are characterized by relatively low language fluency at the time of arrival in the United States, and the latter groups by relatively high language fluency at arrival (see Table 4.1). It appears, therefore, that while education can help overcome language handicaps, its impact depends on the extent of initial language deficiency; education is more important the lower the initial level of proficiency.

The number of years since migration also has a strong positive influence on language fluency for all birthplace regions. The speed of language adjustment is greater among immigrants from Mexico, Cuba, and Other America than for the remaining birthplace groups.[25] These are the least endowed with respect to language skills at arrival. Hence, the story here is akin to that reported

Table 4.2 Selected regression coefficients for English-fluency model with minority-language-concentration interaction terms, adult foreign-born men, United States, 1980

	a	*b*	*c*	*d*
Education	0.027 (58.08)	0.021 (40.21)	0.021 (40.37)	0.021 (41.23)
Years since migration (YSM)	0.020 (34.51)	0.019 (33.99)	0.019 (34.09)	0.017 (28.31)
YSM squared/100	–0.028 (23.93)	–0.028 (23.61)	–0.028 (23.68)	–0.026 (21.41)
Age	–0.004 (16.33)	–0.004 (16.21)	–0.003 (14.86)	–0.003 (10.99)
Minority-language concentration (CONC)	–0.014 (20.59)	–0.030 (26.62)	–0.027 (13.84)	–0.022 (11.24)
CONC * Education	n.e.	0.002 (19.84)	0.002 (19.33)	0.001 (16.04)
CONC * Age/100	n.e.	n.e.	–0.007 (1.90)	–0.036 (8.70)
CONC * YSM/100	n.e.	n.e.	n.e.	0.060 (14.85)
Sample size	32,255	32,255	32,255	32,255
Adj R^2	.3660	.3765	.3766	.3829

n.e. = variable not entered.
Notes: Same as for Table 4.1.
In addition to the variables listed, all other control variables used in Table 4.1 are included in these equations:
Partial derivatives, from column *d*, evaluated at sample means are:

$\partial GOODENG/\partial EDUC = 0.021 + 0.001\ CONC = 0.025$

$\partial GOODENG/\partial Age = 0.026 - 0.00036\ CONC = -0.027$

$\partial GOODENG/\partial YSM = 0.017 - 0.00052\ YSM + 0.0006\ CONC = 0.011$

$\partial GOODENG/\partial CONC = -0.022 + 0.001\ EDUC - 0.00036\ \text{Age} + 0.0006\ YSM = -0.015$

Source: Same as for Table 4.1.

in the earnings determination literature: immigrants having the lowest skill level upon arrival in the United States will be characterized by a relatively rapid adjustment. This consistent pattern is suggestive of an underlying structure, common to both language-capital accumulation and all forms of human capital relevant for the destination, which gets translated in the labor market into earnings.

The minority-language concentration measure is significant and negative in six of the twelve disaggregated analyses and negative but not statistically significant in four others (Table 4.3). The estimated effects for Mexico, Cuba, and Other America are all of the same order of magnitude, suggesting that the Spanish-language groups are fairly homogeneous with respect to the

Table 4.3 Selected regression coefficients for English-language fluency by place of birth, adult foreign-born men, United States, 1980

Birthplace (% fluent)	*Education*	*YSM*	*YSM squared*[a]	*Minority concentration*	*Married overseas*	*Sample size*
Non-English total	0.029	0.022	–0.031	–0.013	–0.035	27,850
(76.63)	(57.28)	(33.59)	(22.13)	(19.74)	(6.13)	
Europe	0.022	0.023	–0.034	–0.019	–0.049	8,971
(87.54)	(27.00)	(21.10)	(16.31)	(7.53)	(5.16)	
Vietnam	0.053	0.017	–0.031	0.012	0.016	335
(70.75)	(7.01)	(1.27)	(0.55)	(1.01)	(0.29)	
Philippines	0.021	0.004	–0.005	–0.009	–0.003	1,181
(95.09)	(7.55)	(1.50)	(0.83)	(1.59)	(0.17)	
China	0.039	0.018	–0.018	–0.009	–0.048	1,289
(75.56)	(18.71)	(5.25)	(2.47)	(0.57)	(1.63)	
South Asia	0.012	0.002	0.003	0.012	0.015	1,007
(98.11)	(3.94)	(0.91)	(0.49)	(2.84)	(1.48)	
Other Asia	0.034	0.023	–0.039	–0.013	–0.030	1,575
(80.83)	(12.98)	(7.48)	(4.80)	(1.22)	(1.23)	
Mexico	0.037	0.025	–0.035	–0.012	–0.069	5,602
(48.30)	(26.86)	(13.60)	(7.94)	(9.32)	(4.57)	
Cuba	0.039	0.031	–0.034	–0.009	–0.059	1,649
(64.71)	(17.32)	(6.97)	(3.36)	(3.76)	(2.26)	
Other America	0.033	0.030	–0.049	–0.013	–0.004	3,121
(75.62)	(18.99)	(13.32)	(9.68)	(10.85)	(0.24)	
Africa	0.011	0.007	–0.013	–0.014	0.036	670
(97.31)	(3.74)	(3.20)	(2.27)	(1.90)	(2.36)	
Middle East	0.017	0.012	–0.017	–0.001	–0.040	804
(90.67)	(6.63)	(3.73)	(2.30)	(0.08)	(1.52)	
Not reported	0.021	0.015	–0.021	–0.020	–0.008	1,646
(83.84)	(9.82)	(6.55)	(4.45)	(10.90)	(0.34)	

Notes: Additional control variables are: age, married, child <6 only, child 6–17 only, children <6 and 6–17, rural, south, and veteran status. The *t* statistics in parentheses were derived using White's "Heteroskedasticity-consistent Covariance Matrix Estimator."
a. Variable divided by 100.

Source: Same as for Table 4.1.

language-enclave effect. As noted earlier, the fact that this enclave variable is significant within birthplace regions for the Spanish-origin groups indicates that it is more than a proxy for country of birth.[26] The insignificance of the language-concentration measure in other birthplace regions (Vietnam, the Philippines, China, Other Asia, the Middle East) may arise because the concentration of those speaking these languages is too small for linguistic enclaves to retard English-language fluency.[27]

Finally, the partial effect of the married-overseas variable is nearly consistently negative (Table 4.3). It is negative and significant (at the 5 percent level) in the case of immigrants from Europe, Mexico, and Cuba, and negative but not significant for most other birthplace groups.[28] The insignificance of this variable for the small sample of Vietnamese immigrants is not surprising: 92 percent of Vietnamese immigrants entered the United States after 1975, and thus few would have married in the United States prior to census enumeration in 1980. Moreover, the concentration of this wave of migration in such a short period implies that the foreign marriage variable may be measured imprecisely (see Appendix 4B).

Dominant-language proficiency in Canada

The analysis of dominant-language fluency in Canada is based on the 1981 Census of Canada. Two data files are available: the one-in-one hundred Household and Family File and the one-in-fifty Individual File. The relevant features of these two files are reported in Table 4.4.

The relative strength of the Individual File lies in the more detailed information available on home-language usage, its larger sample size, and the availability of data on citizenship. The four additional categories of home language coded in the Individual File should allow the impact of the important minority-language concentration variable to be measured more precisely. Where the focus of attention is on whether the impact of this and other variables differ between birthplace groups, the larger sample size (23,741 observations as compared with 11,382) of the Individual File provides a

Table 4.4 Characteristics of 1981 census of Canada data files

Variables available in data	*Household/Family File*	*Individual File*
Foreign marriage	Yes	Yes
Citizenship	No	Yes
Spouse's home language	Yes	No
Children's home language	Yes	No
Presence/age of children	Yes	No
Minority language	4 groups	8 groups
Sample size (adult foreign-born men)	11,382	23,741

Source: Statistics Canada: Census of Canada, 1981, Public Use Samples Tapes, User Documentation.

superior basis for analysis. Finally, the data on citizenship permits the estimation of a model of earnings determination in the next section that corresponds to that estimated using the U.S. data.

The comparative strength of the Household and Family File is that it contains data on the number and age structure of children, the language usage of children, the birthplace of spouse, and spouse's language usage that are not available from the Individual File. Therefore, only the Household and Family File permits an investigation of the key issue of whether the language attainment of adult males is related to characteristics of their spouse and children.

Full use was made of both sets of data. In the first instance a preliminary analysis was conducted using the one-in-one hundred sample from the Household and Family File to estimate the impact of children and of spouse's birthplace in the model of dominant-language proficiency in Canada. The one-in-fifty Individual File is then used to obtain a more accurate measure of the minority-language concentration effect and to examine whether this effect differs among birthplace groups.

The striking feature of the data is the very high rate of dominant-language fluency. Almost 97 percent of immigrants report themselves as able to speak English or French well enough to conduct a conversation There is some variation in dominant-language proficiency across the major birthplace regions. Immigrants from the English- or French-speaking countries have a rate of fluency, for all practical purposes, of 100 percent (two respondents in the sample reported a dominant-language deficiency), while immigrants from Chinese Asia, Southern Europe, and South and Central America have relatively lower rates of dominant-language fluency (87 percent, 92 percent, and 95 percent, respectively). The much higher rate of dominant-language fluency in Canada as contrasted with the United States (where 80 percent of immigrants are classified as fluent in the dominant language) reflects in part the different definitions used; see Appendix 4B.[29] It arises in part also because of the use in Canada of knowledge of one or both of the official languages in the immigration selection procedure.[30]

The examination of the influence of family environment factors on dominant-language proficiency in Canada based on the Household and Family File (not reported here) can be summarized succinctly. First, children do not appear to affect the dominant-language fluency of their parents. This finding may be attributable to the fact that dominant-language fluency is virtually universal in Canada, implying that the exposure factor associated with children's conversations is likely to be of minor importance.[31] Second, foreign marriage reduces the probability of dominant-language fluency in the destination country, and this effect persists when variables for the birthplace of the spouse, her home language, or her mother tongue are included in the estimating equation. This finding suggests that the foreign-marriage variable captures influences on the language outcome other than merely the country of origin of the partner or the language usage within the home.[32] Included here may be custom, cultural factors, and larger family networks in the country of origin. These factors promote a greater propensity to identify with the country of origin through both origin-language retention and eschewal of the dominant language of the destination country.

The remainder of this subsection is based on the Individual File. These analyses have a starting point similar to the analysis of the U.S. labor market presented in Table 4.1. Thus, the results of a baseline specification of a linear probability model of language fluency are presented in column *a* of Table 4.5.

Table 4.5 Regression estimates of dominant-language fluency among adult foreign-born men, Canada, 1981

	a	*b*	*c*	*d*	*e*
Constant	0.909	0.909	0.929	0.919	0.932
	(106.53)	(107.10)	(114.59)	(129.90)	(134.75)
Education	0.007	0.006	0.004	0.004	0.003
	(18.90)	(16.73)	(12.13)	(12.80)	(11.61)
Age	–0.002	–0.001	–0.001	–0.001	–0.001
	(10.98)	(10.16)	(9.43)	(9.47)	(5.69)
Years since migration (YSM)	0.006	0.007	0.007	0.007	0.005
	(13.46)	(13.99)	(14.66)	(14.85)	(10.89)
YSM squared/100	–0.009	–0.009	–0.010	–0.010	–0.008
	(9.68)	(10.39)	(11.32)	(11.56)	(9.27)
Married	–0.001	–0.001	0.001	0.001	0.003
	(0.29)	(0.30)	(0.55)	(0.63)	(1.57)
Married overseas	–0.016	–0.013	–0.013	–0.012	–0.009
	(5.32)	(4.48)	(4.37)	(4.27)	(3.18)
CMA	–0.009	–0.002	–0.001	–0.001	–0.001
	(4.61)	(0.99)	(0.78)	(0.77)	(0.33)
Province Atlantic	–0.001	–0.001	0.001	0.001	0.001
	(0.01)	(0.30)	(0.33)	(0.32)	(0.20)
Quebec	0.012	0.010	0.011	0.011	0.009
	(3.80)	(3.05)	(3.36)	(3.35)	(2.92)
Prairie	0.009	0.005	0.005	0.005	0.003
	(3.58)	(2.09)	(2.13)	(2.21)	(1.44)
British Columbia	0.007	0.007	0.004	0.004	0.006
	(3.11)	(2.77)	(1.56)	(1.58)	(2.81)
Minority-language concentration (CONC)	n.e.	–0.018	–0.059	–0.038	–0.048
		(11.49)	(13.46)	(3.84)	(4.95)
Birthplace					
Western Europe	–0.013	–0.011	–0.015	–0.015	–0.007
	(9.60)	(8.65)	(11.96)	(11.86)	(6.94)
Eastern Europe	–0.004	–0.003	–0.009	–0.009	–0.007
	(1.65)	(1.42)	(3.70)	(3.98)	(2.86)
Southern Europe	–0.057	–0.029	–0.032	–0.032	–0.029
	(17.66)	(9.22)	(10.21)	(10.11)	(9.57)

(Continued Overleaf)

Table 4.5 Continued

	a	*b*	*c*	*d*	*e*
Chinese Asia	–0.095	–0.065	–0.095	–0.094	–0.064
	(11.02)	(7.31)	(10.29)	(10.28)	(7.14)
Other Asia	0.003	0.004	0.006	0.006	–0.003
	(1.06)	(1.40)	(2.03)	(2.31)	(1.15)
Mexico, South & Central America	–0.024	–0.021	–0.023	–0.022	–0.029
	(2.53)	(2.21)	(2.33)	(2.26)	(2.92)
Africa	0.010	0.011	0.013	0.013	0.005
	(3.77)	(4.23)	(4.67)	(4.93)	(1.89)
Other	–0.003	–0.004	–0.006	–0.005	–0.007
	(0.79)	(0.95)	(1.52)	(1.42)	(1.83)
CONC * Education	n.e.	n.e.	0.005	0.004	0.004
			(11.62)	(9.39)	(9.43)
CONC * Age/100	n.e.	n.e.	n.e.	–0.039	–0.179
				(2.37)	(9.62)
CONC * YSM/100	n.e.	n.e.	n.e.	n.e.	0.351
					(15.63)
Sample size	23,741	23,741	23,741	23,741	23,741
Adj R^2	.1058	.1214	.1386	.1395	.1840

n.e. = variable not entered.

Notes: The dependent variable is *GOODLANG*. The *t* statistics in parentheses were derived using White's "Heteroskedasticity-consistent Covariance Matrix Estimator." Partial derivatives from column *e* evaluated at sample means are:

$\partial GOODLANG/\partial \text{Education} = 0.003 + 0.004\ CONC = 0.005$

$\partial GOODLANG/\partial \text{Age} = -0.001 - 0.002\ CONC = -0.002$

$\partial GOODLANG/\partial\ YSM = 0.005 - 0.00016\ YSM + 0.004\ CONC = 0.004$

$\partial GOODLANG/\partial CONC = -0.48 + 0.004\ EDUC - 0.002\ \text{Age} + 0.004\ YSM = -0.009$

Source: 1981 Census of Canada, Public Use Sample, Individual File, 1/50 sample of the foreign born.

In this equation, dominant-language proficiency is related to education, age, years since migration and its square, marital status, overseas marriage, birthplace, and province and region of residence. In the column *b* specification, the minority-language concentration measure is added to the basic estimating equation. Columns *c* through *e* list results for specifications that include interaction terms between the minority-language concentration measure and educational attainment, age, and duration of residence.

The general pattern of results in Table 4.5, column *a* is remarkably similar to the results for the United States. The magnitudes of individual estimated effects differ considerably between the two analyses, however, and these differences are discussed in the following subsection.

Years of education and age exercise major influences on dominant-language skill, with each additional year of education being associated with about one percentage point of improvement in the rate of dominant-language fluency.[33] A negative relationship exists between age and language fluency when other factors, including years since migration, are the same. That is, the older an individual at the time of migration, the less likely he is to acquire dominant-language skills.

The influence of years since migration on dominant-language proficiency is nonlinear. The partial effect of this variable on the probability of being proficient in the dominant language is given by $\partial GOODLANG/\partial YSM = 0.006–0.00018\,YSM$. Evaluated at ten, twenty (approximately the mean), and thirty years' residence in Canada, the partial effect is 0.4, 0.2, and 0.1 percentage points, respectively. While these partial effects may appear small, the years-since-migration factor has a substantial impact on the pattern of dominant-language fluency. There is, for example, an 8 percentage point difference between the rates of dominant-language proficiency of a recent arrival and of a comparable immigrant with the mean duration of residence in Canada. The process of adjustment captured by the years-since-migration variable continues for thirty-five years. Even though the United States and Canada differ greatly in terms of the relative size of their immigrant stock (6 percent versus 25 percent of their work forces, respectively), the nature of their dominant languages (English only versus English and French), and the definition of fluency, the relationship between years-since-migration and dominant-language fluency is remarkably similar.

Marital status per se does not exercise an independent influence on language skills. For individuals who were married prior to migration, however, there is a statistically significant reduction in the probability of dominant-language fluency.

Province of residence appears to exercise an independent impact. Residents of Quebec, British Columbia, and the Prairie provinces have rates of dominant-language skills that are significantly higher than those in the other provinces, but the estimated differences are quite small—about 1 percentage point in each instance.

Finally, the birthplace controls indicate that immigrants categorized as Chinese Asians have a rate of dominant-language proficiency 10 percentage points lower than that of the benchmark group of immigrants from English-speaking countries, other variables being the same. South Europeans are also distinguished by a lower level of language skills, 5.7 percentage points lower than that of the benchmark group. Three other birthplace groups are characterized by small, statistically significant differences in the level of language skills: East Europeans (at the 10 percent level), West Europeans, and South and Central Americans. Immigrants from Africa are shown to have a rate of dominant-language fluency significantly greater than that of the benchmark group of immigrants from English-speaking countries, but this result appears to derive from the application of OLS to a bounded variable having a mean close to a bound.[34] The relatively high rate of dominant-language deficiency among the South Europeans is broadly consistent with previous analyses by deVries and Vallee, who report that immigrants from Mediterranean countries have a high propensity to retain their origin language.[35]

The position of the Chinese Asians, however, does not appear to have emerged as a focal point in previous discussion. The larger coefficient for this group is consistent with the greater linguistic distance between Chinese and the dominant languages and the refugee nature of much of the migration of Chinese from Asia to Canada.

Column *b* in Table 4.5 adds the minority-language concentration variable to the analysis. As is outlined in Appendix 4B, this variable measures the percentage of the population in the region (twenty-three localities, defined by using the Census Metropolitan Area and Province variables) that has the same nondominant, home language as has the respondent. The estimated impact of the language-concentration variable is sizable (–0.018) and is highly significant ($t = 11.49$).[36] Thus, in a region with a concentration 5 percentage points above the national average of people speaking the same non-dominant home language as the respondent, the respondent's probability of being fluent in a dominant language would fall by 9 percentage points. The inclusion of the minority-language concentration variable in the analysis has a negligible impact on the estimated effects of the other regressors, other than for the South European birthplace variable. The disadvantage, ceteris paribus, of a South European birthplace declines by around 3 percentage points from –0.057 to –0.029. This decline probably reflects the explicit identification of the Greek language in the language-enclave measure used in Table 4.5.

The estimated impact of the minority-language concentration variable in Table 4.5 (–0.018) is stronger than that reported in the study of language attainment in the U.S. labor market (– 0.014), a difference that is small but statistically significant ($t = 2.35$). It is possible that this impact is associated with a difference in the method of constructing the variable for the two countries. For the United States, the variable is defined with reference to the state in which the respondent lives. For Canada, however, for approximately one-half of all respondents, the variable is defined with reference to the particular city (Census Metropolitan Area) of residence. The Canadian variable, even though defined for fewer language categories than the U.S. equivalent (eight as compared with twenty), may nevertheless provide a more accurate proxy of the underlying language-enclave effect, and this may be what is reflected in the larger estimated coefficient.

Columns *c* through *e* of Table 4.5 examine the interactions between the minority-language concentration measure and education, age, and years since migration. The findings are similar to those for the U.S. labor market: the minority-language-enclave effect is strongest among recent, adult immigrants who have below-average levels of education. The analysis of the Canadian data reveals these to be the groups possessing fewest dominant-language skills, ceteris paribus.

Table 4.6 presents results from estimation of the model of dominant-language fluency for each of the major non-English- or French-speaking birthplace groups. A number of differences are evident in the relationships

Table 4.6 Selected regression coefficients for dominant-language fluency by place of birth, adult foreign-born men, Canada, 1981

Birthplace (% fluent)	*Education*	*YSM*	*YSM squared*[a]	*Minority-concentration*[b]	*Married overseas*	*Sample size*
Nondominant language total	0.008	0.010	–0.014	–0.017	–0.017	16,092
(95.91)	(16.28)	(13.87)	(9.74)	(10.48)	(3.86)	
W. Europe[c]	0.001	0.001	–0.001	–0.003	–0.000	3,248
(99.94)	(1.37)	(0.96)	(0.97)	(0.83)	(0.07)	
E. Europe	0.012	0.004	–0.006	–0.014	–0.014	2,229
(99.01)	(2.17)	(2.39)	(1.99)	(1.99)	(2.13)	
S. Europe	0.010	0.021	–0.032	–0.015	–0.035	5,511
(91.96)	(9.67)	(9.82)	(6.73)	(8.31)	(3.52)	
Chinese Asia	0.024	0.019	–0.025	–0.008	0.002	1,132
(88.60)	(8.81)	(4.74)	(2.25)	(1.59)	(0.10)	
Other Asia	0.004	0.005	–0.008	–0.022	0.001	2,040
(98.63)	(4.02)	(3.89)	(2.96)	(1.27)	(0.14)	
Africa	0.002	0.001	–0.002	–0.001	0.003	703
(99.72)	(1.33)	(1.24)	(0.84)	(0.40)	(0.39)	
Mexico, South & Central America	0.010	0.014	–0.026	0.002	–0.012	480
(95.00)	(3.67)	(2.79)	(2.01)	(0.13)	(0.49)	
Other	0.003	0.002	–0.001	–0.008	–0.028	749
(98.93)	(2.30)	(0.95)	(0.42)	(0.65)	(2.29)	

Notes: Additional control variables are: age, married, province, and resident of metropolitan area. The *t* statistics in parentheses were derived using White's "Heteroskedasticity-consistent Covariance Matrix Estimator."

a. Variable divided by 100.

b. The eight identifiable languages used in the minority-language-concentration variable are Chinese, German, Italian, Ukrainian, Greek, Netherlandic languages, Polish, Portuguese.

c. Computed F value for equation of 1.778 is at the margin of statistical significance.

Source: Same as for Table 4.5.

between dominant-language fluency and education, duration of residence in Canada, foreign marriage, and the language-enclave variable.

A strong positive association exists between educational attainment and dominant-language fluency for most birthplace groups. The impact is largest for the Chinese Asians, the group with the lowest level of language proficiency upon arrival in Canada. Conversely, for the two groups with the highest level of initial language fluency, West Europeans and Africans, the education variable is insignificant. Hence the conclusion from this analysis parallels that for the United States: education is an important determinant of dominant-language fluency, but its importance is greater at the lower initial level of proficiency.

The impact of years since migration is generally positive, but it differs considerably across the birthplace groups. The ranking of birthplaces in terms of the impact of years since migration on language fluency is approximately the inverse of their ranking in terms of mean level of language fluency. Thus, the impact of years since migration is greatest for immigrants from Chinese Asia and Southern Europe, the two groups with the lowest mean levels of language fluency. Levels of language fluency do not vary significantly with years since migration for immigrants from Northern Europe, Africa, or the "Other" birthplace groups, each of which has a relatively high level of language skill.

As expected, the performance of the minority concentration variable is mixed because of the limited number of languages (eight) separately identified in the 1981 Census. For example, since Spanish is not separately identified, it is not surprising that the variable shows no statistical significance for the Latin American group of countries. The minority concentration variable is generally negative, but it is significant only for the East European and South European birthplace regions—which constitute 48 percent of the non-dominant language sample. While the estimated language-enclave effect for Eastern Europe is close to that derived on the basis of the aggregated data (see Table 4.5), the effect for Southern Europe is twice that reported earlier. This result may reflect the better equality of the data—that is, the use of three important language groups (Italian, Greek, and Portuguese) in the construction of the language-enclave variable relevant to the South European region.

Finally, the foreign marriage variable is significant for three birthplace groups and insignificant for the remaining five. For the cases where it is statistically significant the estimated impact is negative, but it differs by birthplace region. Marriage overseas, for example, reduces the probability of dominant-language fluency by 1.4 percentage points for immigrants from Eastern Europe, but by 3.5 percentage points for immigrants from Southern Europe.

The study of the dominant-language attainment process within each birthplace group yields a pattern of results that is broadly consistent with the aggregate analysis. The aggregate results are not dominated by one birthplace or by subtle country-of-origin interactions. Education, years since migration, foreign marriage, and minority concentration exercise important influences on dominant-language fluency. The larger impact of the human-capital variables (education and years since migration) for birthplace regions with lower initial levels of dominant-language proficiency emerges as a major finding of the disaggregated analysis.

U.S.–Canadian comparisons

Canada and the United States differ appreciably in terms of the fraction of the population who are foreign-born, the source countries of immigrants, and the methods used by the authorities for selecting immigrants. About one-quarter of the Canadian work force is foreign-born, as compared with 6 percent of the U.S. work force. In Canada the immigrant stock is largely of U.K. or European origin (66 percent), although there is a sizable group of recent Asians (13 percent). In the United States the largest immigrant group comes from Latin America. Canada has a skill-based points system for entry, whereas most of the stock of immigrants in the United States enter on the basis of kinship. Under the definitions used for this study, 97 percent of Canada's immigrants report that they are fluent in a dominant language, whereas only 80 percent of immigrants in the United States are fluent in English.[37]

Despite these differences, the immigrant experience in dominant-language fluency is remarkably similar in the United States and in Canada. Education, age at arrival, years since migration, foreign marriage, minority-language concentration, and country of birth affect dominant-language fluency in the hypothesized direction in each country. Dominant-language fluency among adult men increases with years of education and duration in the destination, and it decreases with age at arrival, foreign marriage, and minority-language concentration. Fluency also varies with country of origin in accordance with the extent to which the dominant language of the destination country is used in the origin country.

The magnitudes of the estimated effects of the explanatory variables on language fluency differ between Canada and the United States. Comparing columns *a* and *c* in Table 4.7 indicates that while the model of dominant-language fluency performs similarly in the two countries, the estimated impacts for the United States are consistently two to three times larger than those computed for Canada. The one exception is the impact of minority-language concentration on dominant-language fluency, where the impacts estimated for each country are similar. The differences between the United States and Canada may be substantive or may merely reflect the different definitions of dominant-language usage. To ascertain the weights that should be attached to these explanations, the language-fluency variable for the United States was redefined so that only individuals who spoke English "not at all" are in the not-fluency category. This gives a level of dominant-language fluency for the United States of 95 percent, which is comparable to the 97 percent fluency rate for Canada. Results from the estimation of the

Table 4.7 Partial effects of selected variables on language fluency, United States, 1980, and Canada, 1981

Variable	*United States*	*United States (assuming Canada definition)*	*Canada*
	a	*b*	*c*
Education	0.027	0.009	0.006
Age	–0.004	–0.001	–0.001
Years since migration[a]	0.014	0.006	0.005
Years since migration[b]	0.009	0.003	0.003
Married	0.012	0.009	–0.001[c]
Married overseas	–0.035	–0.012	–0.013
Minority-language concentration	–0.014	–0.005	–0.018

Notes
a. Evaluated at 10 years of residence.
b. Evaluated at 20 years of residence.
c. Estimated effect not significant at the 5 percent level.

Source: Table 4.1, column *b* and Table 4.5, column *b*.

language-fluency model using this alternative definition of the dependent variable are summarized in column *b* of Table 4.7.

For all variables other than the minority-language concentration variable, the column *b* results for the United States are of the same order of magnitude as for Canada. This suggests that the differences in results between Table 4.5 for Canada and Table 4.1 for the United States are largely definitional.

One implication is that the category for fluency in an official language is too broad in the Canadian Census, being equivalent to the "well," "very well," and "not well" categories in the U.S. data. This suggests the desirability of using a question that determines more precisely the degree of language fluency in Canada, comparable to question 13c in the 1980 United States Census (see Appendix 4A).

A striking difference characterizes the language-enclave effects between the two countries. In contrast to the other findings, the estimated impact is considerably stronger for Canada than for the United States. This difference is likely to reflect the information on city of residence used in the construction of this variable for one-half of the respondents and just province for the other half in the Canadian data, as compared with state of residence for all respondents in the U.S. analysis.

Dominant-language fluency, therefore, is amenable to statistical analysis, and such analysis yields consistent patterns for the U.S. and Canadian labor markets.

Language proficiency and earnings

This section examines the importance of proficiency in the dominant language for the explanation of variations in earnings within and across labor market groups. The framework for the analysis follows that developed in the previous section. Initially, statistical analysis of the 1980 U.S. Census of Population is conducted. Then a similar investigation of the 1981 Canadian Census is undertaken. The section concludes with a series of comparisons between the roles of dominant-language proficiency in the two North American labor markets.

The model of earnings determination employed is a human-capital earnings function in which the natural logarithm of earnings is related to years of schooling, labor market experience, weeks worked, marital status, region of residence, and a series of immigrant variables that includes birthplace, duration of residence, proficiency in the dominant language, and citizenship. In this characterization of the earnings-determination process, the duration of residence, proficiency in the dominant language, and citizenship variables capture dimensions of the economic adjustment process among immigrants.

The relationship between earnings and duration of residence is generally held to reflect the following: learning about the institutions and idiosyncrasies of the labor market of the host country; cultural adjustment factors; the development of networks of labor market contacts; and investments in country-specific human-capital skills that lead to labor market success. Included in these actions would be the acquisition of citizenship. Citizenship may open doors to better paying jobs, and it would be expected to lead to a monetary reward sufficient to offset such nonmonetary costs as the forfeiture of citizenship of the country of origin. Naturalization generally requires the demonstration of at least a minimum level of fluency in the dominant language. It also reflects a commitment to the host country. Similarly, learning the language of the host country reflects a commitment to the adopted country and an adaptation to the circumstances of that country. As has been noted previously, learning the language may provide access to better jobs and hence may be associated with higher earnings.

The rates at which different immigrant groups adjust to the labor market have been found to vary considerably. Immigrants who enter North America with relatively few internationally transferable skills (for example, immigrants from non-English speaking countries) or who are less favorably selected for migration (for example, refugees) have fewer destination-specific skills at arrival, ceteris paribus, and consistent with expectations they are typically characterized by a lower earnings profile but a relatively more rapid earnings growth with duration of residence.

Greater dominant-language fluency enhances earnings. Dominant-language fluency, however, is also expected to be related to the gains in earnings associated with language-skill acquisition. In these circumstances, because of correlation between the language-choice variable and the disturbance term, estimation of the earnings equation by least squares would in principle result in inconsistent estimates.

This feature of the data may be accommodated using either an instrumental-variable (IV) estimator or the sample selectivity methods developed by

Heckman, whereby the inverse Mills ratio is added to the estimating equation.[38] The use of the instrumental-variables estimator facilitates a test of endogeneity using the Hausman test, while the significance of the inverse Mills ratio term provides a similar test with the alternative estimator.[39] Both tests suggest that the language variable is endogenous in the earnings equations estimated for both the United States and Canada. That is, the empirical results suggest that better language skills affect earnings, and that the greater the economic return to language skills, the greater the language fluency.

The United States

Results from study of the earnings of foreign-born workers in the United States are presented in Table 4.8. Most of the variables listed in Table 4.8 were introduced in the previous section, and the definitions and measurements presented there are retained here. The new variable LN*WW* is the natural logarithm of the number of weeks worked in 1979, and the citizen variable distinguishes immigrants who became U.S. citizens from those who did not. The race variable distinguishes black immigrants from all other racial groups.

Table 4.8, column *a* presents results for a conventional specification of the human-capital-earnings function in which the explanatory variables comprise years of schooling, years of labor market experience and its square, marital status, locality, weeks worked, birthplace, duration of residence and its square, and citizenship. These results are reasonably standard, and only brief comments are provided.

There is a strong positive relationship between earnings and years of schooling. Each extra year of education is associated with 5.0 percent higher earnings, other things being the same. This coefficient is low relative to that estimated for the native-born (around 7 percent), but consistent with previous analyses of immigrants' earnings.[40]

The impact of labor market experience on earnings differs according to whether the experience was accumulated in the country of origin or in the United States. The partial effect of labor market experience in the country of origin, *EXP*, is given by the coefficients on the experience variables. Hence, $\partial \ln EARN/\partial EXP = 0.030–0.0009EXP$. Evaluated at $EXP = 10$ years this equals 2.1 percent, while after 20 years of labor market activity the earnings growth associated with experience is 1.2 percent.

Under the assumption that the cross section may be used to make longitudinal conjectures, the return to experience in the United States (that is, experience in the origin plus the differential effect of pre-immigration experience) is given as $\partial \ln EARN/\partial EXPUS = 0.053–0.00178EXPUS$. Assuming all labor market activity takes place after migration, the earnings growth with an additional year of experience is 3.5 percent when evaluated at $EXPUS = 10$, and 1.7 percent when evaluated at $EXPUS = 20$. This is larger than the effect of pre-immigration experience.

There is considerable variation in earnings across birthplace regions. In this

Table 4.8 Regression estimates of earnings equations, adult foreign-born men, United States, 1980

	Total Sample			*Fluent in English*		*Not Fluent in English*	
	OLS	*OLS*	*IV*	*OLS*	*Selectivity bias corrected*	*OLS*	*Selectivity bias corrected*
	a	*b*	*c*	*d*	*e*	*f*	*g*
Constant	4.268	4.197	4.028	4.114	3.918	4.922	4.653
	(58.07)	(57.40)	(59.63)	(46.87)	(38.70)	(27.79)	(17.19)
Education	0.050	0.046	0.037	0.053	0.057	0.015	0.010
	(39.12)	(34.99)	(13.05)	(36.12)	(32.73)	(5.29)	(2.02)
Experience	0.030	0.030	0.033	0.036	0.035	0.012	0.013
	(18.55)	(19.18)	(19.73)	(19.72)	(19.16)	(2.97)	(3.21)
Experience squared/100	−0.046	−0.046	−0.047	−0.056	−0.056	−0.020	−0.021
	(15.99)	(16.15)	(16.79)	(16.32)	(16.23)	(3.17)	(3.20)
Years since migration (YSM)	0.023	0.020	0.013	0.018	0.021	0.030	0.025
	(16.19)	(14.02)	(5.71)	(10.77)	(11.28)	(8.30)	(4.86)
YSM squared/100	−0.043	−0.039	−0.029	−0.033	−0.037	−0.062	−0.056
	(13.52)	(12.23)	(7.25)	(9.32)	(9.91)	(6.31)	(5.20)
LNWW	1.062	1.057	1.046	1.084	1.088	0.969	0.969
	(60.08)	(59.88)	(97.15)	(49.97)	(50.24)	(33.36)	(33.37)
Married	0.207	0.207	0.205	0.222	0.222	0.117	0.116
	(17.64)	(17.62)	(17.76)	(16.80)	(16.86)	(4.62)	(4.54)
Citizen	0.054	0.043	0.016	0.045	0.044	0.030	0.030
	(4.94)	(3.92)	(1.26)	(3.70)	(3.62)	(1.17)	(1.14)
Race (black)	−0.224	−0.245	−0.297	−0.259	−0.242	−0.136	−0.151
	(9.08)	(9.95)	(10.66)	(10.15)	(9.35)	(1.35)	(1.49)
Rural location	−0.070	−0.068	−0.065	−0.081	−0.081	−0.056	−0.055
	(3.76)	(3.68)	(3.92)	(3.80)	(3.82)	(1.54)	(1.52)

(Continued Overleaf)

Table 4.8 Continued

	Total Sample			*Fluent in English*		*Not Fluent in English*	
	OLS	*OLS*	*IV*	*OLS*	*Selectivity bias corrected*	*OLS*	*Selectivity bias corrected*
	a	*b*	*c*	*d*	*e*	*f*	*g*
South	–0.065	–0.065	–0.064	–0.062	–0.061	–0.087	–0.085
	(5.43)	(5.42)	(5.53)	(4.68)	(4.60)	(3.25)	(3.20)
Birthplace							
Ireland	–0.178	–0.180	–0.186	–0.163	–0.160	n.e.	n.e.
	(4.14)	(4.21)	(4.05)	(3.81)	(3.70)		
Canada	–0.087	–0.086	–0.085	–0.074	–0.074	n.e.	n.e.
	(3.15)	(3.14)	(2.98)	(2.69)	(2.72)		
West Indies	–0.157	–0.160	–0.165	–0.126	–0.123	–0.494	–0.463
	(4.14)	(4.20)	(3.98)	(3.30)	(3.22)	(0.92)	(0.86)
Europe	–0.140	–0.126	–0.092	–0.120	–0.131	0.077	0.126
	(6.21)	(5.59)	(3.66)	(5.32)	(5.77)	(0.61)	(0.95)
Vietnam	–0.297	–0.281	–0.242	–0.268	–0.287	–0.125	–0.073
	(6.43)	(6.10)	(4.85)	(5.05)	(5.37)	(0.86)	(0.48)
Philippines	–0.310	–0.310	–0.309	–0.328	–0.326	0.110	0.131
	(10.55)	(10.51)	(9.67)	(10.91)	(10.87)	(0.81)	(0.95)
China	–0.364	–0.332	–0.253	–0.289	–0.315	–0.303	–0.231
	(11.87)	(10.81)	(6.80)	(8.85)	(9.43)	(2.30)	(1.62)
South Asia	–0.144	–0.141	–0.133	–0.159	–0.161	–0.136	–0.128
	(4.29)	(4.19)	(3.94)	(4.69)	(4.77)	(0.65)	(0.62)
Other Asia	–0.244	–0.222	–0.169	–0.218	–0.238	–0.001	0.060
	(7.80)	(7.10)	(5.14)	(6.60)	(7.16)	(0.01)	(0.42)
Mexico	–0.333	–0.286	–0.173	–0.273	–0.319	–0.188	–0.105

	(13.45)	(11.47)	(4.43)	(10.32)	(11.04)	(1.51)	(0.75)
Cuba	−0.325	−0.280	−0.174	−0.263	−0.299	−0.089	−0.002
	(11.31)	(9.77)	(4.26)	(8.72)	(9.51)	(0.68)	(0.02)
Other America	−0.335	−0.308	−0.245	−0.265	−0.288	−0.247	−0.181
	(13.01)	(11.99)	(7.73)	(9.93)	(10.60)	(1.94)	(1.30)
Africa	−0.195	−0.186	−0.165	−0.181	−0.187	−0.142	−0.120
	(5.07)	(4.84)	(4.21)	(4.66)	(4.82)	(0.57)	(0.48)
Middle East	−0.219	−0.213	−0.199	−0.199	−0.203	−0.128	−0.099
	(5.72)	(5.57)	(5.55)	(5.15)	(5.27)	(0.73)	(0.56)
Not Reported	−0.298	−0.281	−0.239	−0.266	−0.279	−0.121	−0.061
	(9.75)	(9.17)	(7.51)	(8.19)	(8.55)	(0.91)	(0.43)
Proficient in English	n.e.	0.169	0.571	n.e.	n.e.	n.e.	n.e.
		(12.52)	(5.43)				
Lambda	n.e.	n.e.	n.e.	n.e.	0.399	n.e.	−0.216
					(3.74)		(1.38)
Sample size	32,255	32,255	32,255	25,713	25,713	6,542	6,542
Adj R^2	.3856	.3886		.3632	.3635	.3176	.3177

n.e. = variable not entered.

Notes: Dependent variable is natural logarithm of earnings in 1979. The *t* statistics in parentheses were computed using White's "Heteroskedasticity-consistent Covariance Matrix Estimator."

Source: Same as for Table 4.1.

analysis, Britain is used as the reference group. Each of the fifteen birthplace-dichotomous variables is negative and statistically significant, indicating that members of the particular birthplace have earnings lower than immigrants from Britain. The ranking of birthplaces in terms of decreasing earnings is: Canada, Europe, South Asia, Ireland, the Middle East, Other Asia, Vietnam, Africa, the Phillippines, Not Reported, Cuba, Mexico, West Indies, Other America, and China. The estimated coefficients range from –0.09 to –0.38, indicating a percentage earnings differential of between 7 percent and 32 percent.

The estimating equation (Table 4.8, column *a*) shows that married (spouse present) men have earnings considerably higher than those in other marital statuses, that citizens have a small (5 percent) earnings advantage, and that residents of southern states or of rural areas each have earnings 5 percent lower than residents of other localities, ceteris paribus. Black immigrants have earnings about 20 percent lower than other immigrants (coefficient –0.22), even after controlling for schooling and country of origin. It is not clear to what extent the race differential reflects discrimination, and if so whether it is discrimination in the origin or the destination.

The elasticity of earnings with respect to weeks worked is 1.062, and this is significantly different from one. In other words, full-year workers receive 6 percent higher weekly earnings than part-year workers, ceteris paribus. This difference may reflect dimensions of human capital accumulation by those with a greater attachment to the labor market not captured by the proxy for labor market experience. It may also reflect the effects of an upward-rising labor supply curve (that is, those with higher wages working more weeks) and the positive correlation of hours worked per week and weeks worked per year.

One attribute that has not been accounted for in the Table 4.8, column *a* specification is knowledge of the English language. This knowledge is expected to play a major role in explaining variations in earnings. There is a difference of .611 in the mean logarithmic earnings of foreign-born residents of the United States who are fluent in English and those who have an English-language deficiency, implying an earnings differential of approximately 46 percent. The relatively short period of time those with an English-language deficiency have been in the United States (ten years versus seventeen years), the fewer years of schooling that they possess (eight years versus thirteen years), and the fact that they work, on average, three weeks per year fewer than other immigrants who are fluent in English are factors likely to contribute to the difference in observed earnings.

To isolate the impact on earnings of variables other than English-language deficiency so that the effect of fluency can be estimated, the dichotomous English-fluency variable used in the previous section is added to the conventional human-capital earnings function. Results are presented in column *b* of Table 4.8. Individuals who are fluent in English have 16.9 percent higher earnings than other groups, ceteris paribus. This earnings advantage is of the same order of magnitude as that reported by Fishback and Terza for all workers.[41] As the unadjusted earnings differential was 46 percent, this suggests that differences in measurable endowments account for two-thirds of the observed, unadjusted earnings differential between the two levels of fluency.

When the *GOODENG* variable is included in the estimating equation, there are minor changes to some other coefficients (Table 4.8, column *b*). Several birthplace coefficients fall by moderate amounts—for example, China by 3 percentage points, Cuba and Mexico by 5 percentage points—and the partial effect of duration of residence in the United States is reduced and is given by $\partial \text{In}EARN/\partial YSM = 0.020–0.00078\,YSM$. Evaluated at $YSM = 10$ this yields 1.2 percent, as compared with 1.4 percent when *GOODENG* is excluded from the model. While this decline can be noted, it is important to emphasize that even when the language-proficiency variable is included in the model, years since migration still exercise a pronounced impact on earnings.[42] This is consistent with other studies of immigrant earnings.[43]

A number of other specifications of the earnings equation (not reported here) were estimated. They included interaction terms between duration of residence and the human-capital variables for years of schooling and for experience in the country of origin. Both of these variables were significant and positive. Thus the earnings growth with years in the United States is greater for the better educated than for the less educated, and also greater for immigrants possessing greater levels of overseas labor market experience. The interaction terms therefore provide evidence of complementarity between the human capital represented by schooling and years of pre-immigration experience and that represented by the duration-of-residence variable.

The remainder of Table 4.8 focuses on the potential endogeneity of proficiency in the English language. Column *c* presents results derived using an instrumental-variables (IV) estimator. The instruments for the *GOODENG* variable are all the explanatory variables in Table 4.1, with the identifying instruments being the veteran status, children, foreign-marriage and minority-language concentration measures. There are a number of differences between the OLS and instrumental variables estimates, and as would be expected, the most pronounced change occurs in the *GOODENG* variable. This increases from .169 in OLS to .571 with the instrumental variables approach. Although this is a dramatic change, it is noted that similar changes have been reported elsewhere. In Robinson's analysis of the 1981 Canadian Census, for example, the coefficient recording the wage premium to bilingualism increased by a factor of 2.5 when an instrumental-variables estimator was used instead of OLS.[44] The increase in the wage premium to dominant-language fluency is associated with a reduction in the apparent rewards to other (complementary) types of human-capital investment, such as formal education, labor market experience, and years since arrival.

These results are open to a number of interpretations. They could derive from the endogeneity of dominant-language attainment in earnings determination and thus indicate that the notion of endogeneity should be treated seriously. Alternatively, the dominant-language fluency variable may be measured with considerable random error, which results in a downward bias in the OLS estimates when compared with the value of 0.571 derived using instrumental variables. If so, the self-reported measures of language fluency in the census should be viewed with considerable skepticism. Finally, the large difference between the OLS and instrumental-variables estimates may reflect in part the quality of the instruments available for *GOODENG*: where the instruments have a low correlation with *GOODENG*, the instrumental-variables estimates will be consistent but will have a large variance relative to OLS. This caveat to the method should be kept in mind when interpreting the results.

Further evidence of the endogeneity of dominant-language attainment in earnings determination is found in Table 4.8, columns *d* through *g*. Here the sample is separated according to language proficiency, and separate equations are estimated for each language group, with and without a correction for sample selection bias.[45] There are a number of minor differences between the results for the sample of workers who are fluent in English and the results discussed above, the marginally higher earnings growth associated with both formal education and labor market experience being the most important. The statistical significance of the inverse Mills ratio term (lambda) provides one test of the exogeneity of the language-attainment variable. This sample selection term is significant ($t = 3.11$) and positive. That is, workers become fluent in English if their unobservable skills are more highly rewarded when they are fluent.

The equation estimated for the sample reporting an English-language deficiency Table 4.8, columns *f* and *g* has a number of features. The earnings growth associated with both formal education and labor market experience is markedly lower than for comparable workers possessing English-language fluency. This suggests a degree of complementarity between types of human-capital skills. The premium to labor market experience in the United States, however, as compared with experience in the country of origin, is higher for workers who are not fluent in English. Evaluated at ten years of residence, for example, the partial effect of years since migration on earnings is 1.2 percentage points for individuals with English-language fluency, but 1.8 percentage points for individuals not fluent (see Table 4.8, columns *d* and *f*). The third characteristic of these results is that the earnings differences across birthplace groups is smaller within each of the two broad fluency groups than it is for the sample as a whole.

Finally, among the immigrants with an English-language deficiency, the sample selection term is negative but statistically insignificant ($t = 1.38$). As the lambda variable for this equation is constructed to be negative, the negative sign indicates positive self-selection in this instance. That is, individuals who are not fluent in the dominant language have above-average levels of the unobserved skills that determine earnings in the nonfluent language market. This provides further support for the hypothesis that English language fluency is endogenous.

A summary of the exogeneity issue may be provided by pooling the two samples and estimating an equation that includes the two sample selection terms.[46] The F test on the incremental contribution of the two auxiliary regressors is 29.859, which is significant at the 5 percent level. This indicates that exogeneity of the English-language fluency variable is rejected.[47]

Thus, there is a consistent set of evidence: immigrants in the United States who are proficient in English have higher earnings than individuals with an English-language deficiency, ceteris paribus, and English-language fluency appears to be the outcome of a choice process, determined in part by the economic returns from acquiring language skills. Hence the acquisition of language capital, as with other forms of human capital, is responsive to economic incentives.

Canada

The average annual earnings of immigrant workers in Canada who are fluent in a dominant language are 49 percent higher than the earnings of immigrant workers who lack this skill. Individuals who possess dominant-language skills are also relatively well endowed in most other skills associated with higher earnings. Their average level of schooling is 11.8 years and their average duration of residence in Canada 19.7 years, as compared with the averages of 7.1 and 11.9 years for workers who lack fluency in a dominant language. While workers with a dominant-language deficiency have more years of labor market experience (34 as compared with 26), two-thirds of this experience was accumulated in the country of origin.

The independent effect on earnings associated with dominant-language fluency is analyzed in this section using the earnings functions presented in Table 4.9. These estimates are derived for twenty-five to sixty-four-year-old foreign-born male workers in the one-in-fifty sample Individual File of the 1981 Census of Canada. The approach followed is similar to that adopted to study earnings determination in the U.S. labor market. Hence column *a* presents results for a conventional specification of the augmented human-capital earnings equation, in which the natural logarithm of annual earnings is related to years of schooling, years of labor market experience and its square, marital status, locality, weeks worked, duration of residence and its square, citizenship, and birthplace. The general patterns that emerge from this analysis are consistent with those highlighted in the study of the U.S. labor market.

Earnings increase more than proportionately with weeks worked—the elasticity of earnings with respect to weeks worked is 1.031. This elasticity is considerably higher than that reported in earlier studies, but this difference can be linked to the treatment of workers who reported nonpositive

Table 4.9 Regression estimates of earnings equations, adult foreign-born men, Canada, 1980

	Total Sample			*Fluent in a Dominant Language*		*Not Fluent in a Dominant Language*	
	OLS	*OLS*	*IV*	*OLS*	*Selectivity bias corrected*	*OLS*	*Selectivity bias corrected*
	a	*b*	*c*	*d*	*e*	*f*	*g*
Constant	4.447	4.347	4.105	4.402	4.207	5.011	6.345
	(44.26)	(39.44)	(15.37)	(43.54)	(27.58)	(8.94)	(6.95)
Education	0.045	0.044	0.043	0.045	0.046	–0.014	0.006
	(19.63)	(19.37)	(16.73)	(19.83)	(19.92)	(0.70)	(0.27)
Experience	0.026	0.026	0.026	0.027	0.027	0.012	–0.001
	(10.23)	(10.20)	(10.07)	(10.56)	(10.57)	(0.53)	(0.01)
Experience squared/100	–0.050	–0.050	–0.048	–0.052	–0.052	–0.025	–0.023
	(10.51)	(10.40)	(9.99)	(10.69)	(10.81)	(0.76)	(0.69)
Years since migration (YSM)	0.025	0.024	0.023	0.025	0.026	0.003	0.030
	(8.37)	(8.14)	(7.78)	(8.27)	(8.46)	(0.11)	(1.07)
YSM squared/100	–0.042	–0.042	–0.04	–0.042	–0.043	–0.024	–0.052
	(6.37)	(6.27)	(6.68)	(6.29)	(6.42)	(0.35)	(0.74)
LNWW	1.031	1.029	1.025	1.036	1.039	0.949	0.988
	(42.64)	(42.54)	(59.87)	(42.30)	(42.29)	(9.02)	(9.28)
Married	0.210	0.211	0.213	0.214	0.212	0.029	0.001
	(11.70)	(11.75)	(12.34)	(11.83)	(11.75)	(0.21)	(0.01)
Citizen	0.071	0.067	0.058	0.065	0.066	0.148	0.153
	(4.12)	(3.89)	(2.98)	(3.74)	(3.79)	(1.25)	(1.30)
CMA	0.077	0.078	0.081	0.080	0.078	–0.121	–0.195
	(4.82)	(4.90)	(5.15)	(4.99)	(4.87)	(0.81)	(1.26)
Province							
Atlantic	0.038	0.038	0.037	0.039	0.039	–0.101	–0.136
	(0.80)	(0.80)	(0.83)	(0.81)	(0.81)	(0.18)	(0.24)

Quebec	–0.047	–0.049	–0.053	–0.036	–0.034	–0.388	–0.303
	(2.48)	(2.56)	(2.62)	(1.90)	(1.78)	(2.65)	(2.00)
Prairie	0.104	0.103	0.100	0.110	0.112	–0.080	–0.037
	(5.46)	(5.39)	(5.12)	(5.74)	(5.82)	(0.40)	(0.19)
British Columbia	0.115	0.114	0.112	0.110	0.111	0.534	0.572
	(6.36)	(6.31)	(5.97)	(6.03)	(6.10)	(3.35)	(3.60)
Birthplace							
Ireland	–0.162	–0.162	–0.163	–0.162	–0.162	–	–
	(1.61)	(1.61)	(1.88)	(1.61)	(1.61)		
United States	–0.123	–0.123	–0.124	–0.124	–0.124	–	–
	(3.93)	(3.93)	(4.02)	(3.94)	(3.97)		
West Indies	–0.225	–0.228	–0.234	–0.225	–0.220	0.783	0.816
	(6.92)	(6.99)	(6.73)	(6.90)	(6.76)	(3.77)	(3.93)
France	–0.115	–0.114	–0.110	–0.124	–0.125	1.787	1.876
	(2.51)	(2.48)	(2.08)	(2.70)	(2.73)	(6.92)	(7.04)
W. Europe	–0.138	–0.136	–0.133	–0.137	–0.139	0.821	1.042
	(6.35)	(6.29)	(5.83)	(6.33)	(6.39)	(1.34)	(2.13)
E. Europe	–0.182	–0.182	–0.181	–0.184	–0.184	0.962	1.121
	(6.86)	(6.84)	(7.03)	(6.88)	(6.90)	(4.62)	(5.18)
S. Europe	–0.140	–0.133	–0.117	–0.138	–0.148	0.750	0.812
	(6.86)	(6.49)	(4.30)	(6.73)	(6.97)	(4.26)	(4.53)
Chinese Asia	–0.344	–0.332	–0.304	–0.286	–0.301	n.e.	n.e.
	(10.16)	(9.85)	(6.75)	(8.41)	(8.70)		
Other Asia	–0.237	–0.237	–0.237	–0.234	–0.234	0.206	0.477
	(8.95)	(8.95)	(8.69)	(8.86)	(8.83)	(0.75)	(1.61)
Mexico, South & Central America	–0.296	–0.294	–0.286	–0.291	–0.296	0.484	0.643
	(6.66)	(6.60)	(5.90)	(6.56)	(6.67)	(1.56)	(2.05)
Africa	–0.130	–0.131	–0.133	–0.134	–0.132	0.906	1.331
	(3.73)	(3.75)	(3.27)	(3.82)	(3.76)	(3.06)	(4.34)
Other	–0.103	–0.103	–0.103	–0.105	–0.106	0.973	1.146
	(2.67)	(2.67)	(2.68)	(2.73)	(2.73)	(3.64)	(4.10)

(Continued Overleaf)

Table 4.9 Continued

	Total Sample			*Fluent in a Dominant Language*		*Not Fluent in a Dominant Language*	
	OLS	*OLS*	*IV*	*OLS*	*Selectivity bias corrected*	*OLS*	*Selectivity bias corrected*
	a	*b*	*c*	*d*	*e*	*f*	*g*
Dominant-language proficiency	n.e.	0.122	0.414	n.e.	n.e.	n.e.	n.e.
		(2.43)	(1.34)				
Lambda	n.e.	n.e.	n.e.	n.e.	0.577	n.e.	1.160
					(1.71)		(1.92)
Sample size	23,741	23,741	23,741	23,081	23,081	660	660
Adj R^2	.2217	.2220		.2161	.2162	.2131	.2157

n.e. = variable not entered.
Notes: Dependent variable is natural logarithm of earnings in 1980. The *t* statistics in parentheses were computed using White's "Heteroskedasticity-consistent Covariance Matrix Estimator."

Source: Same as for Table 4.5.

earnings.[48] When this group is excluded from the analysis, the elasticity coefficient drops to 0.917, a level consistent with previous research on Canada.[49]

Earnings increase by 4.5 percent with each additional year of education, and by 1.6 percent with each additional year of labor market experience in the country of origin (evaluated at $EXP = 10$). Labor market experience in Canada is associated with an earnings premium compared with experience prior to migration. Evaluated at ten years of residence in Canada, the premium is a sizable 1.7 percent. Even after twenty years of residence in Canada, an extra year of Canadian labor market experience is worth 0.8 percentage points more in earnings than is experience in the country of origin.

Region of residence exercises an important influence on earnings. Residents of Census Metropolitan Areas have earnings approximately 8 percent higher than those of workers who live outside the major cities. The ranking of immigrants' earnings across provinces is similar to that reported by Chiswick and Miller.[50] Thus the earnings of residents of Quebec are 5 percent lower and the earnings of residents of the Prairie provinces and British Columbia are about 10 percent higher than the earnings of residents of the other provinces. The earnings disadvantage associated with residence in Quebec among immigrants may explain why immigrants tend to avoid this province.

Country of origin is very important for understanding variation in earnings in the Canadian labor market. Each of the birthplace groups has earnings significantly lower than the earnings of immigrants from Britain, ceteris paribus, although the Irish coefficient is at the margin of significance. The ranking of earnings in decreasing order is: Britain (the benchmark), Other, France, the United States, Africa, Western Europe, Southern Europe, Ireland, Eastern Europe, West Indies, Other Asia, South and Central America, Chinese Asia. At the lowest end of the spectrum, the earnings of immigrants from South and Central America and Chinese Asia are approximately 30 percentage points lower than those of the British. The earnings of immigrants from the United States are 12 percentage points lower than those of the British.[51]

The earnings of immigrants who have become Canadian citizens are 7 percent higher than those for noncitizens, other variables being the same. This sizable earnings premium may reflect in part the use of citizenship status as a screen for access to higher paying jobs, or the greater motivation and commitment to the Canadian labor market of individuals seeking citizenship.

In Table 4.9, column *b* the dominant-language proficiency measure is included in the estimating equation. Individuals who are proficient in a dominant language have earnings 12.2 percentage points higher than individuals who lack this skill, other things being the same. The inclusion of the dominant-language proficiency variable has a negligible impact on all other estimated coefficients. In particular, the partial effect of years since migration on earnings is not affected in any material way (a reduction from 1.7 percentage points to 1.6 percentage points, evaluated at ten years of residence in Canada). This finding is consistent with the evidence reported in Abbott and

Beach and Chiswick and Miller for quite different specifications of the language-fluency variable.[52] It appears, therefore, that the economic progress of immigrants in Canada reflected in the duration-of-residence variable arises from a source other than merely the accumulation of language capital.

The results listed in Table 4.9, column *c* are derived using an instrumental-variables method of estimation. In this model the foreign-marriage and minority-language-concentration measures are used as the identifying instruments for the dominant-language proficiency variable. The comparison between the OLS and instrumental-variables coefficients in Table 4.9 is similar to that found in the U.S. data. Hence the coefficient on the dominant-language proficiency variable increases threefold. In this case, however, it is statistically insignificant. In the U.S. study, the language variable was highly significant in the instrumental variables model (t = 5.43).

This difference may indicate that the problems of errors in variables and endogeneity are less serious in the analysis for Canada, where the language question is less subjective (see Appendix 4A) and the level of dominant-language fluency considerably higher. The finding could simply be caused, however, by the identifying instruments being less suitable in the analysis of earnings determination in Canada than in the same model applied to the U.S. labor market. In the U.S. labor market, the coefficient of determination (adjusted R^2) in the model of dominant-language fluency was .37 (see Table 4.1). For the study of Canada, however, the coefficent of determination is only .12 (see Table 4.5). As there is an inverse relationship between the asymptotic variance of the instrumental-variables estimator and the asymptotic correlation between the instruments and the variable instrumented, the instrumental variables would be expected to be less successful when applied to the Canadian data than to the U.S. data.

The application of the control function method (Table 4.9, columns *d* through *g*) yields results that are more consistent with the findings reported previously for the United States. Columns *d* and *e* list estimates of earnings equations for the portion of the sample reporting that they are fluent in a dominant language, while columns *f* and *g* list estimates for immigrants who are not fluent. Both OLS and selectivity corrected estimates are presented.

Individuals who are fluent in a dominant language make up 97 percent of the total sample. Consequently, the OLS estimates for this group do not differ appreciably from those listed for the total sample. The coefficient on the sample selectivity correction term is positive and at the margin of statistical significance (t = 1.71). Thus some, although not overwhelming evidence suggests that the sample of dominant-language speakers is nonrandom. The high representation of this group in the total sample (97 percent) may have an important bearing on this outcome. Correction for sample selectivity does not affect the estimated coefficients in the model.

Columns *f* and *g* in Table 4.9 list results for the portion of the sample that lacks fluency in a dominant language. The sample here is relatively small (660 observations), and the human-capital variables (education, pre- and post-immigration labor market experience) are statistically insignificant. There is considerable variation, however, in earnings across birthplace groups. Because the British birthplace group is not represented in this sample, the Chinese Asia immigrants serve as the benchmark. Compared with this group, all birthplace regions except Other Asia, South and Central America, and Western Europe have higher earnings. The selectivity correction term (lambda) is positive and has a t of 1.92. This provides support for the hypothesis of endogeneity of language skills in the model of earnings determination. Correction for sample selection does not affect the other coefficients, but it is associated with a widening of the earnings differences across birthplace groups; all birthplace variables other than those for Other Asia are significant once the nonrandom nature of the sample is taken into account.

The evidence contained in Table 4.9, while not as conclusive as for the study of the U.S. labor market, suggests that dominant-language fluency is endogenous: that is, it is in part determined by earnings. Further evidence to this effect is provided by the estimation of pooled equations for the two language groups that contain both sample selection correction terms used in the analyses discussed above. In this analysis the F-test on the incremental contribution of the lambda terms was 7.327, which is statistically significant. Hence this summary measure of the endogeneity issue suggests that it is important. Consistent with this finding, the Addison and Portugal test returned an F-statistic of 10.882.[53]

U.S.–Canadian comparisons

The main feature of the comparative study of the determinants of earnings among immigrants in Canada and the United States is the overwhelming similarity of the findings, as summarized in Table 4.10.

In both of the North American labor markets, the earnings growth associated with an extra year of schooling is roughly 4.5 percent. The increase in earnings associated with labor market experience differs according to whether the experience was accumulated in the country of origin or in the destination labor market. An extra year of labor market experience results in earnings roughly 2 percent higher (evaluated at experience of ten years) if the experience was accumulated in the country of origin, and 3.3 percent higher earnings if it was accumulated in the destination. Citizens earn more than noncitizens in each country. Although the U.S. and Canadian labor markets are contiguous, the similarity of these effects in the earnings model is remarkable.

Individuals who are proficient in the dominant language in the United States have earnings 16.9 percent higher than those who lack this skill. In Canada, however, the earnings premium associated with dominant-language fluency is only 12.2 percent. This difference is not statistically significant. When earnings equations are estimated for the United States using the Canadian definition of dominant-language fluency, the earnings premium

Table 4.10 Partial effects on earnings of selected variables, United States, 1980, and Canada, 1981

Variable	*United States*	*Canada*
Education	0.046	0.044
Experience in origin[a]	0.021	0.016
Experience in origin[b]	0.012	0.006
Experience in destination[a]	0.033	0.032
Experience in destination[b]	0.016	0.013
Weeks worked	1.057	1.029
Married	0.207	0.211
Citizen	0.043	0.067
Proficiency in dominant language(s)	0.169	0.122
Proficiency in dominant language(s)[c]	0.127	0.122

Notes
a. Evaluated at experience of 10 years.
b. Evaluated at experience of 20 years.
c. Canadian definition for the United States.

Source: Table 4.8, column *b* and Table 4.9, column *b*.

associated with language skills in the United States is 12.7 percent. When the instrumental variables approach is used, however, the effects of dominant-language fluency are 57.1 percent and 41.4 percent for the U.S. and Canadian labor markets, respectively, but the coefficient is less statistically reliable in the Canadian analysis. The analyses also demonstrate that in each country, dominant-language skills are endogenous with respect to labor market earnings.

Conclusion

This chapter has explored the determinants and labor market implications for immigrants of proficiency in speaking the dominant language in the country of destination. The statistical analysis uses the microdata files on adult foreign-born men in the 1980 and 1981 censuses of the United States and Canada, respectively. The languages treated as dominant are English in the United States and English and French in Canada. The analyses are based on the self-reported responses to questions on spoken-language fluency.

The findings in the analysis of the determinants of language proficiency for the United States and Canada are remarkably similar, and the findings are similar when the analysis is done separately by country of origin for the immigrants. It is shown that in both countries dominant-language fluency varies systematically with exposure, efficiency, and economic variables. Language skills are shown to be determined endogenously with earnings.

Dominant-language fluency can be viewed as produced by the individual. This process is more efficient the greater the exposure is to the dominant language prior to immigration and the younger the age at immigration, apparently because younger people are more efficient in creating language capital. Greater fluency is also achieved by those who have more schooling, presumably because of the complementarity of various types of human capital. The advantageous position of those with more schooling diminishes but does not disappear, however, with a longer duration of residence.

Learning by doing is particularly important for language skills, and a longer duration in the destination enhances fluency. The effect of duration of residence on language skills is larger for those who immigrate at an older age and with less schooling. In general, immigrants with the poorest fluency at arrival undergo the most rapid improvement with experience in the destination.

Family characteristics also appear to matter. Those who are less likely to speak the dominant language at home (for example, because their spouses speak the same nondominant language and there are no children in the households) have lower levels of fluency.

A very important determinant of dominant-language proficiency is the extent to which others in the area in which the respondent lives speak the same nondominant language. That is, immigrants living in communities where their nondominant language of origin is spoken with greater frequency have a lower level of fluency in the dominant language. The adverse effect of a language enclave, however, is not neutral. It is more intense during the initial years in the destination for less educated immigrants and for those who immigrated as adults. These are the very immigrants with the lowest level of language fluency.

The statistical analysis of earnings in the two countries uses as the starting point the standard human-capital earnings function, augmented for immigrant analyses. A dichotomous variable for being fluent in the dominant language is then added to the analysis. Yet the self-reported language variable may be subject to much random measurement error, and language fluency may be determined endogenously with earnings. That is, those who have a greater economic incentive to acquire fluency in the dominant language may have a higher degree of fluency. As a result, the analysis explores alternative statistical methodologies for the two countries, including ordinary least squares, instrumental variables, and sample selectivity techniques.

The determinants of earnings among immigrants are remarkably similar in the United States and Canada; it is as if there is one earnings determination process for the two countries. Using the ordinary least-squares methodology, those in the United States who speak English well or very well have 17 percent higher earnings than those with less fluency, while in Canada those who can carry on a conversation in English or French have 12 percent higher earnings than those who cannot. Converting the U.S. data to a close approximation of the less satisfactory Canadian definition, those who are fluent in English also have 12 percent higher earnings. The instrumental-variables approach indicates an even larger effect of dominant-language fluency—roughly 50 percent.

The sample selectivity test addresses the issue of the endogeneity of fluency.

The test indicates that workers are more likely to become fluent in the dominant language if their unobservable characteristics are more highly rewarded when they are fluent. Thus, the acquisition of language capital appears to be responsive to the economic incentives for acquiring language skills.

The addition of the language-proficiency variable to the earnings equation, whether using the observed value or an instrumental-variables approach, has little effect on the size or statistical significance of the coefficients for the other variables in the analysis. In the instrumental-variables analysis, however, there is a diminution in the partial effect of duration of residence, an important determinant of language fluency. But it remains large and highly significant.

The analysis demonstrates that spoken dominant-language proficiency is an important determinant of earnings and presumably of other measures of economic success among immigrants. This suggests the importance of selecting immigrants who have or who can be expected quickly to acquire this proficiency, if the successful economic adjustment of immigrants is an important policy objective.

Canadian immigration policy explicitly recognizes the importance of this issue by awarding points in their point system for English or French fluency. Current U.S. immigration policy ignores language skills. Even the language requirements for illegal aliens to obtain permanent amnesty under the provisions of the 1986 Immigration Reform and Control Act are so meaningless as to be useless.

The analysis demonstrates the potential counterproductive nature of efforts to shelter immigrants from the economic consequences of inadequate proficiency. Immigrants respond positively to the economic incentives for fluency, thereby making the investment and becoming fluent. The analysis also demonstrates the importance of schooling, age at immigration, and other variables in determining fluency. These findings need to be explicitly recognized in immigration policy and in programs to facilitate immigrant adjustment. Again, the Canadians seem to have done a better job in this regard than the Americans.

This study also generates recommendations regarding the questions asked in the census. The language-related questions in the 1980 U.S. Census are superior to those in the 1981 Canadian Census. In the U.S. data, individuals who speak a language other than or in addition to English at home are asked to report the non-English language and the degree of their spoken fluency in English on a four-point scale (very well, well, not well, not at all). In the Canadian data, however, only those who cannot carry on a conversation in English or French are identified; these are the equivalent of the "not at all" English speakers in the United States. Furthermore, instead of the long list of non-English languages and countries of birth as provided in the U.S. data, the Canadian Census data permit the specific identification of only a handful. This coarseness in the Canadian data hampers the analysis of language. Both countries are repeating their language questions in the 1990–1991 censuses.

On a final note, the knowledge that dominant-language skills are very important for the economic success of immigrants for two countries with different immigration policies suggests the fundamental role of language capital in the labor market. In general, language capital is too obvious to be noticed. Immigration research highlights its role. This research also suggests that even among the native born, fluency is important, and degrees of fluency, not discerned in current data, may be important determinants of economic attainment.

Appendix 4A: The Census Language Questions

UNITED STATES: 1980 CENSUS

13a. Does this person speak a language other than English at home?

○ Yes ○ No, only speaks English (skip to 14)

b. What is this language?
(For example—Chinese, Italian, Spanish, etc.)

c. How well does this person speak English?

○ Very well ○ Not well
○ Well ○ Not at all

Note: The respondents were instructed to report "yes" to Q.13a if a language other than English is spoken at home, even if English is spoken more frequently than the other language. Those who speak only English at home include those who may speak another language at school, work, or elsewhere, but not at home, and those whose use of another language at home is limited to a few expressions or slang.

Those respondents speaking two or more non-English languages at home were asked to report the language spoken most often, or if this could not be determined, the first language learned. The write-in entries were coded into 387 language categories.

Source: US Bureau of the Census, *Census of Population and Housing, 1980: Public Use Microdata Sample, Technical Documentation* (Washington, D.C.: 1983), pp. K26 and K65.

CANADA: 1981 CENSUS

Question 12.
What is the language you first learned in childhood and still understand?

|____| English

|____| French

|____| German

|____| Italian

|____| Ukrainian

|____| Other (specify)

Question 18.
What language do you yourself speak at home now? (If more than one language, which language do you speak most often?)
Mark one box only

|____| English

|____| French

|____| German

|____| Italian

|____| Ukrainian

|____| Other (specify)

Question 19.
Can you speak English or French well enough to conduct a conversation?
Mark one box only

|____| English only

|____| French only

|____| Both English and French

|____| Neither English nor French

Note: The responses to "Other," specified in Q.12 and Q.18, were coded and reported in the Household/Family File as Chinese and Other, whereas Chinese, Greek, Netherlandic languages, Polish, and Portuguese are identified as separate languages in the Individual File.

Source: Form 3: Individual Census Questionnaire. Statistics Canada, *Summary Guide, Total Population,* Catalogue No. 99–902.

Appendix 4B: Definitions of the variables

The variables used in the analysis are defined below. Mnemonic names are also listed where relevant.

Analysis of 1980 U.S. Census of Population

Definition of Population: Foreign-born men aged twenty-five to sixty-four who worked during 1979.

Earnings (LNEARN): The natural logarithm of the sum of wage or salary income and self-employment income (either non-farm or farm). Income data refer to 1979.

Weeks Worked (LNWW): The natural logarithm of the number of weeks the respondent worked in 1979.

Years of Education (EDUC): This variable records the total years of full-time education.

Years of Experience (EXP): This is computed as age minus years of education minus 5 (that is, EXP = AGE – EDUC – 5). A quadratic specification is used.

Years since Migration (YSM): The categorical Census information on year of migration is converted to a continuous measure using the following values: 1975–1980 = 2 years, 1970–1974 = 7 years, 1965–1969 = 12 years, 1960–1964 = 17 years, 1951–1959 = 24.5 years, prior to 1950 = 40 years. A quadratic specification is used for this variable.

Birthplace (BIRTH): A number of birthplace regions were considered in the analyses: Britain, Ireland, Other Europe, Canada, West Indies, Mexico, Cuba, Other America, China, the Philippines, Vietnam, South Asia (which comprises the regions of British influence, namely India, Pakistan, Sri Lanka, Bangladesh, Bhutan and Nepal), Other Asia (for example, Korea and Japan), the Middle East, Africa, and Not reported. For the study of language proficiency, immigrants from Britain, Ireland, Canada, Australia, New Zealand, and the West Indies comprise the omitted English-speaking category, whereas for the study of earnings, the omitted category is restricted to immigrants from Britain.

English Language Proficiency (GOODENG): GOODENG is set to one for individuals who speak only English at home, or if a language other than English is spoken in the home, who speak English either "very well" or "well." The GOODENG variable is set to zero where a language other than English is spoken in the home and the respondent speaks English either "not well" or "not at all."

Citizenship (CITIZEN): This is a dichotomous variable, set to one for individuals who were either naturalized citizens or were born abroad of American parents.

Minority Group Concentration (CONC): Each respondent is assigned a measure equal to the percentage of the population aged eighteen to sixty-four in the state in which he lives, which reports the same non-English language group as the respondent. In the construction of this variable only the twenty largest nationwide language groups are considered. In descending order, these are: Spanish, Italian, German, French, Polish, Chinese, Tagalog, Greek, Portuguese, Japanese, Yiddish, Korean, Arabic, Vietnamese, Hungarian, Russian, Dutch, Hindi, Ukrainian, Czech. These constitute 92 percent of all valid responses. Representation in the other language groups is so small numerically that the proportions are approximately zero, and this value is assigned. Those who reported only English are also assigned a zero value. Appendix Table 4B.3 presents data on the percentage representation in the 8 largest language groups for each state.

Marital Status (MARRIED): This is a binary variable that distinguishes between individuals who are currently married, spouse present (equal to 1), and all other marital states.

Married Overseas (FORMAR): This variable is defined only for the foreign-born who have been married only once. It is constructed from information on age at first marriage and age at arrival in the United States. Individuals currently in their first marriage for whom age at first marriage is lower than age at arrival in the United States are assumed to have married in the country of origin. The variable is zero for all other individuals.

Children: Three variables are included in the estimating equations. The first records whether one or more children aged younger than six years were living in the family, and there were no older children. The second records whether one or more children aged between six and seventeen years, inclusive, were living in the family, and there were no younger children. The third variable records the presence of children aged younger than six years and between six and seventeen years.

Veteran Status (VETSTAT): This is a dichotomous variable, set to one where the respondent is a veteran of the U.S. Armed Forces; otherwise it is set to zero.

Location: The two location variables record residence of a rural area (Rural) or of the South Atlantic, East-South Central, or West-South Central geographic divisions (South). These variables are not mutually exclusive.

Race: This is a dichotomous variable, set to one if the individual is Black, and set to zero for all other racial groups (White, Asian and Pacific Islander groups, other groups).

Note: All variables for the United States are dichotomous except earnings, education, total experience, duration in the destination, weeks worked, and minority-language-group concentration. The means and standard deviations of the variables used in the analyses for the United States are reported in Appendix Table 4B.1.

Analysis of 1981 Census of Canada

Definition of Population: Foreign-born men aged twenty-five to sixty-four who worked during 1980.

Appendix Table 4B.1 Means and standard deviations of variables by region of origin for adult foreign-born men, United States, 1980

			Country of Origin			
	Total Sample		*English-speaking*		*Non-English speaking*	
	Mean	*Standard deviation*	*Mean*	*Standard deviation*	*Mean*	*Standard deviation*
Education	11.981	4.949	13.174	3.598	11.792	5.104
Age	41.108	11.007	44.049	11.493	40.539	10.851
Experience	24.038	12.539	25.876	12.680	23.748	12.492
YSM	15.751	11.994	20.033	13.213	15.074	11.647
Married	0.807	0.394	0.802	0.398	0.808	0.394
Married overseas	0.360	0.480	0.330	0.470	0.365	0.481
Child <6 only	0.144	0.351	0.101	0.301	0.150	0.357
Child 6–17 only	0.277	0.448	0.294	0.456	0.275	0.446
Children <6 & 6–17	0.137	0.344	0.085	0.279	0.145	0.352
Veteran	0.167	0.373	0.237	0.425	0.156	0.363
Rural location	0.078	0.269	0.125	0.331	0.071	0.257
South	0.194	0.396	0.163	0.369	0.199	0.399
Minority concentration	3.808	5.781	0.262	1.305	4.368	6.012
Citizenship	0.482	0.500	0.540	0.498	0.473	0.499
Birthplace						
Britain	0.039	0.194	0.292	0.455	–	–
Canada	0.060	0.237	0.437	0.496	–	–
Ireland	0.012	0.108	0.086	0.281	–	–
West Indies	0.025	0.157	0.185	0.388	–	–
Europe	0.278	0.448	–	–	0.322	0.467
Vietnam	0.010	0.101	–	–	0.012	0.109
Philippines	0.037	0.188	–	–	0.042	0.202
China	0.040	0.196	–	–	0.046	0.210
South Asia	0.031	0.174	–	–	0.036	0.187

(Continued Overleaf)

Appendix Table 4B.1 Continued

			Country of Origin			
	Total Sample		*English-speaking*		*Non-English speaking*	
	Mean	*Standard deviation*	*Mean*	*Standard deviation*	*Mean*	*Standard deviation*
Other Asia	0.049	0.216	–	–	0.057	0.231
Mexico	0.174	0.379	–	–	0.201	0.401
Cuba	0.051	0.220	–	–	0.059	0.236
Other America	0.097	0.296	–	–	0.112	0.315
Africa	0.021	0.143	–	–	0.024	0.153
Middle East	0.025	0.156	–	–	0.029	0.167
Not reported	0.051	0.220	–	–	0.059	0.236
Earnings	17,279	16,559	21,362	19,161	16,633	16,015
Log (earnings)	9.391	0.991	9.639	0.939	9.351	0.994
Weeks worked	46.405	10.893	47.468	10.005	46.237	11.018
Log (weeks worked)	3.779	0.431	3.811	0.399	3.774	0.436
GOODENG	0.797	0.402	0.992	0.088	0.766	0.423
Sample size		32,255		4,405		27,850

Notes: The English-speaking regions include Britain, Canada, Ireland, the British West Indies.

Source: 1980 Census of Population, Public Use Sample, 1/100 Sample of the foreign born.

Appendix Table 4B.2 Means and standard deviations of variables by region of origin for adult foreign-born men, Canada, 1981

			Country of Origin			
	Total Sample		*Dominant language*		*Nondominant language*	
	Mean	*Standard deviation*	*Mean*	*Standard deviation*	*Mean*	*Standard deviation*
Education	11.689	3.851	12.950	3.220	11.090	3.979
Age	42.645	10.542	42.724	10.738	42.608	10.447
Experience	25.956	12.035	24.773	11.706	26.518	12.149
YSM	19.452	10.626	19.725	11.344	19.323	10.266
Married	0.827	0.378	0.814	0.389	0.833	0.373
Married overseas	0.272	0.445	0.280	0.449	0.269	0.443
Child <6 only	0.129	0.335	0.108	0.311	0.138	0.345
Child 6–17 only	0.430	0.495	0.410	0.492	0.444	0.497
Children <6 & 6–17	0.112	0.315	0.091	0.288	0.122	0.327
Metropolitan (CMA)	0.744	0.436	0.694	0.461	0.768	0.422
Atlantic province	0.021	0.145	0.041	0.198	0.012	0.110
Prairie provinces	0.139	0.346	0.134	0.341	0.142	0.349
Quebec	0.143	0.350	0.105	0.306	0.161	0.367
British Columbia	0.159	0.365	0.181	0.385	0.148	0.355
Minority concentration	0.540	1.395	0.009	0.174	0.793	1.631
Citizenship	0.743	0.437	0.679	0.467	0.773	0.419
Birthplace						
Britain	0.200	0.400	0.621	0.485	–	–
United States	0.056	0.230	0.174	0.379	–	–
Ireland	0.005	0.073	0.017	0.128	–	–
West Indies	0.044	0.206	0.137	0.344	–	–
France	0.016	0.127	0.051	0.220	–	–
W. Europe	0.137	0.344	–	–	0.202	0.401
E. Europe	0.094	0.292	–	–	0.139	0.345

(Continued Overleaf)

Appendix Table 4B.2 Continued

			Country of Origin			
	Total Sample		Dominant language		Nondominant language	
	Mean	*Standard deviation*	*Mean*	*Standard deviation*	*Mean*	*Standard deviation*
S. Europe	0.232	0.422	–	–	0.342	0.475
Chinese Asia	0.048	0.213	–	–	0.070	0.256
Other Asia	0.086	0.280	–	–	0.127	0.333
Africa	0.030	0.170	–	–	0.044	0.204
South & Central America	0.020	0.141	–	–	0.030	0.170
Other	0.032	0.175	–	–	0.047	0.211
Earnings	20,218	13,391	22,797	14,427	18,991	12,687
Log (earnings)	9.595	1.096	9.752	1.009	9.521	1.128
Weeks worked	46.472	10.446	47.471	9.607	45.996	10.789
Log (weeks worked)	3.789	0.391	3.819	0.352	3.774	0.408
GOODLANG	0.972	0.164	1.000	0.016	0.959	0.198
Sample size		23,741		7,649		16,092

Notes: The dominant-language regions of origin include Britain, the United States, Ireland, British West Indies, and France.

Source: 1981 Census of Canada, Public Use sample, Individual File, 1/50 sample of the foreign born. The children-variables were derived from the Household and Family File.

Appendix Table 4B.3 Representation of major minority-language groups by state, United States, 1980 (percent)

State	*Spanish*	*Italian*	*German*	*French*	*Polish*	*Chinese*	*Tagalog*	*Greek*
Alabama	0.40	0.03	0.18	0.38	0.00	0.05	0.03	0.10
Alaska-Hawaii	1.72	0.07	0.79	0.22	0.14	1.08	3.66	0.00
Arizona	17.82	0.20	0.43	0.18	0.18	0.23	0.00	0.13
Arkansas	0.30	0.00	0.17	0.08	0.04	0.04	0.04	0.00
California	13.97	0.47	0.71	0.38	0.08	1.08	0.92	0.13
Colorado	6.69	0.27	1.09	0.61	0.10	0.27	0.00	0.00
Connecticut	3.35	2.90	0.73	2.33	1.37	0.13	0.10	0.32
District of Columbia	2.61	0.33	0.65	1.79	0.00	0.65	0.16	0.65
Florida	8.48	0.66	0.83	0.71	0.22	0.06	0.10	0.15
Georgia	0.68	0.11	0.34	0.41	0.05	0.02	0.00	0.13
Illinois	5.02	0.80	0.87	0.31	1.20	0.24	0.20	0.42
Indiana	1.41	0.04	0.75	0.26	0.33	0.04	0.02	0.07
Kentucky	0.47	0.11	0.28	0.28	0.00	0.00	0.03	0.03
Louisiana	1.14	0.26	0.19	6.68	0.02	0.02	0.02	0.07
Maine	0.55	0.20	0.20	6.52	0.35	0.08	0.04	0.12
Maryland	1.18	0.37	0.52	0.45	0.33	0.14	0.17	0.35
Massachusetts	1.91	1.94	0.39	2.90	0.78	0.41	0.02	0.36
Michigan	1.26	0.46	0.63	0.31	0.95	0.12	0.06	0.12
Minnesota	0.76	0.08	1.30	0.24	0.10	0.06	0.04	0.08
Mississippi	0.52	0.00	0.12	0.32	0.00	0.08	0.00	0.00
Montana	2.28	0.18	0.87	0.18	0.05	0.09	0.00	0.00
New Jersey	5.97	2.59	1.10	0.41	1.00	0.25	0.32	0.34
New York	8.12	3.19	0.80	0.94	0.78	0.73	0.15	0.58
North Carolina	0.97	0.07	0.29	0.51	0.03	0.08	0.05	0.12
Ohio	1.02	0.42	0.71	0.31	0.31	0.07	0.05	0.13
Oklahoma	1.04	0.06	0.36	0.26	0.03	0.16	0.00	0.13
Oregon	1.85	0.18	0.78	0.33	0.07	0.22	0.04	0.07
Pennsylvania	1.22	1.10	0.84	0.30	0.64	0.09	0.00	0.17

(*Continued Overleaf*)

Appendix Table 4B.3 Continued

State	*Spanish*	*Italian*	*German*	*French*	*Polish*	*Chinese*	*Tagalog*	*Greek*
South Carolina	0.45	0.13	0.32	0.48	0.03	0.00	0.00	0.06
Tennessee	0.61	0.00	0.33	0.35	0.00	0.07	0.00	0.02
Texas	18.23	0.08	0.57	0.34	0.07	0.16	0.10	0.04
Utah-Nevada	2.90	0.14	0.68	0.50	0.05	0.05	0.27	0.05
Virginia	0.78	0.20	0.43	0.56	0.02	0.06	0.22	0.19
Washington	2.00	0.17	0.92	0.39	0.00	0.36	0.46	0.00
West Virginia	0.51	0.25	0.35	0.10	0.20	0.05	0.00	0.00
Wisconsin	1.37	0.38	1.50	0.11	0.80	0.04	0.04	0.00

Notes: These percentages are defined for the population aged 18 to 64 in each state or group of states. Maine includes Maine, New Hampshire, and Vermont; Massachusetts includes Massachusetts and Rhode Island; Minnesota includes Minnesota, Iowa, Missouri, Kansas, Nebraska, South Dakota, and North Dakota; Maryland includes Maryland and Delaware; Montana includes Montana, Idaho, and Wyoming; Arizona includes Arizona and New Mexico.

Source: 1980 Census of Population, Public Use Sample, C sample, 1/1,000 sample of the population.

Appendix Table 4B.4 Representation of major minority-language groups by region, Canada, 1981 (percent)

Region	*Chinese*	*German*	*Greek*	*Italian*	*Netherlandic*	*Polish*	*Portuguese*	*Ukrainian*
Newfoundland	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nova Scotia								
Halifax	0.1	0.2	0.3	0.1	0.0	0.0	0.1	0.0
Other	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
New Brunswick	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Quebec								
Quebec City	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0
Montreal	0.4	0.2	1.3	3.6	0.0	0.2	0.7	0.2
Other	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0
Ontario								
Ottawa-Hull	0.7	0.3	0.1	1.3	0.1	0.2	0.4	0.1
Toronto	2.1	0.6	1.3	5.9	0.1	0.7	2.1	0.5
Hamilton	0.5	0.8	0.4	3.8	0.1	0.7	0.8	0.5
St. Catherines	0.2	0.8	0.2	3.5	0.2	0.8	0.0	0.6
Kitchener	0.5	2.8	0.3	0.4	0.2	0.3	3.0	0.2
London	0.6	0.3	0.8	1.2	0.1	0.2	1.0	0.2
Other	0.3	0.6	0.1	1.2	0.2	0.3	0.3	0.2
Manitoba								
Winnipeg	0.7	1.4	0.2	0.6	0.1	0.5	0.9	1.4
Other	0.2	5.0	0.0	0.1	0.1	0.1	0.1	2.1
Saskatchewan	0.4	1.1	0.1	0.1	0.1	0.0	0.0	1.0
Alberta								
Calgary	1.9	0.6	0.2	0.8	0.3	0.1	0.2	0.2
Edmonton	1.6	0.8	0.0	0.7	0.2	0.4	0.3	0.8
Other	0.3	1.3	0.0	0.1	0.2	0.1	0.0	0.7
British Columbia								
Vancouver	4.5	0.9	0.3	0.9	0.2	0.1	0.4	0.1
Other	0.5	0.8	0.1	0.4	0.2	0.0	0.2	0.1
Pr. Edward Is.	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Source: 1981 Census of Canada, 1/50 Public Use Sample, Individual File.

Earnings (LNEARN): The natural logarithm of the sum of wage or salary income and self-employment income. Income data refer to 1980.

Weeks Worked (LNWW): The natural logarithm of the number of weeks worked by the respondent in 1980.

Years of Education: This variable records the total years of full-time education.

Years of Experience (EXP): This is computed as age minus years of education minus 5 (EXP = AGE – EDUC – 5). A quadratic specification is employed.

Years since Migration (YSM): The census information on year of arrival in Canada is recorded in individual years between 1971 and 1980, and in intervals of varying length for pre-1971 arrivals. The categorical information was converted to a continuous measure of years since migration using the following values: 1967–70 = 12.5 years, 1966 = 15 years, 1961–65 = 18 years, 1956–60 = 23 years, 1946–55 = 30.5 years, and pre-1946 = 42 years. A quadratic specification is employed.

Birthplace (BIRTH): Previous studies (Meng, "Earnings of Canadian Immigrants," and Chiswick and Miller, "Earnings in Canada") have proposed a range of birthplace groupings for inclusion in analyses of earnings. The present study uses a set of birthplace regions that facilitates comparisons with the study of the U.S. labor market. The following birthplace groups are recognized in this study: Britain (including Northern Ireland), Republic of Ireland, the United States, France, Western Europe (which includes Belgium, Luxembourg, West Germany, the Netherlands, and Austria), Southern Europe (which includes Greece, Italy, Portugal and Yugoslavia), Eastern Europe (which includes Hungary, Poland, the USSR, and Czechoslovakia), Chinese Asia, Other Asia, South and Central America, English-origin West Indies, Africa, and Other. These regions are identified based on the birthplace, ethnic origin, and mother-tongue information in the Census Files. Mother tongue is used to separate immigrants from South and Central America from English-origin immigrants from the Caribbean. Ethnic origin is used to allocate some of the responses to birthplace coded as "Other Europe" to the categories of Northern and Western Europe, Southern Europe, and Eastern Europe, and also to distinguish Chinese Asia from other regions of Asia. For the study of dominant-language proficiency, immigrants from Britain, Ireland, the United States, and the British West Indies make up the omitted dominant-language-speaking category, while for the study of earnings, the omitted category is restricted to immigrants from Britain.

Dominant-Language Proficiency (GOODLANG): Individuals who reported that they could speak English or French well enough to conduct a conversation were classified as proficient in the dominant language.

Minority-Group Concentration (CONC): Each respondent is assigned a measure equal to the percentage of the population aged eighteen to sixty-four in the region (defined using information on residence in a Census Metropolitan Area and province of residence) in which he lives that reports the same home language as the respondent. The nondominant-language groups Chinese, German, Italian, and Ukrainian are identified on the Household and Family File, and Chinese, German, Italian, Ukrainian, Greek, Netherlandic languages, Polish, and Portuguese are identified on the Individual File. The first four language groups constitute 46 percent of nondominant-language responses, and the final four a further 20 percent. Appendix Table 4B.4 presents data on the percentage in each language group for the twenty-three regions distinguished in the construction of the variable.

Marital Status (MARRIED): This is a binary variable that distinguishes between individuals who are married, spouse present (equal to 1), and all other marital states.

Married Overseas (FORMAR): This variable is computed from information on age at first marriage and age at arrival in Canada. Individuals for whom age at first marriage is lower than age at arrival in Canada, and for whom the date of marriage corresponds to that of their spouses, are assumed to have married their present spouses in the country of origin. The variable is unity for this group and zero for all others.

Location: Two location variables are used in the study. The first records province of residence. This information was grouped as follows: Ontario, Atlantic provinces (Newfoundland, Nova Scotia, New Brunswick, Prince Edward Island), Quebec, Prairie provinces (Manitoba, Saskatchewan, Alberta), and British Columbia. The second locality variable records the size of the place of residence. Here, individuals residing in Census Metropolitan Areas (defined as places having 100,000 or more in population) are distinguished from other individuals.

Citizenship (CITIZEN): Individuals who hold Canadian citizenship are distinguished from immigrants who have not yet become Canadian citizens. This information is available only from the Individual File.

Note: All variables for Canada are dichotomous except earnings, education, total experience, duration in the destination, weeks worked, and minority-language-group concentration. The means and standard deviations of the variables used in the analyses for Canada are reported in Appendix Table 4B.2.

Acknowledgements

We are grateful to participants in workshops at Queen's University and the University of Illinois at Chicago for helpful comments, as well as to Chris Robinson, Houston Stokes, and our colleagues on the project. We appreciate the financial support for this project provided by grants from the William H. Donner Foundation, the Sloan Foundation, and the Embassy of Canada, Washington, D.C.

Notes

1 The biblical account of the tower of Babel is instructive (Genesis, chapter xi, v. 1–9). According to tradition all people spoke the same language and gathered at Babel to work together to construct a tower to reach heaven. Offended by this, the Lord inflicted on the people a diversity of languages, thereby increasing transaction costs and halting the progress of the tower.
2 The language questions contained in the census questionnaires are reproduced in Appendix 4A.
3 Calvin Veltman, *Language Shift in the United States* (Berlin: Mouton Publishers, 1983); Calvin Veltman, "Modelling the Language Process of Hispanic Immigrants," *International Migration Review*, vol. 22 (1988) pp. 545–62; Barry R. Chiswick, "Speaking, Reading, and Earnings among Low-skilled Immigrants," *Journal of Labor Economics*, vol. 9 (April 1991), pp. 149–70.
4 The coauthors of this chapter, for example, disagree on the spelling of labor or labour.
5 Veltman, "Modelling the Language Process."
6 Chiswick, "Speaking, Reading, and Earnings."
7 John DeVries and Frank E. Vallee, *Language Use in Canada* (Ottawa: Statistics Canada, 1980).
8 Chiswick, "Speaking, Reading, and Earnings," pp. 149–70.
9 Veltman, *Language Shift*.
10 Perhaps the classic example is the myth, perhaps not too far from reality, that Jewish immigrant parents in Israel learned Hebrew from their children.
11 The equations were also estimated using a logit model. The signs and significance of the estimates were broadly similar for the two methods of estimation.
Diagnostic testing using the Breusch-Pagan test suggested that the residuals were not homoskedastic; T. S. Breusch and A. R. Pagan, "A Simple Test for Heteroskedasticity and Random Coeffcient Variation," *Econometrica*, vol. 47 (1979), pp. 437–41. All *t* values for the linear probability model have been calculated using White's heteroskedasticity-consistent covariance matrix estimator; Halbert White, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, vol. 48 (1980), pp. 817–38.
12 Tests were conducted to determine whether the relationship between English-language fluency and educational attainment was nonlinear. We did not attain any gain in economic insights from attempting to capture this nonlinearity through the use of complex functional forms for education (for example, higher order polynomials, linear splines, or a large number of dummy variables). Accordingly, a simple linear education variable is used.
13 Equations were also estimated with a second-degree polynomial in age. The squared term was not significant at conventional levels, however.
14 Veltman, "Modelling the Language Process."
15 For older cohorts of immigrants (pre-1945), there is a negative partial effect of duration of residence on language skills. Most pre-1945 immigrants in these data arrived during the 1930s, and a disproportionate number were young-adult refugees who may not have been self-selected for acquiring U.S.-specific skills and may have anticipated returning to Europe after the fall of fascism.
16 Veltman, *Language Shift*.
17 Those not fluent in English may have access to a much smaller marriage market and may be less likely to marry. This reverse causation argument would be more compelling for numerically small groups. Yet the same effect appears among a very large group, Mexican immigrants.
18 There is some degree of endogeneity in the veteran status variable, although this would be less intense during the period of conscription.

19 Gilles Grenier and François Vaillancourt, "An Economic Perspective on Learning a Second Languages," *Journal of Multilingual Development*, vol. 4 (1983), pp. 471–83.

20 The children variables record the presence in the household of children younger than eighteen years of age at the time of the census enumeration. Ideally we would like to use information on the number and ages of all children ever in the household in the United States, and not simply those currently living at home.

21 Barry R. Chiswick, "Soviet Jews in the United States: An Analysis of Their Linguistic and Economic Adjustment," *International Migration Review*, vol. 27 (1993), pp. 260–85.

22 Spanish is spoken in the home by 10 percent or more of the population in California, Texas, Arizona, and New Mexico. For further information on home language usage in the various states, see Appendix Table 4B.3.

23 Chiswick, "Speaking, Reading, and Earnings."

24 Barry R. Chiswick, "The Effect of Americanization on the Earnings of Foreign-born Men," *Journal of Political Economy*, vol. 85 (1978) pp. 897–921.

25 Among immigrants from Mexico, the Anglicization process continues for about the same period as reported for the Table 4.1 results. This suggests that the differences in conclusions drawn from Table 4.1 and from Veltman, "Modelling the Language Process," derive mainly from the different methodologies employed. The Table 4.3 finding is similar to that reported by Grenier and Vaillancourt, "An Economic Perspective," also on the basis of a multivariate analysis.

26 The minority-language coefficient is –0.014 in Table 4.1. In Table 4.3 the within-birthplace region estimates of the minority-language coefficients are within two standard errors of that estimate for Other Asia, Mexico, Other America, and Africa. While this is not a valid statistical test because the coefficients are not estimated independently, it does strengthen the point that the Table 4.1 minority-language effect is more than merely a proxy for country of origin.

27 The census reports all Chinese dialects as one category, a practice followed here. Although they share a common written language, the differences in the spoken language among the Chinese dialects is so great that it is as if they were different languages. The minority-language variable is positive and significant for those from South Asia, but fewer than 2 percent of this sample (twenty cases) are not fluent in English.

28 It is, however, positive and significant for the small sample of immigrants from Africa. The reasons for this unexpected result are as yet unclear. The small sample of Africans is heterogenous: 33 percent are white North Africans, 19 percent are white South Africans, 38 percent are black, and 10 percent are other Africans.

29 When only those who report that they cannot speak English at all are considered as lacking in English fluency, the fluency rate increases to 94 percent in the U.S. census data.

30 There are nine factors in the selection procedure, and the maximum points they carry are: education (twelve points), special vocational preparation (fifteen), experience (eight), occupation (ten), arranged employment (ten), demographic factors (ten), age (ten), knowledge of official languages (fifteen), and personal suitability (ten). Thus, of the 100 points in the assessment procedure, fifteen are allocated to knowledge (speaking, reading, and writing) of the official languages. The threshold number of points varies by category of immigrant. It is seventy for independent workers, fifty-five for assisted relatives, and twenty-five for entrepreneurs. Further details can be obtained from the *Immigration Manual: Selection and Control, section 4.08, Employment and Immigration Canada*.

31 Direct information on the language skills of children is available in the family file, and it is possible that this would provide the opportunity to ascertain more definitely whether there is a relationship between the use of the dominant language in

the home by children and adult dominant language. When a variable for children's dominant language use within the home was included in the estimating equation, it was highly significant. This suggests that adults' dominant-language fluency is higher in families where children speak the dominant language. In view of the statistical insignificance of the children variables, however, the direction of causation here is very problematic, and little weight can be attached to this result.

32 There is, of course, possible endogeneity; those not fluent in the dominant language may be less successful in the marriage market for dominant-language speakers. As was found in the United States, however, premigration marriage has an adverse impact on fluency larger than the impact of postmigration marriage among those country groups with the largest ethnic marriage market in the destination (that is, among those most likely to find in the destination a spouse who also speaks the same nondominant language).

33 This positive effect arises even though schooling and language fluency are alternative sources for points in the immigrant rationing system.

34 When the equation was reestimated using a logit model, this anomaly disappeared.

35 DeVries and Vallee, *Language Use in Canada*.

36 This compares with the impact of –0.010 (t=4.14) attributed to this variable on the basis of analysis of the Household-Family File. Using more detailed information in the construction of this variable, therefore, appears to be associated with a stronger estimated impact.

37 As was noted previously, if only those who reported that they spoke English "not at all" are considered as lacking English-language fluency, the U.S. fluency rate is 95 percent.

38 The Heckman technique is developed in James J. Heckman, "Sample Selection Bias as a Specification Error," *Econometrica*, vol. 47 (1979), pp. 153–62.

These methods have recently been investigated in some depth in the union wages effects literature; see, for example, the analyses in John T. Addison and Pedro Portugal, "The Endogeneity of Union Status and the Application of the Hausman Test," *Journal of Labor Research*, vol. 10 (1989) pp. 437–41; Chris Robinson, "The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Models," *Journal of Political Economy*, vol. 97 (1989) pp. 639–67; and Chris Robinson, "Union Endogeneity and Self-Selection," *Journal of Labor Economics*, vol. 7 (1989) pp. 106–12. Both methods are used in this chapter.

A simultaneous equations system in which earnings and languages are both endogenous cannot be estimated because of the absence of instruments that enter an equation for earnings but not for language. While weeks worked might seem to be one such variable, it largely standardizes annual earnings for the amount of time worked. The citizen and race variables are also inappropriate, as citizenship may be determined endogenously with language skills, and the race variable is highly collinear with the country of birth variables.

39 J. A. Hausman, "Specification Tests in Econometrics," *Econometrica*, vol. 46 (1978), pp. 1251–71.

40 Chiswick, "The Effect of Americanization"; and Barry R. Chiswick and Paul W. Miller, "Earnings in Canada: The Roles of Immigrant Generation, French Ethnicity, and Language," *Research in Population Economics*, vol. 6 (1988) pp. 183–224.

41 Price V. Fishback and Joseph V. Terza, "Are Estimates of Sex Discrimination by Employers Robust? The Use of Never Marrieds," *Economic Inquiry*, vol. 27 (1980) pp. 271–85.

42 Recall also that duration of residence is an important determinant of language proficiency.

43 Michael G. Abbott and Charles M. Beach, "Immigrant Earnings Differentials and Cohort Effects in Canada," Institute for Economic Research, Queen's University,

discussion paper no. 705 (1987); Chiswick, "Speaking, Reading, and Earnings"; Chiswick and Miller, "Earnings in Canada."

One study (Walter M. McManus, William Gould, and Finis Welch, "Earnings of Hispanic Men: The Role of English-Language Proficiency," *Journal of Labor Economics*, vol. 1 (1983), pp. 101–30), reports that including language in the earnings function eliminates the Hispanic-non-Hispanic earnings differential. It has been shown, however, that this result is a consequence of a specification error in the equation for predicting English-language fluency (see Chiswick, "Speaking, Reading, and Earnings").

44 See the analysis in Chris Robinson, "Language Choice: The Distribution of Language Skills in Earnings in a Dual-Language Economy," *Research in Labor Economics*, vol. 9 (1988), pp. 53–90.

This result is consistent with the union wage effects literature where, according to Robinson, there is "substantial evidence of a consistent rise in the union differential relative to OLS estimates when the endogeneity of union status is addressed by the instrumental variables or inverse Mills ratio method." Robinson, "Joint Determination of Union Status," p. 659.

45 The selectivity correction factors are computed for logit estimates of the language attainment model developed in the second section. The method applied is outlined in Lung-Fei Lee, "Generalized Econometric Models with Selectivity," *Econometrica*, vol. 51 (1983) pp. 507–12.

46 Robinson, "Joint Determination of Union Status."

47 This finding concerning the assumption of exogeneity of the language-fluency variable is supported by results of an alternative test based on Hausman "Specification Tests," as outlined in Addison and Portugal, "Edogeneity of Union Status." For this procedure, the compared value of the test statistic was 17.687, which exceeds the critical F value at the 5 percent level of significance.

48 Chiswick and Miller, "Earnings in Canada," pp. 183–224, and Ronald Meng, "The Earnings of Canadian Immigrants and Native-born Males," *Applied Economics*, vol. 19 (1987), pp. 1107–19.

49 Differences between the Table 4.9 results and those presented in Chiswick and Miller, "Earnings in Canada," are due to different treatment in the analyses of workers with nonpositive incomes. In Chiswick and Miller these individuals were purged from the sample, and the results obtained are consistent with the Canadian literature. In the present analysis, this small group of workers (2 percent of the sample) are assigned $100 in earnings. The results are very similar to those reported in the U.S. literature where the same procedure is used.

50 Chiswick and Miller, "Earnings in Canada."

51 See Chiswick and Miller, "Earnings in Canada," where it is shown that this is largely a post-1971 phenomenon.

52 Abbott and Beach, "Immigrant Earnings Differentials"; Chiswick and Miller, "Earnings in Canada."

53 The estimates for the Addison and Portugal test differ from the instrumental variables estimates discussed previously in that a logit model is used to predict dominant-language fluency in preference to the linear probability model, and interaction terms with dominant-language proficiency are included in the estimating equation.

Bibliography

Abbott, Michael G., and Charles M. Beach. "Immigrant Earnings Differentials and Cohort Effects in Canada." Institute for Economic Research, Queen's University, discussion paper no. 705, 1987.

Addison, John T., and Pedro Portugal. "The Endogeneity of Union Status and the Application of the Hausman Test." *Journal of Labor Research* 10 (1989): 437–41.

Breusch, T. S., and A. R. Pagan. "A Simple Test for Heteroskedasticity and Random Coefficient Variation." *Econometrica* 47 (1979): 1287–94.

Chiswick, Barry R. "The Effect of Americanization on the Earnings of Foreign-born Men." *Journal of Political Economy* 85 (1978): 897–921.

—— "Soviet Jews in the United States: An Analysis of their Linguistic and Economic Adjustment." *International Migration Review* 27 (1993): 260–85.

—— "Speaking, Reading and Earnings among Low-skilled Immigrants." *Journal of Labor Economics* 9 (April 1989): 149–70.

Chiswick, Barry R., and Paul W. Miller. "Earnings in Canada: The Roles of Immigrant Generation, French Ethnicity, and Language." *Research in Population Economics* 6 (1988): 183–224.

DeVries, John, and Frank G. Vallee. *Language Use in Canada*. Ottawa: Statistics Canada, 1980.

Fishback, Price V., and Joseph V. Terza. "Are Estimates of Sex Discrimination by Employers Robust? The Use of Never Marrieds." *Economic Inquiry* 27 (1989): 271–85.

Greenwood, Michael J., and John M. McDowell. "The Factor Market Consequences of U.S. Immigration." *Journal of Economic Literature* 24 (1986): 1738–72.

Grenier, Gilles, and François Vaillancourt. "An Economic Perspective on Learning a Second Language." *Journal of Multilingual and Multicultural Development* 4 (1983): 471–83.

Hausman, J. A. "Specification Tests in Econometrics." *Econometrica* 46 (1978): 1251–71.

Heckman, James J. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1979): 153–62.

Lee, Lung-Fei. "Generalized Econometric Models with Selectivity." *Econometrica* 51 (1983): 507–12.

McManus, Walter, William Gould, and Finis Welch. "Earnings of Hispanic Men: The Role of English-Language Proficiency." *Journal of Labor Economics* 1 (1983): 101–30.

Meng, Ronald. "The Earnings of Canadian Immigrants and Native-born Males." *Applied Economics* 19: 1107–19.

Reimers, David M., and Harold Troper. "Canadian and American Immigration Policy since 1945", in Barry R. Chiswick (ed.) *Immigration, Language and Ethnicity: Canada and the United States*, pp. 15–54. The AEI Press, Washington, D.C.

Robinson, Chris. "The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Models." *Journal of Political Economy* 97 (1989): 639–67.

—— "Language Choice: The Distribution of Language Skills and Earnings in a Dual-Language Economy." *Research in Labor Economics* 9 (1988): 53–90.

—— "Union Endogeneity and Self-Selection." *Journal of Labor Economics* 7 (1989): 106–12.

Vaillancourt, François. "An Economic Perspective on Language and Public Policy in Canada and the United States", in Barry R. Chiswick (ed.) *Immigration, Language, and Ethnicity: Canada and the United States*, pp. 177–228. The AEI Press, Washington, D.C.

Veltman, Calvin. *Language Shift in the United States*. Berlin: Mouton Publishers, 1983.

—— "Modelling the Language Process of Hispanic Immigrants." *International Migration Review* 22 (1988): 545–62.

White, Halbert. "A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (1980): 817–38.

5 Parents and children talk

English language proficiency within immigrant families

With Yew Liang Lee

Introduction

The tradition in migration research has been to emphasize the individual as the decision maker. In this approach family membership is held implicitly or explicitly to be inconsequential in terms of explaining a given individual's behavior. There is, however, a stream of research emphasizing the role of the family in migration decisions.[1] Jacob Mincer (1978) argued that migration decisions are based on the net economic opportunities open to both the primary income earner (generally the husband), and the secondary income earner (often females). Migration to where the joint opportunities are better will occur, even if each spouse has better opportunities elsewhere. More recent research, for example, by Oded Stark (1991) and Timothy Hatton and Jeffrey Williamson (1997), has widened the scope of the family ties considered in migration decision making to be more consistent with the characteristics of recent international migration flows.

This theme of focusing on the family as the unit of analysis has been developed to explain some post-migration behavior. Thus, Michael Baker and Dwayne Benjamin (1997) develop a family investment model wherein one spouse (generally the wife) will work extensively in the immediate post-migration period in order to finance the human capital investment of another family member (generally the husband). Robert Birrell (1987) sees family decision making as the reason for the high levels of education of some children of immigrants of non-English speaking background in Australia.

A further area of research where there is a growing focus on the family as an influence on post-migration outcomes is the study of dominant language skills. Formal models of the dominant language acquisition process have been developed and tested (e.g., Albert Breton, 1978a, b; Barry Chiswick and Paul Miller, 1992, 1998, 2001; Thomsas Espenshade and Haishan Fu, 1997). In the early literature, these models examined adults (e.g., Chiswick and Miller, 1992, 1998, 2001) and the children of immigrants (e.g., Alejandro Portes and Richard Schauffler, 1994) as separate groups. Recently, Barry Chiswick, Yew Liang Lee and Paul Miller (2005) have provided a framework that enables the quantification of relationships in dominant language skill acquisition among members of the family unit (e.g., adult males, adult females, children). Focusing on spouses in immigrant families in Australia, they show that there is a positive correlation between the personal characteristics within migrating units. These commonalities combine with the similarity of the processes determining the English language skills of family members, and interactions among spouses within the household, to generate strong links between the destination language skills of spouses. In addition to these measured factors, there are other, unmeasured, factors (possibly motivation, aptitude for the learning of languages) that need to be considered, and which reinforce the tendency for the English-speaking skills of spouses to be similar.

The Chiswick, Lee and Miller (2005) study is limited in several regards. It is based on a relatively small survey and examined immigrants who had lived in Australia for only 5 or 6 months. It could therefore not examine the role of duration. Moreover, it focused exclusively on the proficiency of husbands and wives, ignoring the proficiency of children and the interactions between and among mothers, fathers, children and siblings.

This study advances on the previous streams of research by simultaneously considering the links among the dominant language skills of mothers, fathers, their children and sibling relationships among the children. It does this by treating the entire family as the unit of observation. The aims are primarily to establish links between and among the language skills of children and of their parents, and to ascertain whether these links vary according to offspring birth order and the age of the children, among other factors. The data under study are the unit record files from the 1996 Australian Census of Population and Housing, which provides a very large sample size and immigrants of all durations in the country.

The structure of this chapter is as follows. Section 1 provides a brief introduction to the literature analyzing interactions within immigrant families in the post-migration period. Section 2 contains an overview of the data. This includes a discussion of the limitations of using de facto family membership as the underlying categorization. Separate analyses are conducted in Section 3 examining the linkages between the language skills of spouses, and examining the linkages among the language skills of parents and children. Section 4 contains a summary and discussion.

1. Background

Walter Enloe and Philip Lewin (1987) argue that migration affects family members differently, and these effects vary according to the position within the family. Three sets of relationships within the family can be identified, namely the husband–wife relationship, the parent–child relationship and the relationship among siblings (see Wilfried Dumon, 1989).

The husband–wife relationship

The husband–wife relationship has been emphasized in recent studies of both migration decisions (e.g., Mincer, 1978) and post-migration behavior (e.g., Baker and Benjamin, 1997). Key elements of this relationship appear to have changed in recent years. For example, Efie Gavaki (1979), in describing the Greek family both in Greece and Canada, shows that the traditional patterns of family-gender roles have undergone considerable transition. In particular, the fathers/husbands' authority has been reduced, whereas mothers/wives' involvement in decision-making processes has increased. Presumably this transition reflects in large part the wider set of social and institutional changes that have occurred in many Western countries.

The interactions between husband and wife in the model of dominant language proficiency can be analysed using the investment in human capital framework outlined in Shoshana Grossbard-Shechtman and Shoshana Neuman (1991, 1993). There it is argued that one spouse may invest in the human capital of the other, and this will result in positive relationships between an individual's labor market standing, religious practices and other outcomes and the spouse's level of education. Grossbard-Shechtman and Neuman's (1993) study of religious practices is particularly apt, as it has a focus on in-marriage learning, where one spouse influences the religious practice of the other. The distinguishing feature of these models is the inclusion in the estimating equations of variables for the spouse (for example, for earnings, religious practices, spouse's help with career). A further feature is the inclusion of interaction terms between own schooling and spouse's schooling, in order to test if the two are complements or substitutes.[2]

The parent–child relationship

Children have greater exposure to the language and culture of the host country than do their parents. Among other factors, they receive intense exposure to the dominant language while at school. Being younger, they also are able to learn new languages quicker than their parents (see Michael Long, 1990; Elisabet Service and Fergus Clark, 1993). Thus, it is to be expected that children would acquire proficiency more rapidly than their parents.

Four main factors appear to impact on the relationship between children and their parents in the evolution of dominant language proficiency in the family. On the negative side, these are: (i) the children as interpreters factor, where children are encouraged to acquire dominant language skills so that they can help insulate their parents from the host country (see, for example, Suzanne Ziegler, 1977[3]); (ii) the desire among some parents to have their children learn/retain the language and culture of the origin country to enable communication with parents and grandparents, and possibly to facilitate visits or return migration; (iii) children lower mother's labor supply, which results in a lower exposure to the destination language and a smaller economic incentive for the mother to learn this language. On the positive side, (iv) the children as teachers factor, whereby children bring the dominant language into the home and encourage its use.[4]

Of the four sets of influence outlined above, only the last hypothesis, that is, "children as teachers", suggests a positive effect of children on parents' destination language proficiency; the others imply a negative effect. The "children as interpreters" and the "labor supply effects" hypotheses are more likely to depress the mother's language proficiency than that of the father's. If as argued by Dumon (1989) mothers are primarily responsible for the socialization of their children, the links between the characteristics of the mothers and their children's language skills will be stronger than that of the father. These effects on immigrant parents and their children's bilingualism may vary according to socio-economic factors. Portes and Schauffler (1994) note that better-educated parents tend to wish to transmit their mother tongue, but will also be more proficient in the destination language and make available more opportunities for their children to enter the destination's cultural mainstream.

The parent–child relationship can be further analyzed according to different periods in the life cycle.

Early childhood A number of studies (e.g., Bernhard Nauck, 1989) have shown that the parent–child relationship varies with socio-economic status. This may have implications for dominant language acquisition. For example, better-educated parents are more likely to choose formal childcare when their children are very young. This may result in greater exposure to the dominant language from an early age.

Pre-school and school-age children Immigrant children usually face a norm and value system at home which is different from the norm and value system they experience in the school systems of host countries. Moreover, Charles Hirschman (1994, p. 703) notes that "the youngest immigrants may also be more susceptible to peer pressures that are at odds with influences from the home". The impact of this inconsistency on dominant language fluency will depend on the extent to which the parents are oriented to adapting to the host country, and the extent to which the immigrant child interacts with other children of the same origin.

Adolescence The period of puberty is often viewed as the launching stage in the family with potential for conflict. Studies such as Betty Sung (1985) point out that immigrant parents often see their children's ambivalence and "novel ideas" as indication of disrespect and eventual rejection of their values and customs. David Haines, Dorothy Rutherford and Patrick Thomas (1981) note that disagreements between generations often arise as a result of children acculturating more rapidly than their parents to a society that places different values on many aspects of life, including the premium placed on youth. Josephine Naidoo and J. Campbell Davis (1988), for example, identify a generation gap between parents and teenage children as contributing to conflicts with regard to dating for adolescents. While this period is presumably the main period where bilingualism may turn to (practical) dominant language monolingualism, the empirical relevance of this suggestion does not appear to have been tested.

The relationship among siblings Research on the relationships among siblings is not as abundant as research on husband–wife or parent–child relationships. As shown by Judith Blake (1980), among others, the greater the number of siblings the smaller the interaction of any one child with the parents, and the greater the interaction with other children (siblings). Among the native born this shift from parental to children interaction would lower the accumulation of human capital relevant for the country of residence. The picture is less clear for immigrant children and may depend on birth order. The oldest child of immigrant parents may be at the greatest disadvantage, while the youngest child, other things the same, has more assimilated parents and, perhaps more important, older siblings who may have already acquired destination-specific skills, including language proficiency.

2. Census data

The empirical analyses presented below are based on the 1996 Australian Census of Population and Housing Household One Percent Sample File (HSF). These census data are released in the form of a hierarchical file. Thus, information is available on each family within a household, and for each individual within a family. This information can be linked as required for analysis. Thus, the individual information for one person within a family can be merged with the information for another person or persons within the same family. This information can, in turn, be linked with overall characteristics of the family and the household. For example, information on both partners in a married couple can be linked together so that their language skills can be compared. The influence of other characteristics of the partners on their language skills can be determined to the extent that they are collected in the Census. The data compiled in this way enable the husband–wife relationship to be documented in terms of its impact on English language skills. Similarly, information on children living at home can be combined with the information on one or both parents to permit quantification of the parent–child relationship in dominant language acquisition.

This HSF contains information on age, gender, marital status, birthplace, duration in the country, employment status, educational qualification/attainment, occupation, region of residence, relationships in households, and, of primary importance for this study, language spoken at home and English language proficiency, among other variables.[5]

The analyses in this chapter focus on the variables for language spoken at home and proficiency in the English language. Data collection for the proficiency variable was limited to people who indicated that they speak a language at home other than or in addition to English. These people were asked to state how well they speak English. Four categories were identified: very well, well, not well, and not at all. Individuals in the first category (speaks very well), together with those who speak only English at home, are categorized as "Proficient in English" in the analyses that follow. This is the categorization proposed by Barry Chiswick and Paul Miller (1995, p. 253), based on the links between earnings and the various English language proficiency categories and on documentation from the Australian Bureau of Statistics on the likely practical language skills of members of the separate language proficiency categories.[6]

There is one limitation to these data. The detailed information on spouses or on children exist only for those family members resident in the household on Census night. For the study of the language skills of spouses, this might be only a minor limitation, except for spouses living apart. It will be a more serious limitation in the study of the parent–child relationship where there are older children who have left home, and where there are dependent children away at school or who are still in the origin.[7]

3. Empirical analysis

Three separate sets of analyses are presented in detail, namely husband–wife relationship, father–eldest child relationship, and father–youngest child relationship. In the analysis for husband and wife in a family unit, the sample is restricted to opposite-sex adults who were both aged 20–64 years and who were both born overseas in non-English speaking countries. The analyses for both father–eldest child and father–youngest child relationships are limited to pairs where the fathers were between 20 and 64 years old and were born overseas in non-English speaking countries. Parallel analyses are also conducted for mother–offspring combinations. As the results for the parent–offspring analyses are largely invariant with regards to the parent's gender, the findings from the study of mothers and their children are presented only in summary form. In each instance, the sample used represents the maximum data available for the particular family members under analysis. Hence, the sample size for the father–eldest child analysis will differ from the sample size for the mother–eldest child analysis owing to missing values for either the mother or father.

Given the definition of proficiency used in the econometric analysis (speaks only English or speaks English very well), 46 percent of both husbands and wives are proficient in English. Among eldest children, the proficiency level is 88 percent, and it is 89 percent among youngest children.[8]

Analysis of cross-tabulations of the distributions of various family members across categories of the English speaking skills shows clearly that, within the typical household, both husbands and wives, and parents and children, have similar language skill levels. For example, 97 percent of wives who speak only English at home have husbands who also speak only English at home.[9] In the case of wives who speak English "not well", 57 percent of their husbands also do not speak English well.

Fully 70 percent of fathers whose eldest child spoke only English at home also speak only English at home. Very few of these fathers speak English "not well" or "not at all". In comparison, where the eldest child speaks a language other than English at home and speaks English "well", over 40 percent of the fathers are in the "not well" or "not at all" English skills categories. These patterns are repeated when the focus is on the youngest child, or the mother–children relationships rather than the father–children relationships.

Hence, cross-tabulations (available upon request) show both a strong commonality of the languages of family members, and yet considerable divergence in English proficiency within families. Understanding of these patterns can be aided through multivariate analysis.

Bivariate probit models

The variation in the proficiency rates documented above can be quantified using a bivariate probit framework (see Chiswick, Lee and Miller 2005). The model of dominant language fluency used in this study is based in large part on earlier work by Chiswick and Miller.[10] Chiswick and Miller (1992, 1995, 1998, 2001) model an immigrant's proficiency in English (speaking, reading and writing) skills (LANG) as:

$$\text{LANG} = f(\text{economic incentives, efficiency, exposure}). \tag{1}$$

An empirical counterpart to this model is developed using demographic characteristics (e.g., birthplace, age at migration, duration in the destination, educational attainment), and behavioral variables constructed from information on the immigrant's birthplace and mother tongue (e.g., geographic distance of the country of origin from the destination state, linguistic distance of the mother tongue from English).

Economic incentives for the acquisition of English skills among immigrants will depend on the labor market, consumption and community benefits expected to be associated with these skills, and the length of time over which these benefits are expected to accrue. As measurement of the benefits has proven to be difficult on an individual level, empirical research has focused on the length of time over which the various benefits are to be realized. Birthplace variables have often been used as a proxy for the probability of return migration, since origins differ in the extent of permanent and sojourner migration.

Efficiency refers to the extent to which a given amount of exposure to English produces language proficiency. The main variables included in the estimating equation developed by Chiswick and Miller (1992, 1995, 1998) to reflect efficiency aspects are age at migration (those who migrate while young will be more efficient in developing English language proficiency) and level of education (as a general index of the ability to acquire new skills, including language skills).

Exposure has three dimensions: exposure prior to migration, time units of exposure in the destination country, and the intensity of exposure per unit of time in the destination. These have been captured in empirical studies in various ways, depending on the extent of information in the data sets utilized. It is usual for exposure to English prior to migration to be captured by birthplace variables, where birthplaces are categorized according to British or American colonial or military influence. Time units of exposure in the destination country are captured by years since migration. Finally, the intensity of exposure per unit of time in the destination country has been captured by variables reflecting characteristics of the person's home and location, including data on the extent to which the immigrant's mother tongue is used in the region in which he or she lives, and the number and ages of children, and the birthplace, educational attainment and language skills of the spouse.

Thus, the empirical counterpart to the theoretical model used in this study is:

LANG = *f*(age at migration, duration in the destination, education, gender, birthplace, number of offspring/siblings, spouse's/parents' characteristics). (2)

The dependent variable (LANG) is a dichotomous variable, set equal to one where the immigrant speaks only English or speaks English very well (i.e., proficient in English), and to zero in other cases.

Due to the restriction of the data to two-partner families, where rates of marriage are very high (over 96 percent), the model does not include variables for marital status. The variable for the total number of offspring/siblings is used to capture some of the parent–child and within-siblings relationships noted earlier. It is derived from information collected from the female parent on the total number of births. This variable is argued to be superior to the information on the number of children currently living at home, as the total births variable is more likely to reflect the cumulative (or stock) effect on the English skills of children, rather than the flow effect that is likely to be associated with the use of information only on those children currently living at home.

While information on the total number of offspring is entered in the model, this variable is not disaggregated by age. There are two reasons for this. First, only the age of children living at home can be obtained. Second, as separate analyses are to be undertaken for children, and these analyses will then be integrated (i.e., estimated jointly) with the analyses for parents, there is less need to include detailed information on the age structure of children in the estimating equations.

The approach followed in this study differs from the research by Chiswick and Miller (1996, 1999) in several important ways. First, following the conceptual framework of Grossbard-Shechtman and Neuman (1991), variables for spouse's education are included in the estimating equation for the individual, and the education levels of both parents are included in the estimating equations for children.

A second difference is the absence in the current study of variables with behavioral interpretations that have been used in place of birthplace. These variables are usually constructed using information on the immigrant's birthplace or home language and region of residence. They include measures of ethnic concentration, linguistic distance between the immigrant's mother tongue and English, and physical distance between the immigrant's country of origin and the destination.

There are practical reasons for this omission. The data set used contains limited birthplace information. There are only 21 relevant birthplace codes, five of which refer to English-speaking regions.[11] There are only six individual non-English-speaking birthplaces separately identified among the remaining codes, with the balance of the codes being broad aggregates, such as "Other Southern Europe" and "Other Southeast Asia". With so few individual birthplaces identified, the construction of behavioral variables on the basis of this birthplace information is likely to have limited success. Similarly, only seven languages other than English are separately identified.[12] This precludes using the Census language information to create the behavioral variables.

Moreover, the analyses reported by Barry Chiswick and Paul Miller (1996, 1999) show that the birthplace-related variables have limited explanatory power in analyses for Australia. This contrasts sharply with the situation when similar models have been estimated for the US and Canada, where there is greater specificity on birthplace and languages spoken. Furthermore, unless one is specifically interested in behavioral interpretations from knowing country of origin, which is not the purpose of this study, Chiswick and Miller (1996, 1999) show that birthplace dichotomous variables are fine.

A third difference between the current set of analyses and the models used in previous research occurs in relation to the models estimated for "children". Children are identified through the census variable "Relationship in household" which has codes for "Child under 15", "Dependent student (15–24)" and "Non-dependent child". The model used to account for the language skills of children includes all the variables included in the analysis of the language skills of their parents, plus a variable for whether the child is still at school. The schooling variable is also defined differently for this group, recording total years of education for children who have left school, and incomplete years of education for children still at school.

The analyses are conducted in two stages. First, separate analyses are presented for (i) husbands and wives, (ii) fathers and eldest child, and (iii) fathers and youngest child. Then analyses are conducted simultaneously for both parents and the children. For the first set of analyses, the bivariate probit model to be employed may be expressed as[13]

$$\Pr(\text{LANG}_i^{\text{Husband}} = 1, \text{LANG}_i^{\text{Wife}} = 1) = F(\beta^{\text{Husband}} X_i^{\text{Husband}}, \beta^{\text{Wife}} X_i^{\text{Wife}}, \rho) \tag{3}$$

where F is the cumulative standard bivariate normal, X denotes the variables outlined in equation (2), the βs are sets of parameters to be estimated, and ρ denotes the correlation between the disturbances in the estimating equations for spouses. The coefficients in the models of language skills for each spouse are allowed to differ.[14] Bivariate probit analysis for pairs of family members offers a more efficient means of estimation over separate probit analysis for each family member where the disturbances in the individual equations are correlated. The analogy is often made to the advantages that seemingly unrelated regression estimation can provide over ordinary least squares in a multiple-equation setting (see Wiiliam Greene, 1991).

The interpretation of ρ is that it captures the correlation between the effects of unobservables in the models of language skills of two people. Consider a household where the husband had a relatively high ability for learning English. Under positive assortative mating, this implies similar characteristics for the wife. As this ability is not a measured variable, its influence will be captured via the error terms in the estimating equations for both husbands and wives, and a positive correlation between the error terms for spouses would therefore be expected. Alternatively, a model where comparative advantage leads to specialization might see above average English skills of the husband being associated with below average English skills of his wife, measured variables held constant. One spouse may then serve as the translator for the other. A negative correlation between the disturbance terms in the equations would be observed.

To illustrate the general features of the data, Appendix Table 5A.1 lists means and standard deviations of the dependent and explanatory variables used in the first set of analyses for husband and wife. This appendix table also contains definitions of the variables and lists results of the model of dominant language proficiency for spouses. The results are broadly the same for "husbands" and "wives" in the single equation and bivariate equation probits. They show that dominant language proficiency declines with age (or age at migration given that period of arrival is held constant). This effect is similar to previous studies of adult immigrants, and is generally attributed to the greater difficulty that immigrants have acquiring language skills when they migrate at an older age.

Years of (own) education are associated with better English language skills, with the partial effect of each year of schooling on the probit index being about the same as the impact of close to 30 extra years of age at the time of migration. The skills learned at school, or the index of adaptability provided by the years of schooling variable, is obviously of major importance to an understanding of dominant language proficiency.[15]

A person's spouse's years of education are also associated with greater proficiency in English, though the effect is less than that of own educational attainment. It is also seen that the effect of wife's education on the probit index for the husband (0.099 in the bivariate probit model) is greater than that of the husband on the probit index for the wife (0.071 in the bivariate probit model).[16]

It is indicated in the bivariate equation probits that dominant language proficiency is not affected by the number of children. This effect only arises once the spouse's educational attainment is held constant: In models which do not include the spouse's educational attainment, English proficiency decreases with the number of children, with the impact being more intense for females than for males. This pattern is presumably reflecting the greater fertility among the less-well educated.

Compared to immigrants from Australia's traditional source countries for non-English speaking immigrants (Europe), immigrants from the new source regions (Asia) have poorer English speaking skills.[17] The partial effect on the probit index of coming from Asia is the equivalent of around one less year of schooling (Table 5.1, bivariate probit model). Presumably, this is reflecting the

Table 5.1 Bivariate probit model of English speaking skills, 20–64 year old spouses from non-English speaking countries

Variable	*Single equation probit*		*Bivariate probit*	
	Husband	*Wife*	*Husband*	*Wife*
Constant	−2.555 (15.71)	−2.228 (13.13)	−2.625 (17.44)	−2.411 (15.53)
Age	−0.005 (2.56)	−0.011 (5.00)	−0.004 (2.54)	−0.007 (4.14)
Years of schooling (Self)	0.139 (12.18)	0.172 (13.64)	0.136 (12.16)	0.171 (13.96)
Years of Schooling (Spouse)	0.095 (7.90)	0.066 (5.81)	0.099 (8.47)	0.071 (6.36)
Number of children	0.003 (0.19)	−0.009 (0.48)	0.001 (0.04)	−0.020 (1.16)
Birthplace region (Europe)				
Asia	−0.211 (3.56)	−0.242 (4.03)	−0.127 (2.42)	−0.215 (4.11)
Remaining	0.231 (3.75)	0.222 (3.55)	0.247 (4.67)	0.238 (4.53)
Year of arrival (before 1981)				
1981–1985	−0.465 (5.96)	−0.574 (7.24)	−0.419 (6.96)	−0.527 (8.35)
1986–1990	−0.510 (7.37)	−0.549 (7.77)	−0.587 (10.61)	−0.697 (12.46)
1991–1994	−0.676 (8.01)	−0.838 (10.31)	−0.753 (11.06)	−0.882 (13.95)
1995–1996	−0.849 (6.90)	−1.259 (10.71)	−0.917 (8.79)	−1.140 (12.65)
χ^2	676.82	742.21	1064.40	
Prediction success Rate (%)	66.79	68.57	88.35	
Correlation coefficient, ρ	–		0.940 (162.86)	
Sample size	4104	4104	4104	

Note: Numbers in parentheses are '*t*' statistics; reference groups for dichotomous variables are in parentheses.

Source: 1996 Australian Census of Population and Housing, one percent sample file.

greater "linguistic distance" from English of the Asian than of the European languages (Chiswick and Miller, 1998).

Finally, there is a clear negative relation between being a more recent arrival and English speaking skills.

The correlation coefficient between the disturbance terms in the bivariate probit language skills equation for "Husbands" and "Wives" in Table 5.1 is sizeable and highly significant (coefficient of 0.940, with a '*t*' of 162.86).[18] The positive value for this coefficient means that in cases where there are unobservables that lead the husband to have greater (lesser) English speaking skills than predicted by the model, the same or other unobservables will result in the wife having greater (lesser) English skills than predicted by the model. Positive assortative mating on the basis of factors that are not included in the model (motivation, ability, even propensity for language skills development), one spouse learning from the other spouse who is more proficient for unobserved reasons, or a positive correlation in measurement error, will generate a positive correlation in the disturbance terms of the models of English speaking proficiency for husbands and wives.[19]

Interaction terms between the years of schooling of husband and wife were included in the model (results not reported here) to examine the degree of substitutability of these types of human capital in the production of language capital. This interaction term was significant and negative in the equation for each spouse. Moreover, the magnitude of the coefficient was similar for husband and wife, and in the single equation probit and in the bivariate probit. This suggests that the levels of schooling of husbands and wives, representing very similar forms of human capital, are substitutes in the production of language capital.

This evidence sits comfortably with findings concerning the substitutability of spouses' schooling from the study of earnings and spousal support for a person's career. For example, Shoshana Grossbard-Shechtman, Dafna Izraeli and Shoshana Neuman (1994) report a negative sign of an interaction term between own schooling and spouse's schooling in an equation explaining variation in the amount of help that female managers get from their husbands for their careers. The interaction term, however, was insignificant in an equation for male managers. Similarly, Grossbard-Shechtman and Neuman (1991), in a study of earnings of male salaried workers in Israel, report that the schooling of husband and wife are substitutes for Western workers (that is, those born in Europe or America, or born in Israel and father born in Europe, America or Israel).[20]

The results of the language proficiency model for fathers and their oldest child and youngest child living at home are presented in Tables 5.2 and 5.3, respectively.[21,22] The single equation results for adult males living in families where there is at least one child in Table 5.2 are reasonably similar to the single equation results reported for all adult males in Table 5.1, but they differ in one important respect: the coefficient on the Asia birthplace variable in Table 5.1 is considerably smaller (in absolute value) than that in Table 5.2.

Table 5.2 Bivariate probit model of English speaking skills, 20–64 year old fathers and their eldest child from non-English speaking countries

Variable	*Single equation probit*		*Bivariate probit*	
	Father	*Eldest child*	*Father*	*Eldest child*
Constant	−1.624 (7.48)	0.104 (0.34)	−1.630 (7.79)	−0.215 (0.70)
Age	−0.017 (5.68)	−0.050 (5.55)	−0.016 (5.47)	−0.043 (4.49)
Years of schooling (father)	0.157 (11.78)	0.043 (2.35)	0.154 (11.70)	0.041 (2.34)
Years of schooling (mother)	0.078 (5.43)	0.063 (3.10)	0.075 (5.35)	0.075 (3.88)
Years of schooling (eldest child)	–	0.149 (10.96)	–	0.147 (10.83)
Still at school	[a]	−0.254 (1.94)	[a]	−0.194 (1.53)
Female	[a]	0.115 (1.64)	[a]	0.108 (1.56)
Number of children	0.006 (0.25)	−0.028 (0.84)	0.004 (0.18)	−0.037(1.07)
Birthplace region (Europe for father, Australia for child)				
Europe	[a]	−0.321 (1.97)	[a]	0.040 (0.23)
Asia	−0.336 (4.88)	−0.937 (6.22)	−0.253 (3.70)	−0.653 (4.25)
Remaining	0.138 (2.00)	−0.232 (1.32)	0.117 (1.72)	−0.003 (0.02)
Year of arrival (before 1981)				
1981–1985	−0.724 (8.13)	−0.042 (0.22)	−0.678 (7.63)	−0.148 (0.75)
1986–1990	−0.849 (10.14)	−0.263 (1.63)	−0.852 (10.28)	−0.466 (2.90)
1991–1994[b]	−0.932 (8.75)	−0.899 (5.77)	−1.002 (9.99)	−1.048 (6.62)
1995–1996	−1.320 (7.55)	[b]	−1.123 (6.17)	[b]
χ^2	584.70	513.34	982.4	
Prediction success rate (%)	68.36	88.53	63.78	
Correlation coefficient, ρ	–		0.629 (18.69)	
Sample size	2946	2946	2946	

Notes: Numbers in parentheses are ‘*t*’ statistics; reference groups for dichotomous variables are in parentheses.

a Variable not relevant.

b The year of arrival dummy variables, 1991–1994 and 1995–1996, have been combined to form 1991– 1996 for children.

Source: 1996 Australian Census of Population and Housing, one percent sample file.

This implies that compared to fathers from Europe, the English language proficiency of Asian fathers is even lower when children are present.

From Table 5.2 it is clear that there is broad similarity between the results for “father” and “oldest child”. The only instance where there is a significant difference is the “year of arrival” variable. Among children in 1996, it is only arrivals after 1986 who have poorer English skills (in Australia ten or fewer years) whereas for adult males all arrival cohorts have poorer proficiency than longer-term residents (immigrated before 1981). This is consistent with findings reported in the literature to the effect that the young have a far greater capacity to learn languages than the old, and hence immigrants who are children would be expected to acquire English language skills more rapidly than adult immigrants. The school age young also have intensive exposure to English in school.

There is a positive association between years of education and English skill levels for both fathers and their oldest child, and the sizes of the partial effects on the respective probit indexes for the groups are similar. The father's educational attainment is also shown to have a modest positive impact on the language skills of the eldest child. The mother's educational attainment, however, has a larger effect on the language skills of the eldest child (coefficient of 0.075 compared with 0.041 for the father's educational attainment). Similarly, in the equations examining the language skills of mothers and their eldest child, the mother's level of schooling has a larger impact (coefficient of 0.093 in the bivariate probit) than the father's level of schooling (coefficient of 0.029).[23] Moreover, in equations for the eldest child, where either the mother's educational attainment or the father's educational attainment was included, but not both, the coefficient on the mother's educational attainment was greater than that of the father. In this sample, therefore, the wife's human capital is of greater importance than that of the husband in the production of language capital among children.

The correlation coefficient between the disturbance terms in the bivariate probit analysis for father and eldest child (0.629) is sizeable and highly significant (Table 5.2).[24] However, the correlation between the unobserved components in the model for fathers and their eldest child living at home is only about two-thirds of that between the unobserved components in the models for spouses (0.940). This suggests that different sets of unobserved factors may impact on the dominant language acquisition of parents and of their children.

The statistical significance of parent's educational attainments in the language equations for children permits comment on the relative merits of the human capital and marital sorting interpretations of the findings for the strong husband-wife relationships documented above. In particular, while the parent's educational attainment variable could have a human capital interpretation in the child's language equation, it should not have a marital sorting interpretation. Hence, the Table 5.2 results provide for an emphasis on the former explanation for the various sets of results presented in this chapter.

Table 5.3 reports results of the model for fathers and their youngest child living at home. Examination of the coefficients in this table reveals that they are broadly similar to those listed in Table 5.2. One exception is that the coefficients on the schooling levels of father and mother in the equation for the English language skills of the youngest child are about the same, whereas the mother's level of schooling was more important than the father's level of schooling for the language capital of the eldest child.

The correlation coefficient presented in this table between the error terms of the father and the youngest child is positive and significant (0.537).[25] Moreover, it is of about the same order of magnitude as the correlation coefficient in Table 5.2 for the model covering the English language skills of

Table 5.3 Bivariate probit model of English speaking skills, 20–64 year old fathers and their youngest child from non-English speaking countries

Variable	*Single equation probit*		*Bivariate probit*	
	Father	*Youngest child*	*Father*	*Youngest child*
Constant	−1.824 (7.86)	−0.093 (0.28)	−1.813 (8.12)	−0.279 (0.79)
Age	−0.013 (4.17)	−0.041 (3.71)	−0.013 (4.08)	−0.038 (3.14)
Years of schooling (father)	0.156 (11.04)	0.055 (2.78)	0.154 (11.01)	0.053 (2.61)
Years of schooling (mother)	0.074 (4.88)	0.049 (2.22)	0.072 (4.89)	0.058 (2.68)
Years of schooling (Youngest child)	–	0.136 (8.55)	–	0.136 (8.10)
Still at school	[a]	−0.014 (0.10)	[a]	0.014 (0.09)
Female	[a]	0.050 (0.67)	[a]	0.035 (0.47)
Number of children	0.018 (0.74)	0.034 (0.98)	0.019 (0.76)	0.032 (0.92)
Birthplace region (Europe for father, Australia for child)				
Europe	[a]	−0.468 (2.39)	[a]	−0.132 (0.58)
Asia	−0.294 (3.95)	−1.172 (6.28)	−0.239 (3.22)	−0.954 (4.66)
Remaining	0.241 (3.19)	−0.342 (1.60)	0.216 (2.92)	−0.175 (0.75)
Year of arrival (before 1981)				
1981–1985	−0.735 (7.37)	0.113 (0.47)	−0.697 (7.10)	−0.020 (0.07)
1986–1990	−0.794 (8.45)	−0.031 (0.16)	−0.768 (8.12)	−0.212 (1.01)
1991–1994b	−0.866 (7.21)	−0.775 (4.06)	−0.945 (8.23)	−0.900 (4.27)
1995–1996	−1.372 (7.06)	[b]	−1.216 (5.44)	[b]
χ^2	490.20	412.15	800.00	
Prediction success rate (%)	67.98	89.71	61.58	
Correlation coefficient, ρ	–		0.537 (13.72)	
Sample size	2595	2595	2595	

Notes: Numbers in parentheses are '*t*' statistics; reference groups for dichotomous variables are in parentheses.

a Variable not relevant.

b The year of arrival dummy variables, 1991–1994 and 1995–1996, have been combined to form 1991– 1996 for children.

Source: 1996 Australian Census of Population and Housing, one percent sample file.

the father and the oldest child (0.629). That is, the unobserved factors in the equation for the English language skills of the youngest and eldest children have similar relationships with the unobserved factors in the equation for the English language skills of their father, although it is a bit weaker for the youngest child. Thus, the model appears to be robust to the choice of sibling to use in the bivariate probit analysis.

Table 5.4 reports the correlation coefficients between the disturbance terms in the substantive equations for the various bivariate probit models. The correlation coefficient of 0.940 between spouses is much larger than those

Table 5.4 Correlation coefficients between residuals of English speaking skills obtained from Tables 5.1–5.3, 5A.2 and 5A.3 (Appendix A) (bivariate probit analyses)

	Mother	*Eldest child*	*Youngest child*
Father	0.940 (0.006)	0.629 (0.034)	0.537 (0.039)
Mother		0.658 (0.033)	0.617 (0.037)
Eldest child			0.909[a] (0.022)

Notes: Numbers in parentheses are standard errors.
a The analysis for this cell excludes single-child families.

Source: 1996 Australian Census of Population and Housing, one percent sample file.

between parents (male or female) and their children (oldest and youngest), ranging from 0.537 to 0.658. It is also observed that the correlation coefficients between mothers and children (both oldest and youngest) are larger (0.658 and 0.617, respectively) compared to those between fathers and children (0.629 and 0.537, respectively). This would appear to be consistent with the proposition that mothers are more involved in the socialization and home produced human capital of their children than are the fathers. Similarly, the partial efects of mother's educational attainment on the English speaking skills of their children were marginally greater than the effects of father's educational attainment on the English speaking skills of their children.

In summary, these analyses show quite clearly that there are strong links between the observed and unobserved determinants of English language skills within the family among immigrants in Australia. This result holds for each of the parent-offspring and partners combinations considered. The correlations between the disturbance terms in the models are larger between partners than for the parents-offspring combinations and are stronger for the mother-offspring than for the father-offspring analyses. As one would expect, there is also a very high correlation between the disturbance terms in the model for the eldest child and youngest child. These findings are intuitively reasonable.

A four-equation probit model

In the remainder of this section, the models developed above are generalized to a multivariate probit model covering four family members (namely, father, mother, eldest child and youngest child), given as:

$$\Pr(\mathrm{LANG}_i^{\mathrm{Husband}} = 1, \mathrm{LANG}_i^{\mathrm{Wife}} = 1, \mathrm{LANG}_i^{\mathrm{Eldest}} = 1, \mathrm{LANG}_i^{\mathrm{Youngest}} = 1)$$

$$= F(\beta^{\mathrm{Husband}} X_i^{\mathrm{Husband}}, \beta^{\mathrm{Wife}} X_i^{\mathrm{Wife}}, \beta^{\mathrm{Eldest}} X_i^{\mathrm{Eldest}}, \beta^{\mathrm{Youngest}} X_i^{\mathrm{Youngest}}, \rho\mathrm{HW}, \rho\mathrm{HE}, \rho\mathrm{HY}, \rho\mathrm{WE}, \rho\mathrm{WY}, \rho\mathrm{EY})$$

where "eldest" (E) and "youngest" (Y) refer to the eldest child and youngest child living at home, respectively, and ρ_{ij} is the correlation coefficient between the disturbance terms in the equations for persons i and j.

It is to be noted that the sample for this estimation is restricted to two-parent families with at least two children living at home. In other words, couple families with only one child living at home (i.e., the eldest child is also the youngest child) are excluded from the analysis. This restriction reduces the sample size to around 40 percent of that used in the earlier analyses.

Estimates from the four-equation probit model are presented in Table 5.5. The results show that language proficiency declines with age for "mother", "eldest child", and "youngest child". The age effect for father is negative and of the same order of magnitude as that reported above in the bivariate probit analyses, but the 't' statistic is only 0.72.

Years of (own) education are positively associated with English skill levels for every member of the family included in the statistical analysis. The coefficients of the years of schooling variable for "eldest child" and "youngest child" are marginally smaller than those for "father" and "mother". In the model for "eldest child", those who are still at school have lower English language proficiency. This relationship was not evident in the bivariate analyses presented above, where the "eldest child" sample included one-child families. Combining the results of the separate analyses in this manner suggests that the Table 5.5 finding may be associated with the age proximity of children: where the eldest child in a family with at least two children is still at school, and the youngest child is at least 5 years of age, the children may be sufficiently close in age that they interact more than in the case where there is a wider gap between the ages of the youngest and eldest child. This interaction may involve the use of their parents' mother tongue.

Compared to European immigrants, Asian immigrants have poorer English speaking skills. This is especially true for the children. There is a clear negative association between being a more recent arrival and English speaking skills in the models for "father" and "'mother".

The spouse's level of schooling is significant in the equations estimated for the father and mother (both the single equation and the multivariate probit), with an impact about one-half that for the parent's own level of education. Surprisingly, the years of schooling of the father, but not that of the mother, is significant in the equation for each of the children. This is despite the earlier finding that both schooling variables (for father and for mother) were significant when the analysis was based on the separate samples of "eldest child" and "youngest child", and the mother's level of schooling actually had a greater partial effect on the probit index for the analysis of the eldest child. The main difference between the sets of analyses is that the multivariate probit model presented here is fitted to a sample that excludes all singleton families. While larger families may be associated with traditional patterns of family-gender roles among immigrant families (as discussed in Section 2),

Table 5.5 Multivariate probit model of English speaking skills, 20–64 year old spouses from non-English speaking countries and their eldest child and youngest child, 1996

Variable	*Single equation probit*				*Multinomial probit*			
	Father	*Mother*	*Eldest child*	*Youngest child*	*Father*	*Mother*	*Eldest child*	*Youngest child*
Constant	–2.258 (5.87)	–1.240 (3.26)	1.142 (2.28)	0.156 (0.32)	–2.173 (5.93)	–1.556 (4.40)	0.812 (1.75)	–0.300 (0.61)
Age	–0.004 (0.72)	–0.028 (4.42)	–0.074 (5.13)	–0.064 (2.43)	–0.004 (0.93)	–0.018 (3.51)	–0.049 (3.51)	–0.032 (1.21)
Years of schooling (father)	0.143 (6.58)	0.054 (2.50)	0.105 (3.36)	0.088 (3.11)	0.133 (6.17)	0.055 (2.47)	0.079 (2.43)	0.079 (2.48)
Years of schooling (mother)	0.074 (3.20)	0.170 (7.07)	–0.019 (0.60)	0.009 (0.29)	0.078 (3.33)	0.160 (6.75)	0.009 (0.28)	0.012 (0.34)
Years of schooling (child)	–	–	0.115 (4.90)	0.144 (4.79)	–	–	0.089 (4.51)	0.119 (3.99)
Still at school	[a]	[a]	–0.609 (3.31)	–0.127 (0.61)	[a]	[a]	–0.435 (2.44)	0.076 (0.42)
Female	[a]	[a]	0.060 (0.54)	–0.011 (0.11)	[a]	[a]	–0.073 (0.81)	0.075 (0.79)
Number of children	–0.029 (0.68)	0.009 (0.20)	–0.018 (0.31)	0.002 (0.04)	–0.039 (0.92)	–0.016 (0.36)	–0.035 (0.57)	–0.006 (0.11)
Birthplace region (Europe for parents, Australia for children)								
Europe	[a]	[a]	–0.107 (0.43)	–0.323 (0.91)	[a]	[a]	0.553 (0.17)	–0.086 (0.22)
Asia	–0.267 (2.49)	–0.360 (3.27)	–0.772 (3.52)	–0.927 (2.81)	–0.092 (0.89)	–0.200 (1.98)	–0.570 (2.30)	–0.727 (2.04)
Remaining	0.057 (0.51)	–0.097 (0.84)	–0.031 (0.12)	–0.256 (0.75)	0.028 (0.26)	–0.053 (0.50)	–0.043 (0.18)	–0.108 (0.30)
Year of arrival (before 1981)								
1981–1985	–0.448 (3.32)	–0.590 (4.42)	0.006 (0.02)	0.002 (0.01)	–0.403 (3.52)	–0.547 (5.01)	0.137 (0.47)	0.285 (0.72)
1986–1990	–0.322 (2.47)	0.518 (3.99)	–0.414 (1.80)	0.107 (0.32)	–0.463 (3.95)	–0.535 (4.61)	–0.408 (1.69)	0.061 (0.18)
1991–1994[b]	–0.446 (2.87)	–0.510 (3.24)	–0.955 (4.12)	–0.646 (1.97)	–0.560 (3.37)	–0.638 (4.58)	–0.662 (2.58)	–0.445 (1.29)
1995–1996	–0.863 (3.44)	–1.304 (4.56)	[b]	[b]	–0.626 (2.21)	–0.972 (3.32)	[b]	[b]
χ^2	165.14	191.30	167.04	139.77	402.46			
Prediction success rate (%)	65.22	66.67	88.22	86.23	–	–	–	–
Sample size	1104	1104	1104	1104	1104			

Notes: Numbers in parentheses are '*t*' statistics; reference groups for dichotomous variables are in parentheses.
a Variable not relevant.
b The year of arrival dummies, 1991–1994 and 1995–1996, have been combined to form 1991–1996 for children.

Source: 1996 Australian Census of Population and Housing, one percent sample file.

one would still expect this to result in a larger effect of mothers than of fathers on their children's language proficiency.

The six correlation coefficients between unobserved influences on English skills, listed in Panel A of Table 5.6, are positive and significant. The correlation coefficient between fathers and mothers (i.e., opposite-sex couples), 0.931, is much larger than those between parents (fathers and mothers) and their child (either eldest or youngest). The point estimates of the coefficient between fathers and their eldest child (0.637) and between fathers and their youngest child (0.500) suggest that there is a stronger linguistic bond between fathers and their eldest child (who for a while was also an only child), than with their youngest child. It appears that mothers have a stronger bond with their youngest child rather than with their eldest child (0.577 vs. 0.552), though this difference is not statistically significant. Consistent with the literature, there is a strong linguistic bond between siblings, as indicated by the high value of the correlation between the error or residual terms for the eldest child and the youngest child (0.903) in families with at least two children living at home.

The separation of the sample according to whether the eldest child was born in Australia or born overseas yields interesting results (see Panels B and C of Table 5.6). Compared to the case where the eldest children are not disaggregated by birthplace (Panel A), the point estimates of the correlation coefficients between parents (mothers and fathers) and their children (eldest and youngest) have much lower values if the eldest child was born in Australia (Panel B) and much higher if the eldest child is foreign born (Panel C). These results suggest a greater inter-connectedness between parents and their children's destination specific human capital where their eldest children were born overseas. Presumably this also reflects the impact of stronger origin/cultural heritage factors that influence both parents and child where the eldest children were born abroad.

The comparison of the results in Panels B and C also reveal that there is a stronger bond between siblings for the sample where the eldest child was born in Australia (0.995) compared to the case where the eldest child was born overseas (0.872). Where the eldest child was born abroad some of their younger siblings may have been born in Australia. However, where the eldest child was born in Australia nearly all of his/her younger siblings have also been born in Australia. The difference between the correlations between the error terms for two siblings in Panels B and C could therefore be due to foreign-born children having more origin-specific characteristics, including language skills, compared to their Australian-born counterparts.

4. Summary and conclusions

This chapter extends the line of research that views migration and immigrant adjustment as a consequence of a family decision making process, rather than solely as an individual decision. It focuses on the determination of

Table 5.6 Correlation coefficients between residuals of English speaking skills among fathers, mothers, their eldest and youngest children by birthplace of eldest child, (multivariate probit analyses)

	Panel A: Eldest child born in Australia or Overseas (Sample size 1104)			*Panel B: Eldest child born in Australia (Sample size 607)*			*Panel C: Eldest child born Overseas (Sample size 497)*		
	Mother	*Eldest child*	*Youngest child*	*Mother*	*Eldest child*	*Youngest child*	*Mother*	*Eldest child*	*Youngest child*
Father	0.931	0.637	0.500	0.927	0.461	0.423	0.949	0.794	0.612
Mother		0.552	0.577		0.528	0.526		0.883	0.825
Eldest child			0.903			0.995			0.872

Source: 1996 Australian Census of Population and Housing, one percent sample file.

destination language proficiency among members of immigrant families—fathers, mothers and children.

Building on earlier research on the determinants of destination language proficiency among immigrants, the chapter discusses the literature and theoretical relationships among the language skills of spouses/partners, between parents and children, and among siblings. Between spouses there may be a positive assortative mating on the unmeasured determinants of language proficiency (e.g., linguistic ability) as well as the measured determinants (e.g., schooling level, country of origin). Specialization in activities within the marriage might, however, result in negative assortative mating on some relevant dimensions. There may also be language learning between spouses, that is, one spouse learning from the other. Spouses may also invest in the human capital, including language capital, of each other.

Several factors might influence the parent/child proficiency relationship. The presence of children might lower parental destination language proficiency if children serve as translators for their parents (more likely for non-labor market than for labor market activities), if children lower their parents', especially their mother's, labor supply, and if parents use the origin language at home to transmit it and the origin culture to their children. On the other hand, because of the greater exposure to the destination language in school and the greater ability of youths to acquire destination language skills, they may serve as their parents' teachers and role models in the destination language. Moreover, the relationship between parent's and child's proficiency may be stronger for the mother because of the greater time input of mothers in the rearing of children.

Siblings, too, can influence language skills. The larger the number of siblings the greater the linguistic interactions in the home with destination language speakers, and the less the interaction with their foreign-speaking parents, especially for the youngest as distinct from the oldest child, who by definition was for a time an only child.

The econometric analyses are based on probit analyses for each type of person separately (husband, wife, eldest child, youngest child), bivariate probit analyses for pair-wise relationships, and four-equation (multivariate) probit analysis that jointly considers all four relationships. The findings are very robust across statistical techniques.

English language proficiency for all four groups is greater the younger the age at migration, the longer the duration in the destination, and the higher the level of (own) education. Each spouse's language proficiency is positively related to the level of schooling of the other spouse, although the effect is smaller than the effect of one's own schooling. Other than for the special case of larger families, where traditional family–gender roles that favor adult males may dominate, the mother's level of schooling is more important to children's English language proficiency than is the father's level of schooling, though both are important. The positive relationship between parent's level of schooling and children's English language proficiency, especially the larger impact of the mother, implies that a human capital investment model, rather than a marital sorting model, might be used to account for the acquisition of destination language capital.

Compared with those born in Europe, immigrants are less proficient if born in Asia. This is presumably due to the greater "linguistic distance" from English of the Asian than of the European and other immigrant languages.

Other measured variables the same, gender has no separate effect on language proficiency. Boys appear to be as proficient as girls.

In the multivariate probit analysis there is a very high positive correlation in the residuals between the father and the mother (over 0.9), as well as between the eldest child and the parents when this child was born overseas (about 0.8 or more). Because of the very high level of proficiency for children born in Australia, the parent/child correlations are lower if the eldest child is born in Australia (0.4 for the parent/eldest child residual correlation). The correlation between the residuals is stronger in the mother–child analysis than in the father–child analysis, especially when the eldest child is born overseas. This presumably reflects children's language learning at home, which is more likely to come from the mother than the father. The correlation between the residuals for the eldest/youngest child is very high (0.9), regardless of the eldest child's birthplace.

These correlations between and among residuals suggest there is a positive relationship between the unmeasured determinants of proficiency across family members. This may be due to positive "assortative mating" (positive correlations among marriage partners), inherited genetic factors between parents and children, or environmental factors among family members that make them more alike in the unmeasured characteristics that determine proficiency. It may also be due to the learning in the household among family members from each other. If for some (unmeasured) reason one family member acquires greater proficiency, the other family members learn from him or her. It could also arise where there are unobserved dimensions of human capital (e.g., quality of education), that are positively correlated across family members. This would parallel the relationships found between the observed levels of schooling of husbands and wives and the English language skills of all family members.

The strong unmeasured bond between parents and between (eldest and youngest) siblings is not surprising. The stronger relationship between the unmeasured determinants of language skills between mother and child than between father and child may be due to the greater involvement of mothers in child-rearing. This is consistent with the important role of language learning in the home.

Moreover, the analysis also suggests that previous estimates of rates of return from investment in language skills based on individual earnings and language proficiency may be biased downward. If language learning takes place in the home, there is a spillover effect, or externality, from one family member's investment in language training, namely, the improved language skills of other family members.

Appendix 5A

Definitions of variables

This appendix contains brief description of all variables used in the multivariate analyses.

Table 5A.1 Means and standard deviations for variables used in study of spouses' English speaking skills (Table 5.1)

Variable	*Definition*	*Husband*	*Wife*
Proficient in English	Binary indicator, set equal to one where the respondent speaks only English at home, or if a language other than English is spoken at home, English is spoken very well	0.424 (0.49)	0.414 (0.49)
Age	Continuous variable for age	44.097 (10.34)	43.081 (10.62)
Years of schooling (self)	Continuous variable that records the equivalent years of full-time education completed by the individual	12.246 (2.78)	11.708 (2.62)fs
Years of schooling (spouse)	Continuous variable that records the equivalent years of full-time education completed by the individual's spouse	11.708 (2.62)	12.246 (2.78)
Number of children	Continuous variable defined for all family members from the information on total number of children born to the female spouse in the family	2.250 (1.30)	2.250 (1.30)
Europe	Binary variable for respondents born in any European country	0.512 (0.50)	0.498 (0.50)
Asia	Binary variable for respondents born in any Asian country	0.307 (0.46)	0.325 (0.47)
Remaining birthplaces	Binary variable for respondents born in non-European, non-Asian foreign countries	0.180 (0.38)	0.177 (0.38)
Arrived before 1981	Binary variables for immigrants who arrived before 1981	0.617 (0.49)	0.590 (0.49)
Arrived 1981–1985	Binary variables for immigrants who arrived between 1981 and 1985	0.099 (0.30)	0.099 (0.30)
Arrived 1986–1990	Binary variables for immigrants who arrived between 1986 and 1990	0.161 (0.37)	0.156 (0.36)
Arrived 1991–1994	Binary variables for immigrants who arrived 1991 and 1994	0.088 (0.28)	0.108 (0.31)

(Continued Overleaf)

Table 5A.1 Continued

Variable	*Definition*	*Husband*	*Wife*
Arrived 1995–1996	Binary variables for immigrants who arrived between 1995 and 1996	0.036 (0.19)	0.047 (0.21)
Sample size		4104	4104

Note: Numbers in parentheses are standard deviations.

Source: 1996 Australian Census of Population and Housing, one percent sample file.

Acknowledgements

Chiswick acknowledges the research support of the Institute of Government and Public Affairs, University of Illinois. Miller acknowledges the financial support of the Australian Research Council. Helpful comments from the editor and an anonymous referee are appreciated.

Notes

1 The setting of migration and post-migration decision making in a family context is another application of the new household economics, initially developed by Becker and Mincer. For discussion of the ways spouses contribute to each other's (non-language) human capital, labor market performance and home duties, see Grossbard-Shechtman and Neuman (1991, 1993) and Grossbard-Shechtman, Dafna Izraeli and Neuman (1994).

2 Grossbard-Shechtman and Newman (1991) also consider interaction terms between wife's and husband's ethnicity, and between ethnicity and schooling. An attempt is made to use the results to distinguish between human capital and marital sorting interpretations of the earnings determination process.

3 Ziegler (1977, p. 330) notes ". . . many parents expected their children to learn English for them, because they felt too old to learn."

4 The typical Israeli myth in the period of mass immigration following independence was that the parents would learn Hebrew from their children.

5 For those who speak only English at home there are no data on whether they know, or in other contexts speak, another language.

6 Similarly, for the US, Robert Kominski (1989), on the basis of data from "test censuses" conducted by the US Census Bureau, argued that the use of two English skills categories in place of the four reported in the US Census (which are similar to those used in the Australian Census) is supported by the absence of clear differentiation between each of the four levels of English-speaking ability.

7 Leif Jensen and Yoshimi Chitose (1994, p. 717) note that ". . . the corresponding focus only on those children still residing with their parents, imposes an inevitable selectivity problem. To the extent that there are systematic differences between second- and higher-generation children in the rate at which and reasons for which they leave their families of orientation, these comparisons are biased". In the current study, there may also be systematic differences among children in the age at which they leave home that are related to English language skills. This cannot be tested with the data available.

8 The analysis is restricted to children aged 5 years and above (i.e., school-aged or

older children) as the census did not inquire about the language proficiency of younger children.

9 This high proportion may arise because the question does not refer to languages one can speak, but rather to languages actually spoken in the home, other than just a few words.

10 This model has been been applied successfully for the US, Canada and Australia in these Chiswick–Miller studies and to Israel in Barry Chiswick (1998). For applications to Germany and the UK, see Christian Dustmann (1994) and Michael Shields and Stephen Wheatly Price (2002), respectively. The patterns are remarkably similar across countries.

11 In addition to Australia, the 20 foreign birthplace codes are: England; New Zealand; Other United Kingdom and Ireland; Scotland; Italy; Vietnam; Greece; China (excluding Taiwan); Germany; Philippines; Other Southern Europe; Other Southeast Asia; Other Europe and the Former USSR; The Middle East and North Africa; Northern, Central and South America and the Caribbean; Southern Asia; Other Northeast Asia; Other Western Europe; Africa (excluding North Africa); Other Oceania and Antarctica. Although most Western Hemisphere immigrants are from the US and Canada, they cannot be separately identified in the Census.

12 The seven languages other than English that are identified and the proportions of adult immigrants (20–64 years) reporting these languages are: Italian (2.6 percent); Chinese (2.2 percent); Greek (1.8 percent); Arabic (0.9 percent); Vietnamese (0.9 percent); German (0.6 percent); and Spanish (0.6 percent). Fully, 83.5 percent report using only English at home, and 6.8 percent report using non-English languages other than the seven listed above.

13 The expression also applies to the cases of Father–Oldest Child and Father–Youngest Child.

14 The bivariate and multivariate probit models are estimated using full information maximum likelihood. The LIMDEP package is used, and William Greene (2002) contains technical details.

15 The positive effect of education on destination language proficiency cannot be attributed entirely to learning English in school in the origin since among immigrants to Israel, Hebrew language proficiency increases with level of schooling (Chiswick 1998).

16 Grossbard-Shechtman and Neuman's (1991) discussion of cross-productivity effects is based around a similar pattern.

17 Countries that form Europe are: Italy; Greece; Germany; Other Southern Europe; Other Europe and the Former USSR; and Other Western Europe. Asia comprises: Vietnam; China; The Philippines; Other Southeast Asia; Southern Asia; and Other Northeast Asia. The Remaining countries are: Middle East and North Africa; Northern, Central, South America and the Caribbean; Africa; and Other Oceania and Antarctica.

18 This compares with the simple correlations between key observables for the husband and wife of 0.761 for the measure of English proficiency, 0.735 for schooling and 0.619 for age.

19 The positive correlation in measurement error could arise from one person filling in the Census form for both partners. However, the pattern in Table 5.4 for the correlations being much higher for the partner–partner and sibling–sibling comparisons than for the parent-offspring comparisons, suggests the correlations between the disturbance terms does not arise from one person completing the Census form for all household members.

20 Grossbard-Shechtmana and Neuman (1994), however, report that the schooling of husband and wife are complements for non-Western workers in Israel.

21 The language proficiency model for mothers and their oldest child and youngest child are presented in Tables 5A.2 and 5A.3 of Appendix A, respectively.

22 In one-child families, this child is included in the analyses for both the oldest and youngest child. Analyses restricted to families with two or more children are presented later.
23 The two samples (fathers and eldest child and mothers and eldest child) are not independent: however, they do differ to the extent that observations with missing values for the father have valid data for the mother, and *vice versa*.
24 The simple correlation between the measure of English proficiency used in the statistical analysis for fathers and eldest child is 0.317.
25 The simple correlation between the measure of English proficiency used in the statistical analysis for fathers and eldest child is 0.285.

References

Baker, Michael and Dwayne Benjamin. (1997). "The Role of the Family in Immigrants' Labor-Market Activity: An Evaluation of Alternative Explanations." *American Economic Review* 87(4), 705–727.

Birrell, Robert. (1987). The Educational Achievement of Non-English Speaking Background Students and the Politics of the Community Languages Movement. In Lyle Baker, and Paul Miller, (eds.), *The Economics of Immigration: Proceedings of a Conference at the Australian National University*. Canberra: Australian Government Publishing Service, pp. 91–121.

Blake, Judith. (1980). *Family Size and Achievement*. Berkeley: University of California Press.

Breton, Albert. (1978a). *Bilingualism: An Economic Approach*. Montreal: C.D. Howe Research Institute.

Breton, Albert. (1978b). "Nationalism and Language Policies." *Canadian Journal of Economics* 11(4), 656–668.

Chiswick, Barry R. (1998). "Hebrew Language Usage: Determinants and Effects on Earnings Among Immigrants in Israel." *Journal of Population Economics* 11(2), 253–271.

Chiswick, Barry R., Yew Liang Lee, and Paul W. Miller. (2005). "Family Matters: The Role of the Family in Immigrants' Destination Language Acquisition." *Journal of Population Economics* 18(1), 1–17(IZA, Institute for the Study of Labor, Bonn, Discussion Paper No. 460, March 2002.).

Chiswick, Barry R. and Paul W. Miller. (1992). Language in the Immigrant Labor Market. In Barry R. Chiswick, (eds), *Immigration, Language and Ethnicity: Canada and the United States*. Washington, D.C: The AEI Press, pp. 229–296.

Chiswick, Barry R. and Paul W. Miller. (1995). "The Endogeneity Between Language and Earnings: International Analyses." *Journal of Labor Economics* 13(2), 246–288.

Chiswick, Barry R. and Paul W. Miller. (1996). "Ethnic Networks and Language Proficiency Among Immigrants." *Journal of Population Economics* 9, 19–35.

Chiswick, Barry R. and Paul W. Miller. (1998). "English Language Fluency Among Immigrants in the United States." *Research in Labor Economics* 17, 151–200.

Chiswick, Barry R. and Paul W. Miller. (1999). "Immigration, Language and Multiculturalism in Australia." *Australian Economic Review* 32(4), 369–385.

Chiswick, Barry R. and Paul W. Miller. (2001). "A Model of Destination Language Acquisition: Application to Male Immigrants in Canada." *Demography* 38(3), 391–409.

Dumon, Wilfried. (1989). "Family and Migration." *International Migration* 27(2), 251–270.

Dustmann, Christian. (1994). "Speaking Fluency Writing Fluency and Earnings of Migrants." *Journal of Population Economics* 7(2), 133–156.

Enloe, Walter and Philip Lewin. (1987). "Issues of Integration Abroad and Readjustment to Japan of Japanese Returnees." *International Journal of Intercultural Relations* 11(3), 223–248.

Espenshade, Thomas and Haishan Fu. (1997). "An Analysis of English Language Proficiency Among U.S. Immigrants." *American Sociological Review* 62, 288–305.

Gavaki, Efie. (1979). "The Greek Family in Canada: Continuity and Change and the Process of Adjustment." *International Journal of Sociology of the Family* 9(1), 1–16.

Greene, William H. (1991). *Econometric Analysis*. New York: Macmillan Publishing Company.

Greene, William H. (2002). *LIMDEP Version 8.0, Econometric Modelling Guide, Plainview*. NY: Econometric Software Inc.

Grossbard-Shechtman, Shoshana, Dafna N. Izraeli, and Shoshana Neuman. (1994). "When Do Spouses Support a Career? A Human Capital Analysis of Israeli Managers and Their Spouses." *Journal of Socio-Economics* 23(1/2), 149–157.

Grossbard-Shechtman, Shoshana and Shoshana Neuman. (1991). "Cross-Productivity Effects of Education and Origin on Earnings: Are They Really Reflecting Productivity?." *Handbook of Behavioral Economics* 2(A), 125–145.

Grossbard-Shechtman, Shoshana and Shoshana Neuman. (1993). *Religiosity and Investments in Spousal Productivity, Chapter 15 in Shoshana Grossbard-Shechtman The Economics of Marriage, Divorce and Labor*. Colorado, USA: Westview Press 290–302.

Haines, David, Dorothy Rutherford, and Patrick Thomas. (1981). "Family and Community Among Vietnamese Refugees." *International Migration Review* 15(1), 310–319.

Hatton, Timothy and Jeffrey Williamson. (1997). *The Age of Mass Migration: An Economic Analysis*. New York: Oxford University Press.

Hirschman, Charles. (1994). "Problems and Prospects of Studying Immigrant Adaptation from the 1990 Population Census: From Generational Comparisons to the Process of 'Becoming American'." *International Migration Review* 28(4), 690–713.

Jensen, Leif and Yoshimi Chitose. (1994). "Today's Second Generation: Evidence from the 1990 U.S. Census." *International Migration Review* 28(4), 714–735.

Kominski, Robert. (1989). "How Good is 'How Well'? An Examination of the Census English-Speaking Ability Question." *American Statistical Association 1989 Proceedings of the Social Statistics Section*, 333–338.

Long, Michael H. (1990). "Maturational Constraints on Language Development." *Studies in Second Language Acquisition* 12(3), 251–285.

Mincer, Jacob. (1978). "Family Migration Decisions." *Journal of Political Economy* 86(5), 749–773.

Naidoo, Josephine C. and J. Campbell Davis. (1988). "Canadian South Asian Women in Transistion: A Dualistic View of Life." *Journal of Comparative Family Studies* 19(2), 311–327.

Nauck, Bernhard. (1989). "Assimilation Process and Group Integration of Migrant Families." *International Migration* 27(1), 27–47.

Portes, Alejandro and Richard Schauffler. (1994). "Language and the Second Generation: Bilingualism Yesterday and Today." *International Migration Review* 28(4), 640–661.

Service, Elisabet and Fergus I.M. Clark. (1993). "Differences Between Young and Older Adults in Learning a Foreign Language." *Journal of Memory and Language* 32, 608–623.

Shields, Michael and Stephen Wheatley Price. (2002). "The Effect of Language Fluency on the Wages of Immigrant Men in England." *Journal of Population Economics* 15, 137–160.

Stark, Oded. (1991). *The Migration of Labor*. Cambridge, Massachusetts: Blackwell Publishers.

Sung, Betty L. (1985). "Bicultural Conflicts in Chinese Immigrant Children." *Journal of Comparative Family Studies* 16(2), 255–269.

Ziegler, Suzanne. (1977). "The Family Unit and International Migration: The Perceptions of Italian Immigrant Children." *International Migration Review* 11(3), 326–333.

Part II

The effects of language proficiency on labor market outcomes

6 Immigrant earnings

Language skills, linguistic concentrations and the business cycle[1]

1. Introduction

This chapter is concerned with the analysis of the earnings of immigrants. In particular, it focuses on several inter-related and unresolved issues using data from the 1990 Census of Population of the United States.

One set of issues relates to the individual's own language skills. To what extent are the earnings of immigrants in the United States influenced by the respondent's proficiency in English?[2] Has there been a change over time in the effect of language proficiency on earnings? Are different forms of human capital complementary to language capital (language proficiency)? That is, is language capital more productive in the labor market among those with more schooling and other forms of human capital? Moreover, does it appear that investments in language capital are greater among those who expect to receive a greater economic return from English-language proficiency? That is, is language proficiency endogenous to the labor market?

A second and inter-related set of issues has to do with the linguistic characteristics of the area in which the respondent lives. Among those who speak a language other than English, does it matter whether they live in an area in which many others speak their origin language (to be referred to as linguistic concentration), as distinct from an area in which few speak this language? While linguistic concentrations have been shown to lower language proficiency (see Chiswick and Miller 1992, 1995, 1998), are they also associated with lower earnings, other things being the same? Controlling for the individual's own characteristics, the cost of "ethnic goods," broadly defined to include social networks, would be lower the greater the extent to which others speak the same origin language. Then, larger linguistic concentrations would be associated with lower reservation wages, and hence lower observed earnings.

Finally, one of the issues discussed in the literature is whether the stage of the business cycle at entry into the labor market affects the earnings of immigrants. It has been shown by Nakamura and Nakamura (1992) and Stewart and Hyclak (1984) that immigrant earnings are lower among those who enter in a period of high unemployment. Unresolved, however, is whether this is a permanent effect, perhaps due to a lower "quality" immigrant cohort in a recession, or whether this is a temporary disadvantage due to the difficulties new immigrants have in finding a high wage job or a good job match during a recession.

Section 2 presents the model of the earnings of immigrants in the labor market. Particular attention is given to the variables that are of primary focus in this study, namely, the individual's own English language proficiency, the linguistic concentration of the area, and the stage of the business cycle at entry into the labor market. This permits a separation of the effects of the immigrant's own language skills from linguistic concentration effects.

Section 3 describes the data from the 1990 Census of Population of the United States used for this study, as well as the estimating equations. The empirical estimation is presented in Section 4. Various specifications of the variable for the respondent's English language fluency are considered, as are the concentration and business cycle effects, among others. A summary and conclusion, with implications for both immigration policy and assimilation (post-migration) policy closes the paper (Sect. 5).

2. Modelling earnings in the immigrant labor market

The study of earnings in the immigrant labor market has in large part been based on the human capital earnings function (Mincer 1974), which includes schooling and labor market experience, expanded to incorporate the immigrant experience (Chiswick 1978). This equation suggests that variations in earnings across individuals (Y_i) can be explained by variations in the amount of schooling (S_i) individuals have acquired and their labor market experience, in total (T_i) and in the destination (YSM_i). Thus:

$$LnY_i = b_0 + b_1S_i + b_2T_i + b_3T_i^2 + b_4(YSM_i) + b_5(YSM_i)^2 + e_i, \tag{1}$$

where LnY_i is the natural logarithm of the earnings of individual i, schooling (S_i) is a measure of the years of full-time equivalent schooling of the individual, experience (T_i) is the potential length of time the individual has spent in the labor market (in the origin and in the destination), and years since migration (YSM_i) is the number of years in the destination. The experience variables reflect the acquisition of skills through formal and informal training on the job, including merely living in the destination. A quadratic specification is employed as economic theory and prior empirical research have both shown that earnings tend to increase at a decreasing rate with years in the labor market and in the destination (see, for example, Ben-Porath 1967; Mincer 1974; Chiswick 1978). The stochastic error term (e_i) captures the influences of unobserved or unmeasured variables (e.g., motivation, effort, luck) on earnings.

The relationship between earnings and duration of residence is generally held to reflect learning about the institutions of the U.S. labor market, cultural adjustment factors, the development of networks for labor market contacts, and investments in U.S.-specific human capital skills that lead to labor market success (see Chiswick 1978). These effects (investments) are expected to be greatest in the first few years and to diminish with duration in the destination.

Acquisition of skills in the dominant language can be viewed as a form of human capital accumulation. Studies of immigrant language skills in the labor market have tended to treat language skills in the same way as schooling and post-school training. See, for example, McManus et al. (1981); Tainer (1988); Chiswick (1991); Chiswick and Miller (1992, 1995, 1998). Thus, the basic human capital earnings equation can be extended to:

$$LnY_i = b_0 + b_1S_i + b_2T_i + b_3T_i^2 + b_4(YSM_i) + b_5(YSM_i)^2 + b_6LANG_i + v_i, \tag{2}$$

where $LANG_i$, is a measure of the destination language fluency of the individual and v_i is the stochastic error term.

Prior empirical analysis has confirmed that language fluency has an important effect on earnings. Chiswick and Miller (1992), for example, report that among adult foreign-born men in the U.S. 1980 Census of Population, English language fluency is associated with around 17% higher earnings. The earnings advantage associated with fluency in the dominant language in the 1981 Canadian Census is about 12%, while those fluent in English in the Australian 1986 Census have about 8% higher earnings (Chiswick and Miller 1995).[3] Fluency in the dominant language (Hebrew) in Israel (1972 and 1983 Censuses) is associated with about a 12% increase in earnings (Chiswick 1998 and Chiswick and Repetto 2001).

In competitive labor markets, the mobility of workers from low-wage areas to high-wage areas is expected to erode any regional differences in real earnings, up to a cost-of-moving wedge. If there are factors, however, that impede the mobility of individuals across regions, then real earnings variation by region of residence could be observed even beyond short-term disequilibria differentials. In the case of the foreign born, the propensity to cluster in communities formed on the basis of language and ethnicity may be important in influencing inter-regional mobility by affecting the "full cost of living" across areas.[4] These costs would be lower in areas where there are others who speak the same language and observe the same customs.

Consider a situation in which an ethnic group defined by language or some other characteristic has a market basket for goods and services that differs from other groups. These "ethnic goods" could include conventional foods and services, such as ethnic foods and ethnic-specific clothing or household items. More important may be less tangible items such as origin language media (e.g., newspapers, books, radio, television), houses of worship, social clubs, other individuals of the same background to share memories and holidays, and opportunities to interact with others of the same origin in the marriage market. The full price of "ethnic goods" would be lower the larger the size of the ethnic market because of economies of scale and lower networking costs. Then, in a competitive labor market in which internal migration equalizes regional differences in wages, group differences in linguistic and/or ethnic concentrations result in group-specific patterns of regional wage differentials. In other words, if ethnic goods are important, a high wage offer is needed to induce a member of the group to leave an area of concentration (low cost of ethnic goods) to live in an area with few others of the same origin (high cost of ethnic goods). Then, the size of the linguistic concentration is inversely related to observed earnings.

A minority language concentration measure, first introduced in Chiswick and Miller (1992), is used to measure the linguistic concentration. Concentrations of foreign language speakers may have a direct effect on the earnings of individuals through the "ethnic goods" effect. Linguistic concentration may also have an indirect influence on earnings through their negative impact on the acquisition of destination language skills (Chiswick and Miller 1992, 1995, 1998). By including variables for both linguistic concentration and the individual's own language skills in an estimating equation, the relative strengths of these direct and indirect effects may be assessed.

There may also be important interaction effects due to the limitations on internal mobility. Individuals who are not fluent in English, and who live in linguistic concentrations may have higher earnings than their counterparts who, perhaps for family reasons (tied stayers), live in predominately monolingual English-speaking regions. Those not fluent in English may have economic advantages from working in a linguistic minority enclave.

Variables for country of birth are included in the earnings equation for the foreign born to capture the unmeasured differences in the average productivity of immigrants from the various countries of origin, that is, differences in quality not measured in the other variables. Differences in average quality might be expected where the international transferability of skills varies across immigrant groups (for example, immigrants from English-speaking and non-English speaking countries), or where some immigrants are less favorably selected for migration (for example, refugees compared to economic migrants) and so have fewer skills relevant for the destination, *ceteris paribus*. Country of origin differences can also arise if the backgrounds of immigrant groups differ. For example, immigrants from Mexico are disproportionately from rural communities. The skills accumulated in the labor markets of the rural communities of developing economies may be less useful in the U.S. labor market than the pre-immigration skills of immigrant groups that are predominately from urban areas and developed countries. Moreover, expected rates of re-migration differ across origins (Ahmed and Robinson 1994), and the propensity to invest in destination-specific skills, including language skills, would decrease with a higher expected return migration or emigration propensity.[5] Finally, the birthplace, variables will also capture differences in the effects of culture and discrimination on measured earnings.

Citizenship has been shown to affect earnings (Chiswick and Miller 1992). Naturalization generally requires the demonstration of at least a minimum level of English language fluency. It also reflects a stronger commitment to the United States which would be associated with greater investment in U.S.-specific human capital, and thus higher earnings. Moreover, aliens are disadvantaged in the labor market as some jobs require citizenship status, or because of illegal discrimination against lawful resident aliens.

To measure the effect of the business cycle at entry on immigrant earnings, the analysis includes the adult male unemployment rate in the United States in the year of arrival or of entry into the U.S. labor market, whichever is later. This variable permits an examination of whether labor market entrance at a time of high unemployment has a negative effect on an immigrant's future labor market success.[6]

Interacting the unemployment rate at labor market entry with the duration in the United States will indicate whether the disadvantage due to the state of the labor market at the time of labor market entrance diminishes with duration in the destination. This will provide insights into the origins of this particular wage effect. For example, if the wage disadvantage dissipates with duration, the effect is a "temporary blemish" rather than a "permanent scar". The temporary blemish may occur if arrival in a period of high employment makes it more difficult for the new immigrant to obtain any employment or to obtain employment that makes maximum use of the new immigrant's transferable skills. The adverse effects of a poor initial placement would be expected to diminish with duration in the destination.

If the wage effect persists, however, attention needs to be focused on permanent factors for the explanation: either the initial placement in a recession results in a permanent disadvantage, or, more likely, it is the result of negative selection for immigration during recessions. Negative selection for migration during recessions might occur if factors other than economic advantage are relatively more important in the migration decision when job opportunities in the destination are scarce. In such circumstances, family reunification and refugee motivations may be relatively more prevalent in migration flows than purely economic motivations.

3. The data and methodology

The data are from the 1990 Census of Population, Public Use Microdata Sample, and are for the 5% sample of the foreign-born men from non-English speaking countries.[7] The analysis is limited to adult non-aged men to avoid the complexities of modelling labor supply in analyses for women and aged men. The variables are defined in detail and the means and standard deviations are reported in Appendix 6A.

The regression equation used in the empirical analysis relates the natural logarithm of annual (1989) earnings for adult (age 25 to 64) men to a set of explanatory variables. These variables (with their hypothesized partial effects in parentheses) include: years of education (+), total years of labor market experience (+), duration of residence of immigrants in the U.S. (+), English language fluency (+), U.S. citizenship (+), married (+), weeks worked (+), the adult male unemployment rate at the time of arrival or entry in the U.S. labor market (–), the interaction of this unemployment rate variable and duration of residence (+), the minority language concentration measure (–), and the interaction between the minority language concentration measure and the individual's own English language fluency (–). Country of origin dichotomous variables are also included, with Western Europe as the benchmark. Three additional dichotomous variables are unity for those living in a rural area (–), those living in a southern state (–) and those whose race is Black (–).

Separate analyses of earnings are conducted for those fluent in English and for those who lack English-language fluency, using both OLS and selectivity correction techniques.[8] In addition, the data are disaggregated by major birthplace groups, and earnings equations are estimated for each subsample.

4. The earnings of immigrants

The empirical analysis begins with the estimations for foreign-born adult men (Tables 6.1 and 6.2). Table 6.1 reports estimates obtained when the data are pooled across individuals fluent in English and those who lack this skill. A number of specifications of the earnings equation are reported in Table 6.1.[9] The first specification is a standard immigrant earnings function without the variables for English-language skills, the second includes a variable for minority language concentration, the third model adds a variable for the individual's fluency in English, the fourth considers the interaction between English language fluency and the minority language concentration measure, and the fifth augments the estimating equation with information on the unemployment rate in the year of entry into the U.S. labor market. The final model in Table 6.1 explores the effects on earnings of a more detailed measure of English language fluency.

The estimates presented in Table 6.2 are for the separate samples of individuals fluent in English (speaks only English at home or speaks English "very well" or "well") and individuals with English-language deficiencies, that is, they speak English "not well" or "not at all." The separation of the data by English language fluency is not random. Regression equations using non-random samples will result in biased and inconsistent estimates. This problem may be accommodated by employing the methodology advanced by Heckman (1979) that treats the sample selectivity bias as an omitted variables problem. Hence, the equations in columns (ii) and (iv) of Table 6.2 include the coefficient on the sample selectivity (lambda) variable that has been constructed according to Lee (1983). In forming this variable, a reduced form model of fluency in English is estimated that has a large number of explanatory variables, including educational attainment, age, duration of residence, marital status, location, minority language concentration, veteran status, linguistic distance between English and the immigrant's mother tongue, and the geographical

Table 6.1 Regression estimates of earnings equations, adult foreign-born men, 1990

Variable	*(i)*	*(ii)*	*(iii)*	*(iv)*	*(v)*	*(vi)*
Constant	5.056	5.066	5.009	4.969	5.112	5.345
	(172.91)	(173.20)	(171.69)	(167.46)	(157.43)	(163.70)
Education	0.049	0.048	0.045	0.045	0.044	0.043
	(91.15)	(90.21)	(83.07)	(82.99)	(81.61)	(79.09)
Experience (Exp)	0.023	0.023	0.025	0.025	0.023	0.023
	(35.45)	(35.56)	(37.75)	(37.79)	(32.38)	(33.01)
$Exp^2/100$	–0.037	–0.037	–0.038	–0.038	–0.036	–0.036
	(30.90)	(31.03)	(31.82)	(31.78)	(29.36)	(29.67)
Year since migration	0.028	0.028	0.026	0.026	0.022	0.020
(YSM)	(49.79)	(50.25)	(44.81)	(44.78)	(28.30)	(25.36)
$YSM^2/100$	–0.041	–0.041	–0.038	–0.038	–0.035	–0.033
	(32.07)	(32.53)	(29.88)	(29.90)	(25.73)	(24.71)
Log weeks worked	0.968	0.968	0.964	0.964	0.966	0.965
	(135.56)	(135.51)	(134.96)	(134.99)	(134.87)	(134.76)
Married	0.214	0.214	0.210	0.210	0.209	0.210
	(55.30)	(55.57)	(54.36)	(54.52)	(54.11)	(54.36)
Citizen	0.087	0.084	0.077	0.075	0.077	0.074
	(20.66)	(20.02)	(18.19)	(17.85)	(18.14)	(17.65)
Race (black)	–0.185	–0.194	–0.200	–0.201	–0.201	–0.206
	(12.60)	(13.16)	(13.62)	(13.67)	(13.71)	(14.05)
Rural	–0.037	–0.043	–0.044	–0.045	–0.045	–0.048
	(4.68)	(5.54)	(5.59)	(5.76)	(5.77)	(6.18)
South	–0.113	–0.110	–0.111	–0.112	–0.112	–0.114
	(26.40)	(25.71)	(26.07)	(26.31)	(26.20)	(26.58)
Birthplace						
S. Europe	–0.056	–0.053	–0.051	–0.045	–0.051	–0.043
	(5.59)	(5.21)	(5.02)	(4.49)	(5.01)	(4.22)
E. Europe	–0.072	–0.072	–0.069	–0.064	–0.067	–0.057
	(5.97)	(6.03)	(5.71)	(5.32)	(5.52)	(4.69)
USSR	–0.130	–0.131	–0.123	–0.118	–0.122	–0.111
	(7.18)	(7.23)	(6.82)	(6.53)	(6.73)	(6.10)
Indochina	–0.276	–0.278	–0.266	–0.256	–0.255	–0.236
	(22.74)	(22.90)	(22.00)	(21.06)	(20.91)	(18.99)
Philippines	–0.229	–0.223	–0.233	–0.232	–0.234	–0.235
	(21.48)	(20.94)	(21.95)	(21.79)	(21.99)	(21.68)
China	–0.266	–0.263	–0.245	–0.235	–0.238	–0.223
	(23.46)	(23.19)	(21.60)	(20.68)	(20.86)	(19.28)
South Asia	–0.013	–0.015	–0.021	–0.021	–0.026	–0.032
	(1.10)	(1.32)	(1.82)	(1.79)	(2.29)	(2.73)
Other Asia	–0.195	–0.197	–0.198	–0.196	–0.202	–0.191
	(10.13)	(10.22)	(10.28)	(10.17)	(10.50)	(9.91)
Korea	–0.229	–0.230	–0.206	–0.195	–0.195	–0.177
	(14.69)	(14.73)	(13.25)	(12.51)	(12.50)	(11.30)
Japan	0.353	0.352	0.362	0.369	0.362	0.379
	(19.06)	(19.01)	(19.69)	(20.03)	(19.69)	(20.55)

(Continued Overleaf)

Table 6.1 Continued

Variable	*(i)*	*(ii)*	*(iii)*	*(iv)*	*(v)*	*(vi)*
Middle East	–0.089 (7.57)	–0.091 (7.71)	–0.096 (8.16)	–0.095 (8.07)	–0.100 (8.47)	–0.100 (8.40)
Sub-Saharan Africa	–0.054 (2.85)	–0.051 (2.68)	–0.057 (3.01)	–0.056 (2.98)	–0.061 (3.22)	–0.066 (3.48)
Mexico	–0.333 (36.66)	–0.238 (22.46)	–0.229 (21.67)	–0.222 (20.94)	–0.228 (21.37)	–0.224 (21.07)
Cuba	–0.235 (20.95)	–0.173 (14.70)	–0.159 (13.54)	–0.148 (12.45)	–0.146 (12.34)	–0.146 (12.36)
C. & S. America (Spanish)	–0.239 (25.09)	–0.172 (16.71)	–0.168 (16.35)	–0.158 (15.28)	–0.161 (15.51)	–0.159 (15.39)
C. & S. Amer, (Non-Spanish)	–0.076 (3.37)	–0.070 (3.12)	–0.089 (3.97)	–0.089 (3.97)	–0.089 (3.98)	–0.099 (4.41)
Minority Language Concentration (CONC)	(a)	–0.006 (17.54)	–0.005 (14.31)	–0.002 (5.09)	–0.002 (5.28)	–0.004 (12.32)
Fluent in English	(a)	(a)	0.144 (30.72)	0.186 (24.71)	0.183 (24.36)	(a)
Fluent in English* CONC	(a)	(a)	(a)	–0.004 (7.88)	–0.004 (7.69)	(a)
Un. Rate Yr. of Labor Market Entry (U)	(a)	(a)	(a)	(a)	–0.018 (10.37)	–0.021 (11.87)
U* Years in US Labor Market/100	(a)	(a)	(a)	(a)	0.063 (5.34)	0.085 (7.14)
Speaks English Very Well	(a)	(a)	(a)	(a)	(a)	0.018 (2.62)
Speaks English Well	(a)	(a)	(a)	(a)	(a)	–0.088 (11.53)
Speaks English Not Well	(a)	(a)	(a)	(a)	(a)	–0.186 (22.06)
Does Not Speak English	(a)	(a)	(a)	(a)	(a)	–0.194 (18.82)
Adjusted R^2	0.4154	0.4162	0.4186	0.4188	0.4191	0.4204
Sample size	212,384	212,384	212,384	212,384	212,384	212,384

Notes: Dependent variable: Natural logarithm of earnings in 1989.
'*t*' statistics in parentheses computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator.
(a) = Variable not entered.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent sample.

distance (and its square) between the major city in the immigrant's country of origin and the nearest major port of entry (New York or San Francisco) into the United States (Chiswick and Miller 1998). The latter three variables are used as the identifying instruments in the selection (English fluency) equation.[10]

Table 6.2 Regression estimates of earnings equations, by English fluency, adult foreign-born men, 1990

Variable	*Fluent in English*		*Not fluent in English*	
	OLS	*Selectivity corrected*	*OLS*	*Selectivity corrected*
Constant	5.001 (125.34)	4.752 (158.04)	5.091 (62.40)	5.706 (64.38)
Education	0.056 (83.15)	0.067 (81.85)	0.017 (18.16)	0.010 (5.69)
Experience (Exp)	0.026 (29.51)	0.021 (24.54)	0.011 (8.60)	0.015 (9.86)
$Exp^2/100$	–0.042 (25.74)	–0.039 (25.86)	–0.020 (9.34)	–0.021 (9.99)
Year since migration (YSM)	0.019 (22.45)	0.025 (28.48)	0.024 (10.80)	0.015 (4.97)
$YSM^2/100$	–0.032 (21.01)	–0.040 (26.84)	–0.040 (11.60)	–0.034 (9.55)
Log weeks worked	0.995 (106.83)	1.006 (213.97)	0.902 (83.18)	0.894 (145.81)
Married	0.225 (48.52)	0.238 (50.16)	0.159 (23.23)	0.150 (20.49)
Citizen	0.079 (16.41)	0.102 (20.33)	0.033 (3.70)	–0.002 (0.18)
Race (black)	–0.214 (13.43)	–0.197 (12.87)	–0.059 (1.66)	–0.094 (2.51)
Rural location	–0.040 (4.20)	–0.040 (4.48)	–0.073 (5.98)	–0.078 (5.58)
South	–0.114 (22.74)	–0.110 (22.04)	–0.114 (14.32)	–0.122 (15.14)
Birthplace				
S. Europe	–0.049 (4.69)	–0.067 (6.50)	–0.089 (1.14)	0.056 (0.76)
E. Europe	–0.070 (5.48)	–0.096 (7.60)	–0.141 (1.77)	0.001 (0.01)
USSR	–0.111 (5.91)	–0.150 (8.21)	–0.321 (3.56)	–0.169 (2.09)
Indochina	–0.229 (17.33)	–0.285 (21.23)	–0.467 (5.95)	–0.298 (3.90)
Philippines	–0.234 (21.66)	–0.208 (18.62)	–0.432 (5.10)	–0.398 (5.37)
China	–0.194 (15.90)	–0.264 (22.26)	–0.550 (7.04)	–0.367 (4.73)
South Asia	–0.036 (3.03)	–0.028 (2.50)	–0.380 (4.50)	–0.332 (4.48)

(Continued Overleaf)

Table 6.2 Continued

Variable	*Fluent in English*		*Not fluent in English*	
	OLS	*Selectivity corrected*	*OLS*	*Selectivity corrected*
Other Asia	–0.198 (9.85)	–0.208 (11.00)	–0.449 (4.86)	–0.349 (4.06)
Korea	–0.200 (11.37)	–0.288 (19.26)	–0.293 (3.64)	–0.097 (1.22)
Japan	0.330 (16.52)	0.281 (15.48)	0.391 (4.58)	0.544 (6.92)
Middle East	–0.098 (8.16)	–0.092 (8.52)	–0.326 (3.78)	–0.264 (3.64)
Sub-Saharan Africa	–0.052 (2.64)	–0.041 (2.29)	–0.345 (3.57)	–0.318 (3.48)
Mexico	–0.208 (18.13)	–0.256 (21.54)	–0.392 (5.06)	–0.236 (3.17)
Cuba	–0.127 (9.86)	–0.174 (13.06)	–0.291 (3.72)	–0.111 (1.43)
C. & S. America (Spanish)	–0.148 (13.51)	–0.176 (16.01)	–0.323 (4.17)	–0.176 (2.40)
C. & S. America (Non-Spanish)	–0.055 (2.40)	0.004 (0.18)	–0.381 (2.63)	–0.446 (2.54)
Minority Language Concentration (CONC)	–0.004 (10.35)	–0.007 (16.16)	–0.008 (13.29)	–0.005 (5.92)
Un. Rate Yr. of Labor Market Entry (U)	–0.024 (11.89)	–0.028 (13.63)	–0.003 (0.58)	–0.009 (1.93)
U*Years in US Labor Market/100	0.096 (7.33)	0.163 (12.30)	0.070 (1.85)	0.109 (2.91)
Lambda	(a)	0.321 (21.36)	(a)	–0.158 (4.51)
Adjusted R^2	0.3752	0.3770	0.3847	0.3849
Sample size	157,725	157,725	54,660	54,660

Notes: Dependent Variable: Natural logarithm of earnings in 1989.
Columns (i) and (iii) are estimated using OLS, columns (ii) and (iv) are estimated using Lee's (1983) estimator.
'*t*' statistics in parentheses for columns (i) and (iii) computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator.
(a) = Variable not entered.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent sample.

The relationships between earnings and the conventional determinants of earnings will be explored first, and comparisons made between the estimated impacts among individuals who are fluent and those who are not fluent. Then the variables that are of particular interest for this study are discussed.

Additional years of education are associated with around five percent higher earnings when the focus of analysis is the total foreign-born sample (column (i) of Table 6.1). This impact is similar to that recorded in analyses of the 1980 Census (see Chiswick and Miller 1992). When the earnings equation is estimated separately for individuals fluent in English and for those lacking this fluency (Table 6.2, columns ii and iv), the coefficient of schooling is 6.6% among the fluent, but only 1.0% for those who are not fluent in English.

These findings suggest there is a considerable degree of complementarity in the labor market between English-language skills and formal education.[11] In other words, in the United States it is difficult to reap a return to human capital acquired through formal education unless one can speak English. Acquiring English language fluency therefore appears to be a means of increasing the international transferability of previously acquired forms of human capital. This provides a greater economic incentive for the better educated to become fluent in English than their less educated counterparts. The economic incentive also helps explain the very low mean educational attainment among those who are not fluent in English, only 7.9 years of schooling compared to 13.1 years among immigrants fluent in English.

The relative importance of U.S.-specific skills shows up clearly in the analysis of the effect of labor market experience on earnings. The two sets of relevant variables are years of labor market experience and years since migration. The experience variable reflects the impact of labor market experience accumulated prior to migration, while the years since migration variable reflects the premium to labor market experience accumulated after arrival in the United States. For the total foreign-born sample, earnings increase at a decreasing rate with years of pre-immigration experience. When evaluated at 10 years of experience, the earnings growth per year of experience is 1.6% (Table 6.1, column i). The earnings growth with additional years of pre-immigration experience is greater among immigrants who are fluent in English than for those that lack English-language fluency (see Table 6.2).[12]

The earnings premium for a longer duration of residence in the United States is 2.0% per year when evaluated at 10 years of residence for the foreign-born men (Table 6.1, column i). The separate analyses for the two English-language fluency groups reveal a higher premium to U.S. labor market experience (holding total experience constant) for the group not fluent in English when the estimation is by Ordinary Least Squares (OLS), but the expected lower premium for them when the estimation is by the selectivity correction method.[13] As duration of residence lengthens, additional immigrants move from the "not fluent" to the "fluent" states, and this linguistic mobility affects the estimates of the returns to duration in the destination.

A clear pattern is evident from the comparison of the OLS and selectivity-corrected earnings regressions (Table 6.2). If a variable is associated with high levels of English fluency (e.g., educational attainment and years since migration) then the selectivity-corrected estimate of the wage differential is greater than the OLS estimate for the group that is fluent in English, and this pattern is reversed for the group that is not fluent in English. However, if a variable is associated with low levels of English language fluency (e.g., minority language concentration) then the selectivity-corrected estimate of the wage difference is lower than the OLS estimate for the group that is fluent in English, with this pattern being reversed for the group that is not fluent in English.

The estimates also show that married (spouse present) men earn around 20% more than their non-married counterparts.[14] The favorable effect of being married on male earnings is greater among those who are fluent in English (23%) than for the group lacking in English language fluency (15% higher earnings). Citizens have higher earnings than non-citizens (9% for the total sample), and the effect varies by level of fluency in English, 8% for immigrants fluent in English, 3% for immigrants not fluent in English.

The elasticity of earnings with respect to weeks worked also varies by English language fluency. Overall the elasticity is 0.97, which is statistically significantly lower than unity, although the difference has no economic significance. Among the fluent the elasticity is 1.00, but among the non-fluent it is 0.89. That is, a 10% increase in weeks worked in the year is associated with a 10% increase in annual earnings among those who are fluent, but only a 9% increase (i.e., weekly earning decline with weeks worked) for the non-fluent group. This difference would be consistent with the latter group having a backward bending labor supply curve or greater seasonality in their employment.

The effects of the schooling, experience, marriage, citizenship and weeks worked variables suggest that there is a complementarity between these forms of human capital and English language fluency. Those who are fluent receive larger benefits from having more of these favorable characteristics.

Earnings also differ appreciably across birthplace groups, even when account is taken of the individual skills and demographic characteristics. Compared to the benchmark group, immigrants from Western Europe, other immigrants essentially fall into four categories (see Table 6.1): immigrants from Japan who have earnings about 35% greater than the benchmark; immigrants from South Asia whose earnings are not significantly different from the earnings of immigrants from Western Europe; immigrants from countries with earnings 5 to 15% below the earnings of the Western Europeans (Sub-Saharan Africa, Southern Europe, Eastern Europe, the non-Spanish regions of Central and South America, Middle East, USSR); those from countries with earnings more than 15% below that of the benchmark group (Other Asia, Korea, Cuba, Philippines, the Spanish-speaking regions of Central and South America, China, Indochina, Mexico). The earnings disadvantage, *ceteris paribus*, is over 30% for immigrants from Mexico. Note also the tendency for refugee groups (e.g., Cuba, China, Indochina) to have lower earnings even when other measured variables are the same.[15]

The specification in column (ii) of Table 6.1 includes the variable for minority language concentration. Residence in a state that contains a larger concentration of individuals speaking the same foreign language as the respondent is associated with lower earnings. Where 10% of the population speak the same foreign language as the respondent, compared to an area where none speak this language, earnings are lower by 6%. Where 20% of the population speak the same foreign language as the respondent (as is the case of Spanish in a number of states), earnings are lower by 12%.

The addition of the minority language concentration variable reduces the earnings disadvantage of the three Spanish-speaking birthplace groups: Mexico, Cuba, and the Spanish speaking regions of Central and South America. For immigrants from Mexico, the apparent "birthplace" disadvantage is reduced by 10 percentage points to about 24%. For immigrants from Cuba and Central and South America, the reduction in the apparent birthplace effect is six percentage points. Part of the Hispanic birthplace effect on earnings, therefore, arises from Hispanics concentrating in areas where their earnings are lower because of the presence of many other Spanish speakers.

The variable for the individual's fluency in English is added to the model presented in Table 6.1, column (iii). Fluency in English is associated with 14% higher earnings, *ceteris paribus*. This increment in earnings is slightly lower than that recorded in study of the 1980 Census (17% higher earnings), but the difference is at the margin of statistical significance ('*t*' = 1.75). Hence, whereas there has been a rise in the return to language skills in the Australian labor market over the past few years, this is not the case in the United States or in Israel or Canada (Chiswick and Miller 1995, 2000 [2003]; Chiswick and Repetto 2001).

Inclusion of the variable for the individual's English-speaking skills is associated with a small reduction in the estimated impact of the minority language concentration variable (from –0.006 to –0.005). Thus, about one-sixth of the effect of living in a concentration of individuals who speak the same non-English second language at home is linked directly to the individual's own lack of English skills. The greater part of the impact is due to other consequences of a labor market characterized by many minority language speakers. Where you work may be an important determinant of earnings.

In terms of its effects on earnings, fluency in English as defined in this study is the equivalent of around three years of formal education and around five years of U.S. labor market experience evaluated at 10 years of experience. Acquisition of English language skills clearly pays in the U.S. labor market. If English speaking fluency can be attained with the equivalent of between 6 months and one-year of full-time training, provided that the remaining working life is reasonably long, this implies a real rate of return to the acquisition of language skills of between 14 and 28%. There would appear to be few other investments that an immigrant could undertake that would yield such a healthy monetary return.

The inclusion of a measure of fluency in English in the estimating equation is associated with some (albeit minor) changes in the other estimated coefficients. For example, the extra earnings associated with additional years of education declines by almost one-third of one percentage point, and the income growth associated with years in the U.S. declines marginally. These changes indicate that part of the extra earnings associated with education and duration in the U.S. in analyses that do not explicitly consider language skills is in fact due to greater ability to communicate in English among the better educated and those who have resided in the U.S. for a longer period of time.

Table 6.1, column (iv) explores interactions between the minority language concentration measure and fluency in English. It is apparent that residence in a region with a concentration of individuals speaking the same foreign language as the immigrant has a more negative impact on earnings among those fluent in English. The impact of minority language concentration is only –0.002 for those who are not fluent in English, but is –0.006 (= –0.002 + –0.004) for those who are fluent in English. That is, those who are fluent have a larger improvement in their earnings than their non-fluent counterparts if they move to a low concentration area.

Another perspective on these results is gained by focusing on the effect on earnings of fluency in English. This is estimated to be close to 19% for an individual who lives in an area where his origin language is not spoken. Where 20% of the population speak the immigrant's origin language, the return to English-speaking skills would only be 11%. Thus, while possession of English-speaking fluency is important, the language characteristics of the labor market also appear to matter. The favorable effect on earnings of English language fluency is greater in areas in which fewer people speak the worker's origin language. In other words, the economic penalty from not speaking English is smaller among those who live in a linguistic concentration area.

Table 6.1, column (v) uses a linear interaction specification to test the hypothesis that weekly earnings are dependent on the stage of the business cycle at the time of entry into the labor market. Weekly earnings are lower when the unemployment rate is high in the year of U.S. labor market entry. In the linear specification, in the initial year (YRS = 0) the effect is nearly 1.8% lower weekly earnings for each one percentage point increase in the unemployment rate. For example, if the unemployment rate among adult men in the year of labor market entry is 8% (a deep recession), then weekly earnings are initially reduced by 9% compared to a situation where there was a 3% unemployment rate among adult men (full employment).[16] The interaction term between the unemployment rate variable and duration in the U.S. shows, however, that over time this disadvantage is gradually removed. Hence, while individuals who entered the labor market in times of 8% and 3% adult male unemployment rates would initially differ in average earnings by 9%, this earnings gap is 6% after 10 years and is closed after about 30 years of U.S. labor market activity.[17]

A linear interaction specification is arbitrary and alternative specifications were tested to see which would offer the greater explanatory power. The most successful simple specification had the squared unemployment rate in the interaction term, rather than the unemployment rate itself.[18] This increased the adjusted *R*-square from 0.4191 to only 0.4192. In this specification weekly earnings in the initial year (YRS = 0) are lower by 2.9% for each one percentage point increase in the adult male unemployment rate at entry into the U.S. labor market.[19] This implies that initial weekly earning would be lower by 14.5% if a male immigrant entered the United States labor market when the adult male unemployment rate was 8% compared to entry at 3%. The earnings gap as a function of the unemployment rate at entry decreases the longer the duration of residence. At ten years in the United States and a mean unemployment rate of 5.1% the partial effect of a one percentage point increase in the unemployment rate at entry is 1.3% in contrast to the 2.9% in the initial year. The effect disappears (at the mean unemployment rate) at about 18 years in the U.S. labor market.[20]

The analysis of the effect of the stage of the business cycle at entry on subsequent weekly earnings indicates that conditions at entry do matter. Earnings among adult male immigrants are substantially lower initially if they enter during a recession rather than during a period of full employment. This is not a permanent scar or disadvantage. The earnings disadvantage from entering in a recession diminishes with duration and eventually disappears. Thus, entering the U.S. labor market has two disadvantages for immigrants, one is the higher unemployment rate (lower weeks worked in the year) and the other is the lower weekly earnings.

The analyses in Table 6.2 show that the pattern of an initial wage disadvantage from arriving in a year of high unemployment followed by a gradual catch-up pertains only to individuals who are fluent in English. Among those who are not fluent in English, there is no initial earnings disadvantage associated with the stage of the business cycle at entry, nor does the effect of the business cycle vary with the duration of residence. This too suggests that among immigrants who are not fluent in English, there is a form of positive selection for migration during recessions. Perhaps those who are not fluent but who migrate when jobs are scarce have family ties that arrange employment.

The final column of Table 6.1 contains variables that record in finer detail the English-language skills of immigrants who speak a language other than English at home. The first of these variables indicates the wage differential between immigrants who are monolingual English speakers at home (the benchmark group for this analysis, whose English language proficiency is not known) and immigrants who speak a language other than English at home and report they speak English very well. The latter immigrants are assumed to be fully bilingual, and they are shown to receive a small, statistically significant wage premium, of about 2% ('*t*' = 2.6). The second variable identifies individuals who speak a language other than English at home and who self-report that they speak English "well". This group has earnings 9% lower than monolingual English speakers. The final two variables are for individuals who speak a language other than English at home and are not fluent in English. Immigrants in this situation have earnings almost 20% lower than the benchmark group of monolingual English speakers. Individuals who speak English "not well" earn about the same as those who do not speak English at all, and there may be little difference in the English proficiency of these two groups. These results show that the labor market is quite discerning when it comes to language skills.

The statistical significance of the lambda terms in the selectivity-corrected estimates in Table 6.2 indicate that the samples separated on the basis of language fluency are not random. Moreover, the sign of the coefficient on lambda is positive for the sample that is fluent in English and negative for the sample that is not fluent in English.[21] As the selection term in the "not fluent" category is constructed to be negative, a negative coefficient indicates an observed conditional mean that exceeds the population mean. That is, a negative coefficient in the "not fluent" group implies positive selection into that group. This indicates that there is positive selection into each state. In other words, the unobserved characteristics that lead to an individual having a higher than expected level of fluency in English are also associated with a higher than expected earnings in the labor market for individuals who are fluent in English. Moreover, those who have higher than expected labor market earnings for individuals who are not fluent in English are more likely to select into this state than other individuals (i.e., they have a lower than expected level of English language fluency).

This study of the correlation between the residuals in the earnings equations and the model of English-language fluency therefore suggests that English-speaking skills are endogeneously determined in the model of earnings. That is, a worker's language skills are influenced by expectations with respect to income returns.[22]

Table 6.3 presents selected statistics from analyses of earnings for the major birthplace regions. The first column lists the mean logarithm of earnings for each birthplace region. The second column treats immigrants from Western Europe as the reference group and reports the differences between the mean logarithm of earnings for each birthplace region and the reference group. These figures range from a high of 0.132 for Japan (14% higher earnings) to a low of –0.922 for Mexico (60% lower earnings). Other relatively low earnings origins are Indochina, Central and South America, Sub-Saharan Africa, Cuba, China and Korea. The data in the third column are the birthplace effects, other variables held constant, from column (i) of Table 6.1. Comparison of the data in columns (ii) and (iii) of Table 6.3 reveals the extent to which factors such as differences in educational attainments, years of labor market experience, duration in the U.S., and location account for the unadjusted earnings differences measured in the Census data. In many cases the narrowing of the earnings differentials between the unadjusted and adjusted figures is pronounced. For example, in the case of immigrants from Southern Europe, the earnings differential falls by over two-thirds when adjustments are made for the impact of the determinants of earnings other

Table 6.3 Selected statistics for analyses of earnings for birthplace regions, adult foreign-born men, 1990

Birthplace region	*Mean log of earnings*	*Unadjusted earnings difference*	*Adjusted earnings difference*	*Fluency rate*	*Returns to fluency*	*Returns to bilingualism*	*Impact of minority language concentration*	*Mean level of education*	*Coefficients of education*	*Sample size*
	(i)	*(ii)*	*(iii)*	*(iv)*	*(v)*	*(vi)*	*(vii)*	*(viii)*	*(ix)*	*(x)*
Western Europe	10.320	–	–	0.990	0.136	–0.016	0.003	14.42	0.078	13,283
					(1.73)	(0.94)	(0.65)		(25.91)	
Southern Europe	10.116	–0.204	–0.056	0.849	0.071	–0.058	0.001	10.96	0.035	16,899
		(18.01)	(5.59)		(4.00)	(2.95)	(0.71)		(18.01)	
Eastern Europe	10.121	–0.199	–0.072	0.859	0.102	0.041	0.008	13.50	0.050	7,467
		(13.93)	(5.97)		(3.82)	(1.34)	(1.09)		(15.71)	
USSR	10.007	–0.313	–0.130	0.841	0.148	–0.050	0.001	14.43	0.058	2,776
		(13.13)	(7.18)		(2.45)	(0.96)	(0.09)		(9.87)	
Indochina	9.647	–0.673	–0.276	0.722	0.161	–0.032	–0.013	11.95	0.025	8,747
		(48.65)	(22.74)		(7.12)	(0.55)	(1.13)		(9.68)	
Philippines	9.949	–0.371	–0.229	0.962	0.071	0.047	–0.011	14.32	0.073	11,508
		(30.26)	(21.48)		(1.70)	(1.79)	(3.60)		(14.32)	
China	9.838	–0.482	–0.266	0.742	0.241	–0.069	0.005	14.15	0.052	12,763
		(36.03)	(23.46)		(10.24)	(1.60)	(0.30)		(21.55)	
South Asia	10.153	–0.167	–0.013	0.959	0.055	–0.072	–0.007	15.83	0.082	10,502
		(12.32)	(1.10)		(1.36)	(2.68)	(1.01)		(25.27)	
Other Asia	9.937	–0.383	–0.195	0.918	0.154	–0.041	0.041	14.80	0.079	2,389
		(15.97)	(10.13)		(2.37)	(0.80)	(1.99)		(14.80)	
Korea	9.859	–0.461	–0.229	0.695	0.188	0.065	–0.010	14.45	0.036	6,438
		(26.42)	(14.69)		(5.50)	(0.97)	(0.43)		(14.45)	
Japan	10.452	0.132	0.353	0.798	0.270	0.027	–0.007	15.14	0.064	3,043
		(6.55)	(19.06)		(6.41)	(0.46)	(0.89)		(10.17)	

(Continued Overleaf)

Table 6.3 Continued

Birthplace region	*Mean log of earnings*	*Unadjusted earnings difference*	*Adjusted earnings difference*	*Fluency rate*	*Returns to fluency*	*Returns to bilingualism*	*Impact of minority language concentration*	*Mean level of education*	*Coefficients of education*	*Sample size*
	(i)	*(ii)*	*(iii)*	*(iv)*	*(v)*	*(vi)*	*(vii)*	*(viii)*	*(ix)*	*(x)*
Middle East	10.035	–0.265 (21.02)	–0.089 (7.57)	0.952	0.144 (3.28)	0.004 (0.15)	–0.001 (0.19)	14.72	0.074 (24.67)	12,062
Sub-Saharan Africa	9.801	–0.519 (30.08)	–0.054 (2.85)	0.971	0.072 (1.07)	–0.088 (2.85)	0.001 (0.15)	15.09	0.077 (14.81)	4,413
Mexico	9.387	–0.932 (96.17)	–0.333 (36.66)	0.535	0.145 (23.11)	0.095 (4.86)	–0.002 (5.45)	7.90	0.025 (27.09)	61,700
Cuba	9.841	–0.479 (36.50)	–0.235 (20.95)	0.710	0.087 (3.88)	–0.114 (2.91)	–0.004 (2.16)	12.01	0.051 (18.59)	10,859
Central & South America (Spanish)	9.647	–0.673 (62.96)	–0.239 (25.09)	0.683	0.150 (13.15)	0.094 (4.03)	–0.009 (14.04)	11.54	0.041 (29.01)	25,926
Central & South America (non-Spanish)	9.828	–0.491 (21.10)	–0.076 (3.37)	0.989	0.126 (1.74)	0.126 (1.74)	–0.005 (1.00)	12.56	0.034 (5.56)	1,601

Notes: Column (i) is the mean logarithm of earnings by country of birth, column (ii) is the difference between the country's mean log earnings and that of immigrants from Western Europe (10.320), while column (iii) is the birthplace regression coefficient from Table 6.1, column (i). Column (iv) is the observed fluency rate and column (viii) is the mean level of education for the birthplace group. Columns (v), (vii) and (ix) are regression coefficients estimated from equations similar to Table 6.1 column (iii) for each birthplace separately. Column (vi) is the partial effect of speaking English "very well" from an equation similar to Table 6.1, column (vi), estimated for each birthplace separately. Column (x) is the sample size for each birthplace group.
'*t*' statistics in parentheses computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent sample.

than birthplace. Similarly, for immigrants from Mexico, the earnings differential falls by over two-thirds, while in the case of immigrants from Cuba, it is reduced by one-half.

Table 6.3, column (iv) reports information on the mean fluency rate for each birthplace region, while column (v) lists the partial effect of fluency on earnings for each birthplace. This is computed from an equation with the same structure as Table 6.1, column (iii) estimated separately for each birthplace. It is only among immigrants from South Asia, Sub-Saharan Africa and from non-Spanish speaking parts of Central and South America that fluency in English is not associated with significantly higher earnings in the U.S. labor market. These groups have very high levels of English language fluency (96%, 97% and 99%, respectively); nearly all of the respondents are fluent in English. Among the other birthplace regions, the effect on earnings associated with fluency in English ranges from lows of 7% for Southern Europe and the Philippines, to highs of 14 to 16% for Western Europe, USSR, Indochina, Other Asia, Middle East, Mexico and the Spanish-speaking regions of Central and South America.

The data were also examined for the returns to "bilingualism". That is, do those who speak a language other than English at home and who report they speak English "very well" have higher earnings than monolingual English speakers? Table 6.3, column (vi) reports the coefficient on the "very well" variable for equations with the same structure as Table 6.1, column (vi), but estimated separately for each birthplace group. In only four birthplace groups is there evidence of the labor market rewarding this skill, namely the Philippines (5% higher earnings), Mexico (10% higher earnings), the Spanish-speaking regions of Central and South America (also 10% higher earnings) and the non-Spanish speaking regions of Central and South America (13% higher earnings).

The impact of the minority language concentration variable on earnings within country of birth categories is examined in the seventh column of Table 6.3, estimated from birthplace-specific equations. This variable has negative and significant impacts on earnings for the three Spanish-speaking birthplace groups, and also for the Philippines. The size of the estimated coefficients for the Spanish-speaking regions are interesting: the smallest partial effect is estimated for immigrants from Mexico and the largest partial effect is estimated for immigrants from the Spanish speaking regions of Central and South America. The coefficient is negative, although not statistically significant, for most of the Asian countries.[23] The coefficient is positive, although not statistically significant, for the European countries. Thus, there is a clear negative effect for Spanish and Tagalog, with weaker negative effects for most of the other Asian languages. This is consistent with the ethnic goods model as the cultural characteristics of the Hispanic and Asian immigrants differ far more from the American mainstream than that of the European immigrants.

The final two columns of estimates in Table 6.3 give the mean level of schooling and the estimated partial effect of schooling on earnings, estimated from birthplace-specific equations similar to Table 6.1, column (iii). On the basis of the previous study of the determinants of English language fluency, the positive relationship between the mean level of education and the fluency rate is to be expected. What is surprising, however, is the lack of association between the returns to fluency and the returns to education. Examination of the correlation coefficient between the two returns shows a weak negative association (correlation coefficient –0.25). While there are obvious complementarities between the two types of skills (see in particular Table 6.2), they have sufficiently different impacts on earnings that separate analyses of each is required.

5. Summary, conclusions and implications

This study of the determinants of earnings among adult foreign-born men from non-English speaking countries uses the microdata files from the 1990 Census of Population (5% sample). The study focuses on the effects on earnings of the respondent's own English language skills, the extent to which others who live in the area in which the respondent lives speak the same foreign language, and the stage of the business cycle at the time of entry into the U.S. labor market.

The analysis demonstrates the importance for earnings of English language fluency. *Ceteris paribus*, the foreign born from non-English speaking countries who are fluent in English earn about 14% more than those lacking this fluency. This is about the same effect as that observed in the 1980 Census (17%). The earnings effect is found for immigrants from all non-English speaking countries of origin.

The effect of fluency varies with the extent to which others in the area where the respondent lives speak the same non-English language. Earnings are lower among the foreign born the greater is the intensity of the linguistic concentration. The adverse effect of living in an area with a greater concentration of origin-language speakers is more pronounced among those immigrants more fluent in English. That is, compared to those who are fluent, those who are not fluent in English have relatively greater earnings opportunities inside the linguistic concentration area than outside such an area.

There is strong evidence of endogeneity between language and earnings, as found in the selectivity correction analysis. Those who are fluent are found to be favorably selected for higher earnings in this sector, while the same is true for the unmeasured characteristics of those who are not fluent. Thus, as with other forms of human capital, language skills are acquired, in part, in response to the expected increment in earnings.

There is also strong evidence of complementarity among forms of human capital. The partial effects on earnings of schooling, total labor market experience, duration in the U.S., being married, being a citizen, and weeks worked in the year are greater among those more fluent in English, and are generally very low among those lacking fluency.

Among the foreign born, the stage of the business cycle on entry into the U.S. labor market does influence current earnings overall and among those who are fluent. Earnings are lower for those who enter during a period of high unemployment. Yet this effect is not a permanent scar as the adverse effect of a high unemployment rate at entry diminishes with duration in the U.S. Among those who are not fluent, however, the stage of the business cycle at entry has no effect on earnings.

This chapter highlights the important relationship between earnings and English language fluency among the foreign born. English language fluency enhances earnings, and at the same time is itself, in part, a consequence of the expectation of higher earnings. Where one lives matters as earnings are higher, especially for those fluent in English, among those who do not live in a linguistic concentration area.

These findings have important implications for public policy. They emphasize the value to be had from including English language fluency, or the correlates of English language fluency, in the criteria for rationing immigration visas. They also point to the value of encouraging immigrants to participate in English-as-a-second-language programs either prior to or after arrival in the United States.

Those who live in linguistically concentrated areas have lower earnings, even after controlling for their own language skills, among other variables. This may be reflecting an "ethnic goods" effect, that is, that immigrants sort themselves across the country to equalize real incomes and that "ethnic goods", including community ties and networking, have a lower cost the greater the concentration of people speaking the same origin language. If so, the regional wage differentials may merely reflect ethnic-group specific cost of living differentials, rather than a "crowding effect".

The finding of a temporary adverse effect on earnings from entering the U.S. labor market in a period of high unemployment will add fuel to the debate on whether business cycle considerations should be explicitly included in the allocation of immigration visas. This finding may reflect the temporary negative labor market effects of an initial poor job placement of those who came in a recession, or a lesser degree of skill transferability among those who arrive in a recession as they rely more heavily on family ties. Yet the adverse effect is temporary as it diminishes and eventually disappears the longer an immigrant has lived in the United States.

Appendix 6A: Definitions of variables

The variables used in the statistical analyses are defined below. Mnemonic names are also listed where relevant. The means and standard deviations are reported in Appendix Table 6A.1 for the total sample and separately for the fluent and not fluent groups.

Table 6A.1 Means and standard deviations of variables[a]

Variable	*Total sample*	*Fluent in English*	*Not fluent in English*
Annual earnings	27061 (28589)	31092 (31013)	15128 (14110)
Log of annual earnings	9.787 (1.03)	9.954 (1.01)	9.292 (0.94)
Education	11.79 (4.92)	13.10 (4.20)	7.926 (4.84)
Experience	22.76 (11.46)	21.46 (10.98)	26.58 (12.00)
Years since migration	15.43 (10.85)	16.95 (11.22)	10.92 (8.13)
Log weeks worked	3.75 (0.47)	3.78 (0.44)	3.66 (0.55)
Married	0.673 (0.47)	0.696 (0.46)	0.604 (0.49)
Citizen	0.417 (0.49)	0.494 (0.50)	0.192 (0.39)
Race (black)	0.033 (0.18)	0.041 (0.20)	0.009 (0.09)
Rural	0.057 (0.23)	0.057 (0.23)	0.058 (0.23)
South	0.240 (0.43)	0.237 (0.43)	0.247 (0.43)
Southern Europe	0.078 (0.27)	0.089 (0.28)	0.047 (0.21)
Eastern Europe	0.036 (0.19)	0.041 (0.20)	0.020 (0.14)
USSR	0.013 (0.12)	0.015 (0.12)	0.008 (0.09)
Indochina	0.041 (0.20)	0.040 (0.20)	0.046 (0.21)
Philippines	0.053 (0.22)	0.068 (0.25)	0.008 (0.09)
China	0.061 (0.24)	0.060 (0.24)	0.062 (0.24)
South Asia	0.051 (0.22)	0.066 (0.25)	0.008 (0.09)
Other South-East Asia	0.011 (0.11)	0.014 (0.12)	0.004 (0.06)
Korea	0.031 (0.17)	0.029 (0.17)	0.038 (0.19)
Japan	0.015 (0.12)	0.016 (0.12)	0.012 (0.11)

Middle East	0.059 (0.24)	0.075 (0.26)	0.011 (0.10)
Sub-Saharan Africa	0.024 (0.15)	0.032 (0.18)	0.003 (0.05)
Mexico	0.279 (0.45)	0.200 (0.40)	0.514 (0.50)
Cuba	0.050 (0.22)	0.048 (0.21)	0.057 (0.23)
C. & S. America (Spanish)	0.127 (0.33)	0.116 (0.32)	0.160 (0.37)
C. & S. America (non-Spanish)	0.009 (0.09)	0.012 (0.11)	0.001 (0.01)
Unemployment rate at time of labor market entry	5.14 (1.59)	5.06 (1.63)	5.39 (1.43)
Minority language concentration (CONC)	7.44 (8.93)	5.65 (8.20)	12.76 (8.89)
Fluent in English	0.748 (0.43)	1.00	0.000
Speaks English very well	0.369 (0.48)	0.493 (0.50)	0.000
Speaks English well	0.268 (0.44)	0.358 (0.48)	0.000
Speaks English not well	0.191 (0.39)	0.000	0.755 (0.43)
Speaks English not at all	0.062 (0.24)	0.000	0.245 (0.43)
Sample size	212,384	157,725	54,660

Note: (a) The data are for men who worked and had earnings in 1989, were 25 to 64 years in 1990 and were born in non-English speaking countries.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent sample.

Data source. 1990 Census of Population, Public Use Microdata Sample, 5% sample of the foreign born, except where noted otherwise.

Definition of population. Except where otherwise stated, foreign-born men aged twenty-five to sixty-four with earnings in 1989, born in countries other than the English-speaking developed economies (UK, Ireland, Canada, Australia, New Zealand).

Dependent variable

Earnings (LNEARN). The natural logarithm of the sum of wage or salary income and self-employment income (either non-farm or farm) received in 1989.

Explanatory variables

English language fluency (LANG). LANG is set equal to one for individuals who speak only English at home, or if a language other than English is spoken in the home, who speak English either "very well" or "well". The variable is set to zero where a language other than English is spoken in the home and the respondent speaks English either "not well" or "not at all". The categories "very well", "well", "not well" and "not at all" are separately identified in the more extensive specifications of language skills included in some models.

Years of education (EDUC). This variable records the total years of full-time education. It has been constructed from the Census data on educational attainment by assigning the following values to the Census categories: completed less than fifth grade (2.5 years); completed fifth through eighth grade (7 years); completed ninth grade (9); completed tenth grade (10); completed 11th grade (11); completed 12th grade or high school (12); attended or completed college (14); Bachelor's degree (16); Master's degree (17.5); Professional degree (18); Doctorate (20).

Years of experience (EXP). This is computed as age minus years of education minus 5 (that is, EXP = AGE – EDUC – 5). A quadratic specification is used.

Years since migration (YSM). The categorical Census information on year of immigration is converted to a continuous measure using the following values: 1987–1990 (1.75 years); 1985–1986 (4.25 years); 1982–1984 (6.75 years); 1980–1981 (9.25 years); 1975–1979 (12.75 years); 1970–1974 (17.75 years); 1965–1969 (22.75 years); 1960–1965 (27.75 years); 1950–1959 (35.25 years); before 1950 (49.75 years).

Birthplace (BIRTH). A number of non-English speaking birthplace regions are considered in the analyses: Western Europe; Southern Europe; Eastern Europe; former Soviet Union; China; Indochina; Philippines; South Asia (which comprises the regions of British influence, for example, India, Nepal, Pakistan); Other (South-East) Asia; Korea; Japan; Middle East and North Africa; SubSaharan Africa; Mexico; Cuba; Central and South America (Spanish influence); Central and South America (non-Spanish influence). The benchmark group (omitted category) in the regression analysis for the foreign born is Western Europe.

Log of weeks worked (LNWW). The number of weeks worked in 1989 is used in natural logarithmic form.

Minority language concentration (CONC). Each respondent is assigned a measure equal to the percentage of the population aged eighteen to sixty-four in the state in which he lives who reports the same non-English language as the respondent. In the construction of this variable, only the twenty-four largest language groups nationwide are considered. In descending order there are: Spanish; French; German; Italian; Chinese; Tagalog; Polish; Korean; Vietnamese; Japanese; Portugese; Greek; Arabic; Hindi; Russian; Yiddish; Thai; Persian; French Creole; Armenian; Hebrew; Dutch; Hungarian; Mon-Khmer (Cambodian). These constitute 94 percent of all responses that a language other than English is used at home. Representation in the other language groups is so small numerically that the proportions are approximately zero, and this value is assigned. Those who reported only English are also assigned a zero value.

Marital Status (MARRIED). This is a binary variable that distinguishes individuals who are married, spouse present (equal to 1) from all other marital states.

Location. The two location variables record residence of a rural area (Rural) or of the Southern States (South). The states included in the latter are: Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia.

Race. This is a dichotomous variable, set to one if the individual is Black, and set to zero for all other racial groups (White, Asian and Pacific Islander groups, other groups, American Indian).

Unemployment rate in the year of labor market entry (U). These data are the unemployment rate of males aged 20 years and over. Year of labor force entry is obtained from the data on year of entry for the foreign born for whom duration in the U.S. is less than total labor market experience (age minus schooling minus five), and it is year of entry into the labor market (year in which age is years of schooling plus five) for other immigrants. The unemployment rate data are from the Council of Economic Advisers, *Economic Report of the President, 1996*, Washington: U.S. Government Printing Office, Table B-38, p. 324.

Appendix 6B: English language fluency equation

Table 6B.1 Logit estimates of English-language fluency, foreign-born males, United States, 1990([a])

Variable	*Estimate*	*t-ratio*
Constant	0.753	5.07
Age	–0.054	67.32
Educational attainment	0.176	113.20
Years since migration (YSM)	0.160	44.95

(Continued Overleaf)

Table 6B.1 Continued

Variable	*Estimate*	*t-ratio*
YSM squared/100	–0.137	23.81
Married	0.194	13.93
Veteran	0.478	10.71
Black	0.617	9.18
Citizenship	0.545	34.90
Rural	0.092	3.20
South	0.135	8.71
Log weeks worked	0.175	14.08
Minority language concentration	–0.047	39.51
Linguistic distance	–0.673	4.50
Miles of origin country from US (10 thousands)	–0.051	0.15
Miles squared (10 millions)	0.325	6.68
Birthplace		
S. Europe	–2.224	23.60
E. Europe	–2.237	22.75
USSR	–2.502	23.02
Indochina	–4.215	25.02
Philippines	–1.658	11.73
China	–3.471	27.65
S. Asia	–2.092	13.65
Other Asia	–3.247	16.90
Korea	–3.365	24.65
Japan	–2.500	18.17
Middle East	–1.445	13.11
Sub Saharan Africa	–1.494	10.14
Mexico	–2.014	20.54
Cuba	–2.305	20.40
C & S American (Spanish)	–1.875	19.24
C & S American (non-Spanish)	1.555	6.19
Un rate yr. of labor market entry	0.083	10.30
U × YSM	0.003	5.12
Sample size	212,385	
McFadden's R^2		0.3426

Note: (a) The data are for men who worked and had earnings in 1989, were 25 to 64 years in 1990 and were born in non-English speaking countries. The dependent variable is LANG.

Notes

1 The research for this chapter was funded in part by the Bureau of International Labor Affairs (ILAB), U.S. Department of Labor. We appreciate the comments from Shirley Smith, the ILAB Project Officer. This chapter has benefited from

comments received at the European Science Foundation Conference on Migration and Development, Espinoh, Portugal, April 1998, the Population Association of America Annual Meeting, New York, March 1999, the Canadian Economics Association Annual Meeting, Toronto, May 1999, the Center for Economic Policy Research Conference on Marginal Labour Markets in Metropolitan Areas, Dublin, October 1999, and the Midwest Economics Association Annual Meeting, Chicago, April 2000. Christian Dustmann's comments were especially helpful.

2 One of the earliest studies to consider this issue is the McManus et al. (1981) paper on Hispanics in the United States.

3 This is greater than the 5.3% higher earnings for the fluent in the 1981 Australian Census (Chiswick and Miller 1995).

4 For analyses of the internal migration of immigrants see, for example, Bartel and Koch (1991).

5 Chiswick and Miller (1998) find that, other variables the same, English language fluency among immigrants in the United States is lower the higher is the emigration or return migration rate of immigrants in the United States from their country of birth.

6 Nakamura and Nakamura (1992) report that a higher national unemployment rate in the year of entry into the labor market is associated with a lower current wage among the foreign born in both the U.S. (1980 Census) and Canadian (1981 Census) labor markets. Stewart and Hyclak (1984) find that a higher annual growth rate in real GNP in the period of entry is associated with higher earnings among the foreign born in the 1970 U.S. Census. These studies did not test for whether the "scarring effect" varies by duration of residence in the destination. Chiswick et al. (1996), however, find that, using the Current Population Survey, the labor market conditions at the time of entry have an adverse effect on employment shortly after arrival, but do not have an effect on the future employability (employment ratio or unemployment rate) of immigrants in the United States.

7 Immigrants from English-speaking countries are excluded as the language issues do not exist for this group, while they may exist for immigrants from non-English speaking countries.

8 To the extent that there is measurement error in the language variable, there is misclassification in the fluent/non-fluent dichotomy.

9 The analyses in Table 6.1 were also computed with state fixed effects, that is, state level dummy variables, with the South variable deleted from the equation. There are no substantive differences in findings among any of the variables. For the specification (Table 6.1, column iii) and variables of primary interest (*t*-ratios in parentheses):

Variable	*Table 6.1 column iii*	*Same Specification State-Fixed Effects*
EDUC	0.045	0.045
	(83.1)	(82.4)
LANG	0.144	0.148
	(30.7)	(31.6)
CON	−0.0048	−0.0054
	(−14.3)	(−12.4)

The coefficient on the minority language concentration index is the variable most affected since its construction is based on state level data. Even so, the coefficient and *t*-ratio hardly change.

10 Estimates of the logit equation used in the computation of the sample selectivity correction terms are presented in Appendix 6B. See Chiswick and Miller (1998) for the model of the determinants of English language proficiency that serves as the basis for the selection equation. There are three identifying variables. The veteran status variable is used because experience in the United States Armed Forces would enhance the English-language proficiency of veterans. The U.S. Armed Forces accepts immigrants (resident aliens and naturalized citizens) as volunteers and drafted immigrants during periods of conscription. "Linguistic distance" between the origin language and English reflects the difficulty immigrants would have learning English. Geographic distance is a proxy for the probability of the immigrant returning to the origin country (emigration rate) and hence their incentive to learn English. A logit selection equation is used for consistency with earlier studies of the determinants of English language fluency. As logit and probit models of language fluency with these data yield similar results, the specification of the selection equation employed will have results similar to the more familiar specification based on a probit selection specification (Heckman 1979). Estimation of both types of models confirms this.

11 Similar findings regarding the complementarity of language skills with education and labor market experience are found for Canada (Chiswick and Miller 2000 [2003]).

12 Using the equations with the selectivity corrections (Table 6.2, columns ii and iv), the partial effect of pre-immigration experience at 10 years is 1.7% for those who are fluent in English and 1.1% for those who are not fluent.

13 Evaluated at 10 years duration it is 1.3% per year for the fluent and also 1.1% per year for those who are not fluent when the selectivity correction is applied (Table 6.2, columns ii and iv).

14 See Korenman and Neumark (1991) for a discussion of the effect of marriage on earnings among men.

15 Chiswick (1997) shows, using data from the 1980 and 1990 U.S. Censuses, that post-World War II immigrants from the Soviet Union, predominately refugees, have very low earnings initially but their earnings rise rapidly with duration of residence.

16 This is obtained (Table 6.1, column v) from 0.018(8 – 3) = 0.09 or 9%.

17 The partial effect of the unemployment rate on the natural logarithm of earnings (Table 6.1, column v), is:

$$\partial Ln\,Y/\partial \mathrm{U} = -0.018 + 0.00063\,(YRS),$$

which diminishes linearly in absolute value as years in the U.S. increases, and reaches zero at about 29 years.

18 More complex specifications are precluded by the limited number of period of immigration intervals and the small variability in the adult male unemployment rate in the post-World War II period.

19 The partial effects and *t*-ratios in this specification which is otherwise the same as Table 6.1, column (v) are:

Variable	*Coefficient*	*t-ratio*
UR	–.0.0288	–10.42
UR squared*	0.000159	6.88
Years in U.S. Labor Market		
Adjusted R square = 0.4192		

20 The partial effect is: $\partial LnY/\partial U = -0.0288 + 2(0.000159)\ (U)\ (\text{YRS})$. $\partial LnY/\partial U$ at the mean unemployment rate of 5.1 equals 1.26 at ten years in the U.S. labor market and equals zero at 17.8 years. The mean years in the U.S. labor market is 13 years.

21 The selection variables in the fluent and not fluent samples are defined as f_i/F_i and $-f_i/(1 - F_i)$ where f_i and F_i are the standard normal density and standard normal distribution functions, respectively, evaluated at the level of the index of fluency for individual i computed using Lee's (1983) methodology.

22 An alternative methodology is to estimate the specification presented in Table 6.1 using an Instrumental Variables (IV) estimator. Results using this procedure on 1980 Census data for the U.S. and Canada are reported in Chiswick and Miller (1992) and for several countries in Chiswick and Miller (1995). Note that identification in the selectivity-corrected estimating equation is obtained through both the identifying instruments and the non-linearity of the selection-correction term (lambda). This may account for the relatively greater reliability (robustness) of the selectivity corrected compared to the IV estimates.

23 The variable is associated with higher earnings (at the margin of statistical significance) among the small sample of immigrants from the heterogenous Other South-East Asia ('t' = 1.99). A positive effect on earnings could arise where immigrants can work in ethnic labor markets that afford protection against discrimination experienced in the general labor market. It is not clear, however, why the estimated impact for immigrants from Other South-East Asia is so different from that estimated for other birthplace regions.

References

Ahmed B, Gregory Robinson J (1994) *Estimates of Emigration of the Foreign-Born Population: 1980–1990*. Technical Working Paper No. 9, Washington, D.C.: U.S. Bureau of the Census

Bartel A, Koch M (1991) Internal Migration of U.S. Immigrants. In: Abowd JM, Freeman RB (eds) *Immigration, Trade and the Labor Market*. University of Chicago Press, Chicago 121–134

Ben-Porath Y (1967) The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy* 75(4):352–365

Chiswick BR (1978) The Effect of Americanization on the Earnings of Foreign-born Men. *Journal of Political Economy* 86(5):897–921

Chiswick BR (1979) The Economic Progress of Immigrants: Some Apparently Universal Patterns. In: Fellner W (ed) *Contemporary Economic Problems*. American Enterprise Institute, Washington 357–399

Chiswick BR (1997) Soviet Jews in the United States: Language and Labour Market Adjustments Revisited. In: Lewin-Epstein N, Ro'i Y, Ritterband P (eds) *Russian Jews on Three Continents: Migration and Resettlement*. Frank Cass, London 233–260

Chiswick BR (1998) Hebrew Language Usage: Determinants and Effects on Earnings Among Immigrants in Israel. *Journal of Population Economics* 11(2):253–271

Chiswick BR, Cohen Y, Zach T (1997) The Labor Market Status of Immigrants: Effects of the Unemployment Rate at Arrival and Duration of Residence. *Industrial and Labor Relations Review* 50(2):289–303

Chiswick BR, Miller PW (1985) Immigrant Generation and Income in Australia. *Economic Record* 61(173):540–553

Chiswick BR, Miller PW (1988) Earnings in Canada: The Roles of Immigrant Generation, French Ethnicity, and Language. *Research in Population Economics* 6:183–224

Chiswick BR, Miller PW (1992) Language in the Immigrant Labor Market. In: Chiswick BR (ed) *Immigration, Language, and Ethnicity: Canada and the United States*. American Enterprise Institute, Washington D.C. 229–296

Chiswick BR, Miller PW (1995) The Endogeneity between Language and Earnings: International Analyses. *Journal of Labor Economics* 13(2):246–288

Chiswick BR, Miller PW (1998) English Language Fluency Among Immigrants in the United States. *Research in Labor Economics* 17:151–200

Chiswick BR, Miller PW (2000 [2003]) The Complementarity of Language and Other Human Capital: Immigrants Earnings in Canada. RIIM Working Paper No. 00–08, Vancouver, July (Republished under the same title in *Economics of Education Review* 22(5):469–480)

Chiswick BR, Repetto G (2001) Immigrant Adjustment in Israel: The Determinants of Literacy and Fluency in Hebrew and Their Effects on Earnings. In: Djajic S (ed) *International Migration: Trends, Policies and Economic Impact*. Routledge, London 204–228

Heckman JJ (1979) Sample Selection Bias as a Specification Error. *Econometrica* 47:153–162

Korenman S, Neumark D (1991) Does Marriage Really Make Men More Productive. *Journal of Human Resources* 26(2): (Spring) pp. 282–308

Lee L-F (1983) Generalized Econometric Models with Selectivity. *Econometrica* 51:507–512

Mincer J (1974) *Schooling, Experience and Earnings*. National Bureau of Economic Research, New York

Nakamura A, Nakamura M (1992) Wage Rates of Immigrant and Native Men in Canada and the United States. In: Chiswick BR (ed) *Immigration, Language and Ethnicity: Canada and the United States*. American Enterprise Institute, Washington, D.C. 145–166

Stewart JB, Hyclak T (1984) An Analysis of Earnings Profiles of Immigrants. *Review of Economics and Statistics* 66(2):292–303

Tainer E (1988) English Language Proficiency and Earnings Among Foreign-Born Men. *Journal of Human Resources* 23(1):108–122

White H (1980) A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48:817–838

7 Schooling, literacy, numeracy and labour market success[1]

With Yew Liang Lee

I Introduction

When attempting to explain labour market outcomes in Australia, economists have generally based their analyses on the human capital model. According to Schultz (1962) and Becker (1964), human capital includes schooling, on-the-job training, medical care, migration and the acquisition of information. The search for better explanations of labour market outcomes has seen this list expanded to include language capital (Chiswick & Miller 1995) and other types of knowledge, such as computer skills (Miller & Mulvey 1997). Recent research by the member countries of Organisation for Economic Co-operation and Development (OECD) (1997) has developed this theme further by focusing on literacy and numeracy skills. It argues that the role of literacy and numeracy skills in the economy is not fully appreciated.

This chapter seeks to contribute to this literature by quantifying the links among schooling, literacy and numeracy skills and labour market outcome measures (labour force participation and unemployment) in Australia. This is done through the analysis of data from the *Aspects of Literacy* survey conducted in 1996 by the Australian Bureau of Statistics (Australian Bureau of Statistics 1997). The information in this data set relates to "functional literacy and numeracy skills". The use of test-based measures represents a significant advance over the self-assessments employed in earlier research (e.g. Chiswick & Miller 1995), while the use of information on quantitative literacy extends the range of skills incorporated into studies of labour market outcomes in Australia. The approach taken recognises the multiplicity of skills that constitute literacy in the modern economy.

Section II presents a brief review of some methodological issues. In Section III relevant features of the *Aspects of Literacy* survey are presented as well as some descriptive statistics for the measures of literacy and numeracy. Section IV presents and discusses the multivariate statistical analyses for ascertaining the contributions of educational attainment/qualifications and literacy and numeracy to the determinants of labour market success. The chapter closes (Section V) with a summary and conclusion.

II Methodological issues

According to the OECD (1997), literacy refers to the ability to understand and employ printed information to achieve one's goals and to develop one's knowledge and potential. A multiplicity of skills are held to constitute literacy in the modern economy. The OECD (1997) concentrates on prose, document and quantitative literacy. They show that there are strong connections between educational attainment and adult literacy: most people learn basic quantitative skills and to read at school, and this environment offers ongoing opportunities to use and improve literacy. Yet they also show that there are routes to literacy other than through formal education, that schooling need not always be successful in generating literacy, and consequently, the relationship between educational attainment and literacy is complex.

Similarly, the links between educational attainment, literacy and numeracy, and labour market outcomes, will not be straightforward. The human capital model suggests that the skills acquired at school may make more educated workers better at performing tasks ("worker efficiency") and may make them better decision-makers ("allocative efficiency") (Schultz 1975). In either case, the labour market outcomes for educated workers will be superior to those for their less-educated counterparts. However, the actual process through which education delivers these benefits is not clear. Included here is the extent to which the superior labour market achievements of the better educated can be linked to specific skills learned at school, such as literacy and numeracy.

There are a number of scenarios that can be considered. In the first, formal education is synonymous with literacy and numeracy: people attend school mainly to acquire literacy and numeracy, or literacy/numeracy may facilitate higher educational attainment. In this instance, information on either formal education or literacy/numeracy could be used to understand labour market outcomes; where the researcher has access to both sets of information, one will be redundant. The analyses by the OECD (1997) draw attention to the growing role of adult education and training in the development of literacy and numeracy, and hence suggest that this scenario is unlikely to be a complete description of the processes through which literacy and numeracy are developed. If this scenario were relevant, econometric analyses would result in unbiased but imprecise/unstable estimates when models include both measures of schooling and literacy/numeracy (i.e. multicollinearity), and the explanatory power of the model would be no greater when both schooling and literacy/numeracy are included in the estimating equation than when only either one is included.

A second scenario is where the functional literacy and numeracy that are important to labour market outcomes are developed mainly in the home and on-the-job. Schools may be important for teaching other skills, such as the affective skills of cooperation, perseverance and delayed gratification (Weiss 1995). In this case both literacy and numeracy, and schooling as a measure of these affective skills, should be included in explanations of labour market outcomes. If literacy and numeracy and education refer to essentially different skill sets that are orthogonal to each other, then the addition of either measure to an estimating equation that initially includes only one set of skills should not result in any major changes to the initial set of estimates. However, as with the first scenario, this description of the way that literacy and numeracy are developed is unlikely to be complete. As noted above, the OECD (1997) shows that most people learn basic quantitative skills and to read at school, and this environment offers ongoing opportunities to use and improve literacy.

A third, more general scenario, is where schooling enhances a range of skills, including literacy and numeracy. These skills may also be developed outside the schools sector, as has long been recognised in the education production function literature (e.g. Hanushek 1986). In this situation, literacy/numeracy measures and years of schooling should be included in estimating equations used to explain labour market outcomes. In this type of model, education can be viewed as having both direct and indirect (via literacy and numeracy) effects on labour market outcomes. The direct and indirect effects may vary across levels of education.

This discussion suggests that a number of issues can be examined in statistical analysis. First, what impacts do literacy and numeracy have on labour market outcomes? How large are these impacts compared to the effects of attributes such as educational attainment and experience that have been the focus in much of the literature on labour market outcomes in Australia? Second, are robust findings obtained from estimating equations that include both years of schooling and measures of literacy and numeracy? The estimates may not be robust where the measures are highly collinear, which might arise under the first scenario outlined above. Third, is extra explanatory power gained when measures of literacy/numeracy are added to models that include years of schooling, and is the estimated impact of years of schooling affected by this augmentation of the model? Similarly, what happens when the years-of-schooling variable is added to a model that contains literacy and numeracy measures? If the levels of literacy/numeracy and education are uncorrelated with each other, then even though each affects labour market outcomes, the estimated effects should be similar in the simple and augmented models described here. Fourth, do the ways that literacy and numeracy impact on labour market outcomes vary by level of education? This might be expected under the third scenario outlined above.

While these analyses will be informative of the processes outlined, it will not be possible to sort out the "causal" relationships between schooling and literacy/numeracy with these data. All three are measured at the time of the interview which is also when labour force status is measured. While schooling was completed some time earlier, with data only on skills measured after schooling is completed it is not possible to determine whether schooling caused or was caused by the literacy/numeracy skills. Moreover, the data are not sufficiently rich to develop an instrumental variables or simultaneous equations system to resolve the dilemma. In spite of these limitations, the results can be instructive.

III The Aspects of Literacy Survey

The *Aspects of Literacy* survey was a national survey designed to assess directly the literacy and numeracy skills of Australia's adult population (Australian Bureau of Statistics 1997). The survey was conducted between May and July 1996, and consisted of personal interviews administered to a representative sample of 10 700 people (aged 15–74) across Australia.

Data were collected on the demographic characteristics, labour force status, educational attainment and languages spoken by the individual, among other socioeconomic variables. The demographic variables include sex, age, state or territory of usual residence, birthplace, and, for the foreign born, year of arrival in Australia, age on arrival in Australia, and whether English was the respondent's first language. There are three broad categories of labour force status: "employed", "unemployed", and "not in the labour force".[2]

All respondents were asked to provide perceptions of their English reading and writing skills, along with similar perceptions on their mathematical skills. In addition, three types of literacy are assessed in an objective manner in the *Aspects of Literacy* survey, namely prose literacy, document literacy and quantitative literacy. *Prose literacy* is defined as the knowledge and skills required to understand and use information from texts, including texts from newspapers, magazines and brochures. *Document literacy* refers to the knowledge and skills needed to process information in materials such as tables, schedules, charts, graphs and maps. Document literacy depends on the individual's abilities in locating, integrating, generating and transferring information. *Quantitative literacy* (numeracy) involves the ability to perform arithmetic operations using numbers embedded in printed texts or documents. A quantitatively literate (or numerate) person must be able to locate and extract numbers from different types of documents that contain similar but irrelevant information, and also be able to perform the appropriate arithmetical operations when the operations to be used must be inferred from printed directions. The scores for prose literacy, document literacy and numeracy (quantitative literacy) available in the survey have been classified by the Australian Bureau of Statistics into five levels, level 1 (lowest) to level 5 (highest). Appendix 7A, available upon request, contains additional details on the literacy and numeracy data.

In the survey there tends to be a strong positive relationship between the level of educational attainment and the rate of participation in the labour force, and a strong inverse relationship between the level of educational attainment and the rate of unemployment among labour force participants.[3] Hence, the participation rate among males and females (combined) aged 15–74 years ranges from 57 per cent among individuals who did not complete secondary school, to 93 per cent for individuals with a higher degree

(PhD, Masters, etc.). The unemployment rate ranges from a low of only 2.2 per cent among those with a higher degree to 11.2 per cent for individuals who did not complete secondary school.

Self-reported English reading skills are classified into four levels, "excellent", "good", "moderate" and "poor". There is a very strong positive relationship between labour force participation rates and the level of reading skills, with the participation rate among males and females combined with "excellent" skills, at 78 per cent, almost 2.5 times that of those with "poor" reading skills (33 per cent). The unemployment rate of these groups differs by a factor of almost four: 5.5 per cent for individuals with "excellent" reading skills, and 20 per cent for those with "poor" reading skills.[4]

Prose skills are reported in five levels, with level 1 being the lowest and level 5 the highest. Labour force participation rates rise with prose skill level, with the rate being 48 per cent at the lowest level of skill, and 92 per cent at the highest level of skill. Conversely, unemployment rates fall with level of prose skill, from 16 per cent at the lowest level of skill and only 2 per cent at the highest level of skill.

Quantitative literacy skills are recorded in five levels, similar to prose skills. Labour force participation rates are positively related to quantitative skills, rising from 46 per cent at level 1 to 94 per cent at level 5. In comparison, unemployment rates fall with quantitative skill level, from 18 per cent at the lowest skill level, level 1, to only 1.7 per cent at the highest level of skill, level 5.

It is readily apparent from these data that level of educational attainment, literacy and numeracy are closely linked to the various labour market outcomes considered. The difficulty for understanding these links is that level of educational attainment and skills (such as, literacy and numeracy) are, as discussed in Section II, related. There is a positive association between schooling level and both literacy and numeracy. Nevertheless, analysis of the determinants of labour market outcomes that takes account of literacy and numeracy skills, as well as educational attainment, as developed in Section II, may permit improved insights into the reasons why the better educated have economic outcomes superior to those of the less-well educated.

IV Econometric analyses

Studies of the determinants of labour market outcomes have generally employed probability models where the participation decision is examined separately from the employment/unemployment outcome, conditional on being in the labour force. This is generally done for expositional purposes. Few additional insights have been gained from the studies that have adopted a more general multinomial specification of the probability model where the allocation of workers across the employment, unemployment and not-in-the-labour force states is considered simultaneously (see, for example, Brooks & Volker 1985; Wooden 1991). Single equation logit models of participation and of unemployment among labour force participants will be estimated in the present study.

The decision to enter the labour force is a major one for many individuals. It will be affected by a large range of factors. Research into these has largely concentrated on the labour supply decisions of females because of the nearly universal participation of nonaged adult males who are not disabled or enrolled in school. Chiswick and Miller (1994) and Kenyon and Wooden (1996) provide overviews of the Australian literature. Consistent with findings for other countries, Kenyon and Wooden (1996) report that cross-sectional studies show that female labour force participation increases with wages and with educational attainment, and decreases with the number of children at home.

The study that is probably of most relevance to the current research is by Chiswick and Miller (1994). They model female labour supply using a standard reduced form specification in which the participation decision is related to the respondent's age, educational attainment, location of current residence, marital status, presence and age structure of children, husband's income, country of birth, and, among immigrants, duration of residence in Australia, citizenship and English language skills. They show that among their sample of 25–64-year-old females, labour force participation rates decline with age, particularly from age 40 onwards. The degree of participation in the labour force increases with educational attainment, tends to be lower in non-metropolitan areas than in metropolitan areas, and is lower if children less than 15 years of age are present in the household, with the children's effect being more pronounced if they are aged less than 2 years.

Chiswick and Miller (1994) also include a dichotomous variable for English language proficiency in some of their specifications. This is set to one for individuals who speak only English at home, or if a language other than or in addition to English is spoken in the home, they speak English "very well". The variable was set to zero where a language other than English is spoken in the home and the respondent speaks English "well", "not well" or "not at all". It was shown that immigrants possessing English language fluency had participation rates about 4 per cent higher than other groups. This effect was the equivalent of that of about 1.5 years of schooling.

This study also models labour supply within a reduced form context.[5] Thus the person's tendency to join the labour force is expressed as:

$$PR_i^* = X_i\beta + \varepsilon_i \tag{1}$$

where PR_i^* is a latent index that captures the propensity of individual i to join the labour force, X is a row vector of observed factors (e.g. educational attainment, potential labour market experience, birthplace, etc.) that are held to influence labour supply decisions, β is a column vector of coefficients to be estimated, and ε is a stochastic error term that captures the net influences on labour supply decisions of all unobserved factors and also the influence of purely chance events. The explanatory variables in this model will be restricted to those used in the typical study, as the primary aim is to ascertain the extent to which the effects of schooling on labour supply decisions can be linked to literacy and numeracy.

Two outcomes are derived from PR^* with reference to an arbitrary threshold of zero. Thus, the individual is held to be a labour force participant ($PR = 1$) where PR^* exceeds zero, and is outside the labour force ($PR = 0$) otherwise. With the logit model to be employed here, the natural logarithm of the odds ratio of the probability of labour force participation (PR) to the probability of non-participation in the labour force ($1-PR$) is expressed as $\log\left[\frac{PR_i}{1-PR_i}\right] = X_i\beta$ The parameter estimates in the logit model therefore record the impact on the logarithm of the odds ratio of a small change in the explanatory variables.

The model specifications adopted here are constrained by the way the data from the *Aspects of Literacy* survey are made available to the research community. While information is available on educational attainment, potential labour market experience, birthplace, duration of residence, sex, disability and location in Australia, there are no data on marital status or family details such as the presence of children and the income of the spouse. In addition, there is a lack of detail on labour market history that might facilitate an approach along the lines of that used by Le and Miller (1999). Participation rates and unemployment rates are therefore both related only to the contemporaneously measured personal characteristics of educational attainment, potential labour market experience, disability, location and, where appropriate, literacy and numeracy, although the birthplace and duration of residence variables used provide some control for the history effects that are a feature of inertia models (see Le & Miller 1999). Separate equations are estimated for males and for females to allow for the effects of the full set of explanatory variables to vary by gender. The analysis is limited to those aged 15–64 years. Means and standard deviations for all variables are presented in Appendix 7B. This Appendix, available upon request from the authors, also contains a brief description of the variables.

Results from conventional models (which do not contain information on literacy and numeracy) of labour force participation and of unemployment are presented in Table 7.1.[6] The estimates of the determinants of male participation rates are presented in the first column, and of male unemployment rates in the second column of the table. There is a positive relationship between labour force participation and educational attainment. Evaluated at the mean of the variables, an extra year of schooling increases the male participation rate by 0.53 per cent.[7] Potential labour market experience also has a strong effect on labour force participation rates, with participation rates rising with this measure of experience for the first 17 years of potential labour market activity. Beyond this threshold participation rates

Table 7.1 Logit model of labour force participation and unemployment, males and females aged 15–64 years

	Males		*Females*	
Variable	*Participation*	*Unemployment*	*Participation*	*Unemployment*
Constant	2.150	1.019	–1.206	0.362
	(5.26)	(2.17)	(4.62)	(0.60)
Years of education	0.049	–0.237	0.184	–0.229
	(1.89)	(7.34)	(10.81)	(5.63)
Experience	0.106	–0.065	0.026	–0.003
	(6.66)	(3.80)	(2.73)	(0.13)
Experience squared (÷ 100)	–0.308	0.068	–0.123	–0.108
	(9.50)	(1.73)	(6.03)	(1.84)
Birthplace (Australia)				
Overseas – English-speaking country	–1.777	1.469	–2.268	0.942
	(4.06)	(2.91)	(7.36)	(1.42)
Overseas non-English-speaking country	–3.059	2.262	–2.615	1.425
	(7.76)	(4.76)	(8.69)	(2.13)
Period of residence (POR)	0.305	–0.151	0.267	–0.001
	(3.86)	(1.77)	(5.26)	(0.01)
POR squared (÷ 100)	–0.803	0.357	–0.645	–0.184
	(3.10)	(1.30)	(4.06)	(0.59)
Disabled	–1.536	0.650	–0.498	0.712
	(12.12)	(4.31)	(6.19)	(4.13)
Location (S. Australia and Tasmania)				
Capital city in other States	–0.047	–0.259	0.219	–0.129
	(0.23)	(1.11)	(1.85)	(0.46)
Non-capital city in other States	0.023	0.169	0.104	0.374
	(0.11)	(0.71)	(0.84)	(1.32)
χ^2 (10)	536.09	130.98	612.73	96.12
Pseudo R^2	0.211	0.076	0.115	0.069
Prediction success (%)	88.65	91.97	72.53	93.30
Sample size	3 621	3 176	4 285	2 894

Note: '*t*' statistics in parentheses.

Source: Australian Bureau of Statistics (1997).

decline with potential labour market experience. These patterns of labour market attachment are typical in the literature (see, for example, Miller & Neo 2002 [2003]).

Both birthplace and duration of residence matter when labour market attachment is being considered. Labour force participation rates are lower among the overseas born than among the native born, with the differentials being much greater among immigrants from non-English-speaking countries. The participation rate differences between the foreign born from English-speaking countries and the native born are quickly reduced with duration in Australia, with the participation rates of male immigrants from English-speaking countries exceeding those of the native born after about 7 years of residence. The participation rates of male immigrants from non-English-speaking countries, however, never gain parity with those of the native born.

It is not surprising that participation rates are also relatively low for those who report a disability. Location in Australia, however, does not appear to have an impact on labour force participation rates once other differences in the sample are taken into account.

Estimates of the unemployment model for males are listed in the second column of Table 7.1. Evaluated at the mean of the variables, an extra year of education lowers the unemployment rate by 1.7 per cent. The estimated coefficients on the experience variables reveal a pronounced U-shaped relationship between potential labour market experience and unemployment status. However, as the turning point in this relationship occurs at 48 years, for almost all of the sample[8] unemployment rates decrease with potential labour market experience.[9] Unemployment rates are higher among the overseas born than among the native born, and particularly so among the overseas born from non-English-speaking countries. The unemployment rates of the foreign born decrease with duration of residence in Australia for the first 21 years of residence. As a result, the unemployment rates of the foreign born from English-speaking countries gain parity with those of the Australia born after 15 years of residence in Australia. However, despite the pronounced reduction in unemployment rates with duration of residence, the unemployment rates of the foreign born from non-English-speaking countries do not catch up with those of the native born.[10]

Results for the models of labour force participation and unemployment for females are presented in the final two columns of Table 7.1. Labour force participation rates increase strongly with level of education (by 4.0 per cent per year of education, evaluated at the means), and unemployment rates are much lower among better educated females than among their less-well educated counterparts (by 1.4 per cent per year of education). Of some note is the apparent greater strength of the effect of education on the participation rate among females compared to males.

What happens when account is taken of the differences in literacy and numeracy across the levels of educational attainment? To examine this issue, the specification contained in Table 7.1 can be augmented with variables for literacy and numeracy. There is a wealth of information in the survey on literacy and numeracy. Indeed, as will be shown, the large number of highly intercorrelated literacy and numeracy variables creates multicollinearity problems if all are included in the analysis. While multicollinearity does not bias the logit coefficients, it results in large standard errors. Multicollinearity is not a problem if the sole purpose of the analysis is prediction, but it can be a problem for hypothesis testing. These considerations shape the procedures that are followed.

The analysis was undertaken as follows. First, the focus was on only six of the measures available in the survey, the three objective measures and the three subjective measures noted in the earlier discussion. Second, dichotomous variables were created for each of the possible categories within each of these measures. The "excellent" category was selected as the benchmark group in the case of the self perceptions measures, and the highest level of skill (level 5) was selected as the benchmark group in the case of the test-based measures. Third, variables for each of the measures of literacy and numeracy were simultaneously entered into the equations determining labour force status. There is a total of 21 variables; four each for prose skill, document skill and quantitative skill, and three each for the self-reports on reading, writing and mathematical skills.

When all of these variables are entered into the equation, many of the variables for literacy and numeracy were statistically insignificant or incorrectly signed, especially in the models of unemployment (regression results available upon request). This outcome is not surprising because of the very high intercorrelation among these explanatory variables. Multicollinearity is created when they are all included.

To illustrate the relationships between the various measures of literacy and numeracy, a correlation matrix is presented in Table 7.2 for the female sample used in the analysis of participation rates.[11] These correlations are polychoric correlations that are appropriate when each of the variables under consideration is categorical.[12] From the data in Table 7.2 the correlation between the self perceptions of reading and writing skills is 0.882, while that between the self perceptions of reading and mathematical skills is 0.611. Two patterns are apparent here. First, the correlations among the objective measures are higher than those among the self assessments. Specifically, the correlations among the test-based measures range from 0.895 to 0.966 while those among the self-reported measures range from 0.611 to 0.882. Second, the correlations between the self assessments and the objective measures are

Table 7.2 Correlations between measures of literacy and numeracy, female participation sample

	Self-perceptions data			*Test-based data*		
	Reading	*Writing*	*Mathematical*	*Prose*	*Document*	*Quantitative*
Reading	1.000					
Writing	0.882	1.000				
Mathematical	0.611	0.633	1.000			
Prose	0.637	0.593	0.460	1.000		
Document	0.575	0.549	0.483	0.926	1.000	
Quantitative	0.544	0.527	0.545	0.895	0.966	1.000

Note: Polychoric correlations based on 4285 observations.

Source: Australian Bureau of Statistics (1997).

lower than either those computed between various self assessments, or those computed between various objective measures. The mean of the correlation coefficients between the self-reported and test-based measures is 0.546. The multiple correlation when all of the variables are considered is, of course, even higher than the "pair-wise" correlations reported in Table 7.2, and it is these multiple correlations that are responsible for the multicollinearity in the logit equations.

Given the high degree of correlation among the measures of literacy and numeracy, all the measures of literacy and numeracy cannot be entered simultaneously into a single equation for labour market outcomes. There are alternative techniques available for addressing this issue. For example, various measures of literacy (or numeracy) can be combined into a single index. There is no unique way to combine the various measures into a single index. To do this, however, would be to lose insights into which of the highly intercorrelated measures would give the highest explanatory power when taken separately. This would result in a loss of information that may be particularly relevant for the cost-effective design of future questionnaires. That is, information is lost on which of the members of the set of literacy and numeracy variables gives the best fit.

Therefore, for the purposes of the present study a "general-to-specific" modelling strategy was employed. In this alternative, variables that are insignificant or have "perverse" signs are eliminated from the estimating equation in a sequential manner until the model contains only statistically significant terms with economically meaningful signs and magnitudes. Application of this general-to-specific modelling approach needs to confront the issue that there is not necessarily a unique path from the general model to the specific model. It was found, however, that it was generally only possible to include one of the sets of self-perception measures (i.e. self-perception data on either reading skills, writing skills or mathematical skills) and one of the sets of test-based measures (i.e. prose skill, document skill or quantitative skill) if the aim is to have significant, and economically meaningful, estimates of the literacy and numeracy parameters. Moreover, the results also show that estimating equations that contain only one of the sets of self-perceptions measures *or* one of the sets of test-based measures will be misspecified. That is, the subjective and objective measures each contain useful information. Drawing upon this, nine models were estimated that included the various combinations of self-perception measures (of either reading, writing or mathematical skills) and test-based measures (i.e. prose, document or quantitative skill). This procedure was followed for the models of labour force participation and unemployment for both males and females.

The likelihood functions for these nine models, for each gender and for each dependent variable, were then compared. Table 7.3 gives the χ^2 statistics for the test of overall goodness of fit for the nine models of female participation rates to illustrate the approach taken. Note that in Table 7.3 the highest overall χ^2 figure is found when self-perceptions of mathematical skills and

Table 7.3 χ^2 Statistics of overall goodness-of-fit, female participation model

Self-perception of	*Prose skill*	*Document skill*	*Quantitative skill*
Reading	718.05	735.42	733.35
Writing	743.81	760.22	756.79
Mathematical	755.88	767.38	758.88

Source: Australian Bureau of Statistics (1997).

test-based document skills are included in the estimating equation. Similar tabulations were computed for the other three categories of gender/employment status. This comparison shows that for the models estimated for females the combination of information on self-perceptions of mathematical skills and test-based measures of document skills maximised the likelihood function. For males the combination of self-perceptions of mathematical skills and the test-based measures of quantitative skills maximised the likelihood function.[13] For uniformity, and because it makes little difference to the argument, models based on the self-perceptions of mathematical skills and the test-based measures of document skill will be discussed here. This model will be termed the "restricted model".

Estimates from the restricted models are presented in Table 7.4. There are two reasons for presenting these estimates. The first is to report the effects literacy and numeracy have on labour force participation and unemployment. The second is to provide an examination of the impact that inclusion of measures of literacy and numeracy have on the partial effects of educational attainment and labour market experience on labour market outcomes. It is useful to provide an outline of the way these partial effects are to be interpreted.

The conventional model of labour force status listed previously can be re-written as:

$$PR_i^* = a_0 + a_1 E_i + \ldots \qquad (2)$$

where E denotes the level of education. The coefficient a_1 in this model can be viewed as capturing the *total* effect of educational attainment on labour market outcomes. When the model is augmented with variables for literacy (L) and numeracy (N) we have

$$PR_i^* = \gamma_0 + \gamma_1 E_i + \gamma_2 L_i + \gamma_3 N_i + \ldots \qquad (3)$$

In equation 3, γ_1 provides a measure of the *direct* effect of education on labour market outcomes independent of its indirect effects on literacy and numeracy. The difference in coefficients (a_1–γ_1) provides an estimate of the *indirect* effect of education on labour market outcomes that occurs via the measures of literacy and numeracy. The effect of education can be viewed as

Table 7.4 Logit model of labour force participation and unemployment including literacy and numeracy variables, males and females aged 15–64 years

	Males		*Females*	
Variable	*Participation*	*Unemployment*	*Participation*	*Unemployment*
Constant	2.884	–1.016	1.353	–2.451
	(4.38)	(1.05)	(2.34)	(1.62)
Years of education	–0.015	–0.134	0.105	– 0.153
	(0.50)	(3.69)	(5.59)	(3.45)
Experience	0.111	–0.070	0.025	0.003
	(6.89)	(4.04)	(2.52)	(0.14)
Experience squared (÷ 100)	0.316	0.076	– 0.117	–0.128
	(9.63)	(1.91)	(5.58)	(2.12)
Birthplace (Australia)				
Overseas – English-speaking country	–1.788	1.210	–2.145	0.772
	(4.03)	(2.35)	(6.75)	(1.11)
Overseas – non-English speaking country	–2.920	1.660	–2.161	0.871
	(7.24)	(3.38)	(6.90)	(1.24)
Period of Residence (POR)	0.314	–0.124	0.257	0.016
	(3.94)	(1.42)	(4.90)	(0.15)
POR squared (÷ 100)	–0.839	0.302	–0.626	– 0.215
	(3.21)	(1.08)	(3.83)	(0.66)
Disabled	–1.489	0.582	–0.374	0.536
	(11.59)	(3.79)	(4.50)	(2.97)
Location (South Australia and Tasmania)				
Capital city in other States	–0.078	–0.304	0.192	–0.142
	(0.38)	(1.29)	(1.59)	(0.50)
Non-capital city in other States	0.022	0.135	0.113	0.382
	(0.10)	(0.56)	(0.89)	(1.33)
Self-perception of mathematical skills (excellent)				
Good	–0.514	–0.291	–0.532	0.428
	(3.38)	(1.69)	(5.86)	(2.28)
Moderate	–0.515	0.139	–0.783	0.023
	(2.62)	(0.66)	(6.98)	(0.09)
Poor	–1.230	0.591	–1.336	1.287
	(4.18)	(1.73)	(6.69)	(3.34)
Document skill level (five = maximum)				
Four	0.271	0.521	–0.974	1.575
	(0.61)	(0.66)	(1.94)	(1.16)
Three	0.584	0.611	–1.050	1.452
	(1.35)	(0.78)	(2.12)	(1.08)
Two	0.360	1.117	–1.248	1.682
	(0.81)	(1.42)	(2.51)	(1.25)
One (minimum)	0.102	1.745	–1.659	2.480
	(0.22)	(2.19)	(3.29)	(1.82)
χ^2 (17)	572.53	178.10	767.38	135.57
Pseudo R^2	0.225	0.103	0.144	0.097
Prediction success (%)	88.84	92.00	74.10	93.30
Tests of incremental explanatory power				
Self-perception of mathematical skill χ^2 (3)	20.94	10.23	74.83	14.60
Objective measure of document skill χ^2 (4)	8.69	32.09	34.39	18.74
Sample size	3621	3176	4285	2894

Note: '*t*' statistics in parentheses.

Source: Australian Bureau of Statistics (1997).

$$\frac{\partial PR^*}{\partial E} = \gamma_1 + \gamma_2 \frac{\partial L}{\partial E} + \gamma_3 \frac{\partial N}{\partial E}. \tag{4}$$

The expression $\left(\gamma_2 \frac{\partial L}{\partial E} + \gamma_3 \frac{\partial N}{\partial E}\right)$ records the impact of education on labour market success that occurs because those with higher levels of education have higher levels of literacy and numeracy, and literacy and numeracy are themselves associated with superior labour market outcomes, as shown by γ_2 and γ_3, respectively.

The results listed in Table 7.4 indicate that both self-perceptions of mathematical skills and the objective measure of document processing skills exercise important influences on the labour market outcomes of males and females. The χ^2 tests of whether each set of variables adds to the explanatory power of the model are all significant at the 5 per cent level.

The higher the level of literacy and numeracy, the higher the labour force participation rate and the lower the unemployment rate. Consider self-perceptions of mathematical skills. Each of the three levels, "good", "moderate" and "poor", are associated with negative and significant effects on participation rates compared to the benchmark group with self-reported "excellent" mathematical skills. The partial effects for males, who have a mean labour force participation rate of 87.7 per cent, are around 6 per cent lower participation rates for "good" compared to the benchmark group that reported their mathematical skills as "excellent", 6 per cent for "moderate" and 13 per cent for "poor". For females, where the mean labour force participation rate is 67.5 per cent, the respective partial effects are 12, 17 and 29 per cent. These effects are substantial, and it is useful to bear in mind the size of the groups to which they apply. Table 7.5 lists the distribution of the population and of the labour force for both males and females across the self-perceptions of mathematical skills and objective measure of document skills data. While the group with "poor" mathematical skills is relatively small, each of the other groups has a relatively large representation, implying that the significant and sizeable labour force participation effects associated with the self-perceptions of mathematical skills apply to a large segment of the population.

The links between self-perceptions of mathematical skills and unemployment are weaker than those recorded for labour force participation. Two coefficients are significant at the 5 per cent level, and both of these are in the model for female unemployment rates. These show that the 43 per cent of female labour force participants who self-report their mathematical skills as "good" have unemployment rates 3 per cent higher than those that report their mathematical skills as "excellent".[14] The small group of females who self-report their mathematical skills as "poor" have unemployment rates approximately 11 per cent higher than the benchmark group with self-reported mathematical skills of "excellent".

Table 7.5 Distribution of population and labour force across self-perceptions of mathematical skills and objective measure of document skills, males and females aged 15–64 years

	Males		*Females*	
Variable	*Population*	*Labour force*	*Population*	*Labour force*
Self-perception of mathematical skills				
Excellent	0.412	0.433	0.340	0.408
Good	0.418	0.413	0.438	0.427
Moderate	0.142	0.133	0.185	0.146
Poor	0.028	0.021	0.037	0.019
Document skill level (five = maximum)				
Five (maximum)	0.025	0.026	0.017	0.024
Four	0.172	0.181	0.142	0.173
Three	0.376	0.394	0.372	0.418
Two	0.264	0.261	0.294	0.277
One (minimum)	0.163	0.138	0.175	0.108
Sample Size	3 621	3.176	4.285	2.894

Source: Australian Bureau of Statistics (1997).

The relationships between the objective measures of document skills and labour market outcomes are weaker than those discussed above for the self-perceptions of mathematical skills. Only those with the lowest level of document skills (level 1) have unemployment rates that can be distinguished in the analysis. The relevant partial effects are, however, very large: 13 per cent higher unemployment rates for males and 16 per cent higher unemployment rates for females. While the document skills variables as a group are significant in the model of male labour force participation, none of the individual document skill variables is significant. In comparison, all of the document skill variables are significant at the 6 per cent level in the model of female participation rates. The estimated participation rate effects indicate a strong monotonic relationship between the probability of participating in the labour force and the objectively measured levels of document skills.

The effect of literacy and numeracy skills on labour market outcomes noted above is quite large relative to the effects of education. Consider a person with poor document skills (level 1) and poor mathematical skills ("poor") compared to a person in the top categories for literacy and numeracy, other measured variables being the same. The impact on male unemployment is the equivalent of 17 years of education. This is a very large effect considering that the mean level of education is 13 years, and the range in the data is 11 years of schooling.[15] It is noted that the unemployment rates associated with the extreme levels of literacy and numeracy in the cross-tabulations discussed in Section III differed by a factor of between four and nine, while the unemployment rates associated with the education categories of "did not complete high school" and "Higher Degree" differed by a factor of around five. Hence the effects of literacy/numeracy derived from the multivariate approach are in fact often smaller than the unstandardised effects found in the cross-tabulations. It might be argued that the pronounced effects associated with literacy/numeracy are due to the small numbers in the minimum document skills level 1 (see Table 7.5). However, the lowest school leaving age of "13 years and under" also has small representation (2.8 per cent of the sample) which should similarly impact on the estimated effects shown for educational attainment that are used as the yardstick to assess the magnitudes of the labour market effects of literacy/numeracy. Finally, in a study of earnings for a number of countries, the OECD (1997) reports that for the United Kingdom and the USA, countries that Australia might be argued to resemble in terms of position in the knowledge economy, "the net return to (literacy) skill is large and is the same order of magnitude as the return to formal educational qualifications".

Finally, a comparison of Tables 7.1 and 7.4 shows that the inclusion of the literacy and numeracy variables lowers the estimated effects of the education variables. For males, holding constant the influences of literacy and numeracy skills, education is no longer statistically significant as a determinant of labour force participation rates, while the impact of education on unemployment, once account is taken of literacy and numeracy skills, remains significant but is approximately 45 per cent less than the impact attributed to education when literacy and numeracy skills are not taken into account. For females, the effect of education when account is taken of literacy and numeracy skills is approximately 40 per cent less than that derived without standardisation for literacy and numeracy skills. In other words, approximately one-half of the *total* effect of educational attainment on labour market outcomes appears to be due to the *indirect* effect of education that occurs via literacy and numeracy skills, and about half persists even when these crude measures of skills are held constant.[16]

The other measure of human capital included in the analysis is potential labour market experience, or years since leaving school, measured as age minus schooling minus 5 (Mincer 1974). The coefficients of the labour market experience and labour market experience squared variables reflect the combined effects of the share of potential earnings invested in on-the-job training and the rate of return from this training. Experience is apparently uncorrelated with the levels of literacy and numeracy. Additional years of potential labour market experience do not enhance literacy or numeracy scores. The coefficients on labour market experience and its square do not change when the literacy and numeracy variables are added to the analyses (compare Tables 7.1 and 7.4).

While the various measures of literacy and numeracy are highly correlated, they are not perfectly correlated. Each may therefore be thought of as containing elements of specific information. It would therefore be expected that, when all of the literacy/numeracy variables are included in the estimating equation, the coefficients on the education variables that record only the *direct* effects of education would be reduced even further compared to Table 7.4. Table 7.6 contains estimates that illustrate the main findings for the male unemployment rate.

The first column of Table 7.6 lists the estimate of the *total* effect of education from the model that does not incorporate any information on literacy or numeracy (see Table 7.1). The second column of the Table 7.1 lists the estimate of the *direct* effect of education from the restricted model that includes only variables for self-perception of mathematical skills and document skill (see Table 7.4). The third column of data in Table 7.6 contains an estimate of the *direct* effect of the education variable from a specification that contains *all* 21 of the variables for literacy and numeracy. Reading across Table 7.6, it is seen that while there is a pronounced change in the effect of education on male unemployment when the first, more limited, set of literacy and numeracy variables is included in the estimating equation (a change from –0.24 to –0.13), there is no further change when all the information on literacy and numeracy considered in this study is utilised in the estimated model (unchanged at –0.13). This type of effect is typical of the various models estimated for the other three gender/labour market status groups because of the very high intercorrelation among the literacy and numeracy variables.

A conclusion that can be drawn from this analysis is that the education effect in the typical study of labour force participation and of unemployment is an over-estimate of the *direct* contribution of schooling per se after its effects on literacy and numeracy are held constant. Approximately one-half of the effect commonly attributed to education is in fact due to the superior

Table 7.6 Selected coefficient from logit models of unemployment, males aged 19–64 years

	Simple model[a]	*Restricted model*[b]	*Full model*[c]
Years of education	–0.237	–0.134	–0.135
	(7.34)	(3.69)	(3.59)
χ^{2}[d]	130.98	178.10	193.83
Pseudo R^2	0.076	0.103	0.112
Prediction success (%)	91.97	92.00	92.03
Sample size	3 176	3 176	3 176

Notes: '*t*' statistics in parentheses.

a The Simple model does not include any literacy or numeracy variables. The coefficient is from Table 7.1.

b The Restricted model includes three variables for self-perceptions of mathematical skills and four variables for document skills. The coefficient is from Table 7.4.

c The Full model includes variables for self-perceptions of reading skill, writing and mathematical skills, and for prose, document and quantitative skills. Twenty-one variables are used for these influences.

d The degrees of freedom for the χ^2 tests are 10 for the Simple model, 17 for the Restricted model and 31 for the Full model.

Source: Australian Bureau of Statistics (1997).

literacy and numeracy skills of the better educated, which may, at least in part, be a consequence of their higher level of education. Moreover, only a limited set of variables on literacy and numeracy is required to standardise for the effects that these skills have on labour market outcomes.

It is also of interest to examine whether this phenomenon varies with the level of education. To examine this, the continuous years of education variable was replaced by eight dummy variables for the highest level of education completed. Selected findings are reported in Appendix 7C, available upon request. These estimates show that the findings discussed above carry across to each of the levels of education considered. For example, the coefficients on the bachelor degree variable for the four sets of gender/outcome equations are between 27 and 41 per cent smaller when literacy and numeracy are held constant than when the equations do not include information on literacy and numeracy. For the higher degree category, the impact when information on literacy and numeracy is included in the estimating equation is between 31 and 37 per cent smaller than that obtained from the specification that does not take account of literacy and numeracy. In other words, when separate educational categories are considered, between one-quarter and 40 per cent of the *total* effects associated with formal education appear to be due to the *indirect* effect of education that occurs via literacy and numeracy skills.

The analysis was also computed when the data were split into high-education (post-school qualifications) and low-education (all other) categories. Within each category there is a sharp reduction in the coefficient of education going from the simple model to the restricted model (when self-perceptions of mathematical skills and the document skills are added), but little further reduction going to the full model.[17] Yet the education coefficients remain statistically significant. The results from this disaggregated analysis indicated, however, that literacy and numeracy have stronger impacts on the labour market outcomes of the less-well educated sample than on the better educated. This might be expected, given the higher levels of participation and lower levels of unemployment for the better educated. For example, the unemployment rate among the better educated was approximately 5 per cent, while for the less well educated it was approximately 10 per cent. Another reason for the different impacts of literacy and numeracy might be that literacy and numeracy skills are higher on average among the better educated segment of the population (with post-school qualifications), and also more uniformly distributed. Employers may also exaggerate the extent of this uniformity, expecting those with post-school qualifications to be literate and numerate. In this case, as argued by the OECD (1997), "differences in skill are smaller and more difficult to detect, evaluate and reward". In other words, while, as noted below, the market rewards basic observable skills, there are premiums when "the comparatively large variation in literacy skills . . . facilitates differential recognition by employers" (OECD 1997).

As a further test, the analysis was computed separately for three major birthplace categories: Australia, English-speaking developed countries, and other countries. Again, it is found that within each of these birthplace categories there is a sharp decline in the coefficient on education when the self-perceptions of mathematical skills and the measures of document skills are added to the equation (restricted model), but little or no change when the other literacy and numeracy variables are added to the analysis (full model).[18] The education coefficient remains statistically significant, except among the relatively small sample from English-speaking developed countries.

Thus, it appears that the patterns observed for the pooled data regarding the effects on the coefficient of education from including literacy and numeracy in the participation and unemployment equations are also found when the analyses are computed separately within broad education and birthplace categories. This provides further support for the robustness of the findings.

Finally, in Section II the question was raised "What happens when a years-of-schooling variable is added to a model that contains literacy and numeracy measures?". Viewing equation 3 as an expression of the true model, the omission of the years-of-schooling variable would bias the estimated coefficients on the measures of literacy (L) and numeracy (N) by the amounts $\gamma_1 \frac{\partial E}{\partial L}$ and $\gamma_1 \frac{\partial E}{\partial N}$, respectively. These expressions are measures of the *indirect* effects of literacy/numeracy, and provide a means of inferring the *direct* effect of these skills.

Most interest in this analysis centres on a comparison of the magnitudes of these direct and indirect effects with those associated with years of schooling discussed previously. In order to focus the discussion, the analysis is restricted to the specification based on the self-perceptions of mathematical skills and the objective measure of document skills. Results from models of labour market outcomes that include these dimensions of literacy/numeracy and exclude the years-of-schooling variable are compared in Table 7.7 for female labour force participation with the results presented in Table 7.4.[19]

Examination of the estimated impacts of literacy and numeracy when the years-of-schooling variable is added to the model reveals an average change (decline in absolute value) of 17.6 per cent across the literacy/numeracy variables in the models of labour force participation and unemployment for males and females (compare, for example, columns (i) and (iii) of Table 7.7). As a number of the literacy/numeracy variables are statistically insignificant, it is of interest to focus only on significant terms: in this case there is a mean change (decline in absolute value) of 11.2 per cent in the estimated effects of literacy and numeracy following inclusion of the years-of-schooling variable.[20] Note, however, that other than for the male participation rate model discussed previously, the education variable is significant in the augmented model, indicating that it adds to the understanding of labour market outcomes.

Table 7.7 Selected coefficients from logit model of labour force participation, females aged 15–64 years

Variables	*Restricted Model_A*[a]	*Restricted Model_B*[b]	*Restricted Model_C*[c]
Years of education	[d]	0.184	0.105
		(10.81)	(5.59)
Self-perceptions of mathematical skills (excellent)			
Good	–0.563	[d]	–0.532
	(6.22)		(5.86)
Moderate	–0.836	[d]	–0.783
	(7.51)		(6.98)
Poor	–1.421	[d]	–1.336
	(7.17)		(6.69)
Document skill level (five = maximum)			
Four	–1.028	[d]	–0.974
	(2.06)		(1.94)
Three	–1.203	[d]	–1.050
	(2.45)		(2.12)
Two	–1.493	[d]	–1.248
	(3.04)		(2.51)
One (minimum)	–2.009	[d]	–1.659
	(4.03)		(3.29)
χ^2	735.33	612.73	767.38
Pseudo R^2	0.138	0.115	0.144
Prediction success (%)	73.65	72.53	74.10
Sample size	4 285	4 285	4 285

Notes: '*t*' statistics in parentheses.

a Restricted Model_A: Estimation omitting years of education variable but including literacy/numeracy variables.
b Restricted Model_B: From Table 7.1, column (iii).
c Restricted Model_C: From Table 7.4, column (iii).
d Variable not included.

Source: Australian Bureau of Statistics (1997).

From the perspective of the *direct* and *indirect* distinction drawn earlier, this finding implies that between 82 and 89 per cent of the effect that literacy/numeracy has on labour market outcomes can be categorised as a *direct* effect. Very little of this effect, only 11–18 per cent, arises because those with higher levels of literacy/numeracy skills have higher levels of schooling. Literacy/numeracy skills, which can be acquired through formal schooling and through informal means (at home, on-the-job), are rewarded in the labour market independently of whether they were initially associated with higher levels of education.[21]

Can the differing indirect effects of schooling and literacy/numeracy skills be rationalised? They suggest that additional years of schooling without improvements in literacy and numeracy will have a more modest impact on labour market outcomes. However, improvements in literacy and numeracy that are not accompanied by extra years of schooling can still have a more marked impact on labour market outcomes. This suggests that the labour market rewards basic observable skills, of which literacy and numeracy are but two dimensions. The mechanisms through which these skills are acquired may be of secondary importance.

V Conclusion

Many studies of labour market outcomes have focused on the effects of formal education and labour market experience. It is now taken more or less for granted that formal education and labour market experience are associated with more favourable labour market outcomes, whether measured by the degree of labour force participation, the extent of unemployment, or the wage and occupational status among the employed. The present study explores one reason why the better educated have such advantages.

It is shown in the current study using data from the Australian *Aspects of Literacy* survey that higher levels of education are associated with greater labour market success, as measured by participation and unemployment rates. It is also shown that higher levels of numeracy/literacy, whether measured from self-reported data or from objective test data, are also associated with greater labour market success. As the better educated also have higher levels of numeracy and literacy, it is possible that part of the improvement in labour market outcomes conventionally attributed to the attainment of more education could in fact be due to achievements in literacy and numeracy.

The estimation of models of labour market outcomes that include variables for both level of education and literacy and numeracy shows that approximately one-half of the *total* effect of education is in fact an *indirect* effect of education that arises due to the higher literacy and numeracy skills of the better educated. Education appears to be associated with improvements in skills (here literacy and numeracy) that are rewarded well in the labour market. Hence education affects labour market outcomes through its effects on human capital skills that are embodied in people and which are not measurable in most other studies. Only a limited set of variables on literacy and numeracy is required to standardise for the effects these skills have on labour market outcomes. Quite robust findings are generated with this limited set of variables, both for the full sample and for analyses computed separately within broad education and birthplace categories.

An alternative way of expressing the results is that while approximately half of the effect of education on labour market status is due to its association with literacy and numeracy, the other half is due to other factors. Some of this "other half" may be a consequence of greater measurement errors in literacy and numeracy than in years of schooling, the labour market value of other things learned in school (e.g. decision-making skills, sciences, affective or behavioural skills, etc.) and other omitted factors correlated with years of education (e.g. parental income and education, curiosity about the world, perseverance, etc.).

However, the effects of potential labour market experience on the labour market outcomes considered here do not vary with the inclusion of the literacy and numeracy variables. Experience has its effects on participation and unemployment independent of literacy and numeracy, perhaps because these skills are formed prior to, concurrent with, or after schooling, but not systematically with labour market experience.

When years of schooling is added to the models of labour market outcomes that include the measures of literacy/numeracy, there is only a modest decline in absolute value (11–18 per cent) in the partial effects of the literacy/numeracy variables. This implies that it is the literacy/numeracy skills that are of primary importance and not the setting (school or elsewhere) in which they are acquired.

This analysis also has implications for the design of surveys that include questions on literacy and numeracy. The first is that it is possible to ask objective (test-based) questions and subjective (self-reported proficiency) questions on literacy and numeracy. The second is that, due to the high intercorrelation among various dimensions of literacy and numeracy, most of the possible explanatory power can be obtained by including one measure for literacy and one for numeracy.

These findings demonstrate the importance of education. They show that part of the effect of schooling arises because of the greater literacy and numeracy skills, and presumably other unmeasured skills, of the more highly educated. They also reveal an asymmetry in that very little of the superior labour market outcomes of individuals with higher levels of literacy/numeracy arise because these individuals necessarily have higher levels of formal schooling. Literacy/numeracy skills, which can be acquired through formal schooling and through informal means (at home, on-the-job), are rewarded in the labour market independently of whether they are associated with higher levels of education.

Notes

1 This chapter is based on a study funded under Australia's Department of Education, Training and Youth Affairs' Evaluations and Investigations Program. Financial Assistance from the Australian Research Council is also acknowledged. Helpful comments have been received from two anonymous referees and Jeff Borland. The views are those of the authors and should not be attributed to the sponsoring agencies.

2 Although the survey also contains information on earnings, these data are available only in deciles. The cut-off points for the deciles are known, and hence the data can be thought of as having been made available in 10 categories. However,

attempts to use these data did not generate findings consistent with the Australian literature that has had access to superior measures of earnings. The results from study of the labour force status data are, however, consistent with findings for conventional models reported in the literature.

3 For further details regarding the descriptive statistics in this and the next three paragraphs, see Miller and Chiswick (1997) and Lee and Miller (2000). Other studies using these data include Norton (1997) and Cumming (1997).

4 Information on labour market outcomes by self-reported writing and mathematical skills and by document skills is available from the authors.

5 The study of unemployment is conducted here within a similar framework. Le and Miller (2000) provide a review of modelling issues and provide comparison results.

6 The information on location was made available in three categories: (a) capital city for the States of New South Wales, Victoria, Queensland, Western Australia, the Northern Territory and the Australian Capital Territory, (b) the balance of the State for these areas, and (c) the whole of South Australia and Tasmania.

7 The partial effects are calculated as $\beta(1 - \overline{PR})\overline{PR}$, where $\overline{PR}$ is the mean participation rate and β is the coefficient of interest.

8 Slightly less than one per cent of the sample has more than 48 years of potential labour market experience.

9 In Miller and Neo's (2002 [2003]) study based on the 1991 Australian Census of Population and Housing, there is also a U-shaped relationship between age and unemployment rates, although the lowest unemployment rates are experienced by those aged around 45.

10 Miller and Neo (2002 [2003]) report a similar result on the basis of their study of data from the 1991 Australian Census of Population and Housing. These findings are in contrast to the pattern in the USA where, other measured variables being the same, immigrants reach unemployment rate parity with the native born at approximately 5 years duration (see Chiswick & Hurst 2000). The difference may be due to the greater flexibility of wages within the USA.

11 Correlation matrices for each of the other three samples used in this study are available from the authors.

12 The usual correlations are Pearson product moment correlations, which are appropriate for variables measured on a continuous scale. The literacy and numeracy measures (denote two of these by L_1 and L_2), while being presented as categorical data, can be thought of as having underlying continuous indices (say L_1^* and L_2^*). With the assumption that L_1^* and L_2^* have a bivariate normal distribution, their correlation is referred to as a polychoric correlation coefficient (see Neale & Cardon 1992).

13 For males, the second highest value of the likelihood function is for the same combination of variables that maximises the likelihood function for the models estimated for females. For females, the models based on self-perceptions of mathematical skills and test-based measures of quantitative skills are associated with the third-highest value of the likelihood function.

14 The mean unemployment rate for females is 6.7 per cent; for males it is 8 per cent.

15 The data set includes a few individuals who reported that they did not attend school. These are not included in the computation of the range reported in the text. Including them yields a range of 19 years. It was not clear whether those with zero years of schooling are reporting errors or reflect true values.

16 The marginal (percentage point) effects of an extra year of schooling on labour market outcomes are as follows. Insignificant effects have been set to zero in this presentation.

Model	*Males*		*Females*	
	LFP	*Unempl.*	*LFP*	*Unempl.*
Simple Model—Total Effect	0.53	–1.74	4.04	–1.43
Restricted Model—Direct Effects	0.0	–0.99	2.30	–0.96

17 For example, for the coefficient on years of education for male unemployment:

Education	*Simple Model*	*Restricted Model*	*Full Model*	*Sample Size*
Low	–0.266	–0.156	–0.147	1 408
High	–0.159	–0.038	–0.017	1 768

Note: All coefficients are statistically significant.

Source: Appendix 7C.

18

Birthplace	*Simple Model*	*Restricted Model*	*Full Model*	*Sample Size*
Australia	–0.253	–0.133	–0.151	2 376
English-speaking	0.194	–0.103[a]	–0.027[a]	430
Non-English-speaking	–0.283	–0.217	–0.191	370

Note: a Coefficient not statistically significant.

Source: Appendix 7C.

19 Results for the other three labour market outcomes are not reported here, but are available from the authors.

20 This change is less than the change to the estimated impact of years of schooling when the literacy/numeracy variables are added to the model. (Compare, for example, columns (ii) and (iii) of Table 7.7) This arises from both the smaller partial effect of education γ_1 on labour market outcomes and the smaller partial derivatives $\frac{\partial E}{\partial L}$ and $\frac{\partial E}{\partial N}$ that determine the indirect effects of literacy/numeracy than the terms γ_2, γ_3 (partial effects of literacy and numeracy on labour market outcomes), $\frac{\partial L}{\partial E}$ and $\frac{\partial L}{\partial E}$ (partial derivatives) that enter into Equation 4 for the indirect effect of education.

21 This suggests that developing pathways other than formal schooling for the acquisition of these skills will have positive labour market effects.

References

Australian Bureau of Statistics (1997), *Aspects of Literacy: Assessed Skill Levels, Australia 1996*, ABS Catalogue 4228.0, Australian Government Publishing Service, Canberra.

Becker, G.S. (1964), *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. The University of Chicago Press, Chicago.

Brooks, C. and Volker, P.A. (1985), "Labour Market Success and Failure: An Analysis of the Factors Leading to the Workforce Destinations of the Australian Population", in Volker, P.A. (ed.), *The Structure and Duration of Unemployment in Australia*. Australian Government Publishing Service, Canberra, 43–71.

Chiswick, B.R. and Hurst, M.E. (2000), "The Employment, Unemployment and Unemployment Compensation Benefits of Immigrants", in Bassi, L.J. and Woodbury, S.A. (eds), *Long-Term Unemployment and Reemployment Policies, Research in Employment Policy*. Vol. 2. JAI Press, Stamford, CN.

Chiswick, B.R. and Miller, P.W. (1994), "Language and Labor Supply: The Role of Gender among Immigrants in Australia", in Slottje, D.J. (ed.), *Research on Economic Inequality*. Vol. 5. JAI Press, Greenwich, CT; 153–89.

Chiswick, B.R. and Miller, P.W. (1995), "The Endogeneity Between Language and Earnings: An International Analysis", *Journal of Labor Economics* **13**, 246–88.

Cumming, J. (1997), "The Quantitative Literacy Performance of Australians: Implications of Low Skill Levels", in Australian Bureau of Statistics (1997), *Aspects of Literacy: Assessed Skill Levels, Australia 1996*, ABS Catalogue 4228.0, Australian. Government Publishing Service, Canberra.

Hanushek, E.A. (1986), "The Economics of Schooling", *Journal of Economic Literature* **24**, 1141–77.

Kenyon, P. and Wooden, M. (1996), "Labour Supply", In: Norris, K. and Wooden M. (eds), *The Changing Australian Labour Market*. Australian Government Publishing Service, Canberra.

Le, A.T. and Miller, P.W. (1999), "A Risk Index Approach to Unemployment: An Application Using the Survey of Employment and Unemployment Patterns". Australian Bureau of Statistics Occasional Paper 6293.0.00.001. Ausinfo, Canberra.

Le, A.T. and Miller, P.W. (2000), "Australia's Unemployment Problem", *Economic Record* **76**, 74–104.

Lee, Y.L. and Miller, P.W. (2000), *Literacy, Numeracy and Labour Market Success*. Evaluations and Investigations Programme. Report 00/9. Department of Education, Training and Youth Affairs, Canberra.

Miller, P.W. and Chiswick, B.R. (1997), "Literacy, Numeracy and the Labour Market", in Australian Bureau of Statistics (1997), *Aspects of Literacy: Assessed Skill Levels, Australia 1996*, ABS Catalogue 4228.0, Australian. Government Publishing Service, Canberra.

Miller, P.W. and Mulvey, C. (1997), "Computer Skills and Wages", *Australian Economic Papers* **36**, 106–13.

Miller, P.W. and Neo, L. (2002 [2003]), *"Labor Market Flexibility and Immigrant Adjustment"*. Mimeo. Department of Economics, The University of Western Australia, Perth, (Republished under the same title in *Economic Record* **79**, 336–56).

Mincer, J. (1974), *Schooling, Experience and Earnings*. National Bureau of Economic Research, New York.

Neale, M.C. and Cardon, L.R. (1992), *Methodology for Genetic Studies of Twins and Families*. Kluwer Academic, Dordrecht.
Norton, M. (1997), "So—How Many People Can't Read?", in Australian Bureau of Statistics (1997), *Aspects of Literacy: Assessed Skill Levels, Australia 1996*, ABS Catalogue 4228.0, Australian Government Publishing Service, Canberra.
Organisation for Economic Co-operation and Development (OECD) (1997), *Literacy Skills for the Knowledge Society*, OECD, Paris.
Schultz, T.W. (1962), "Reflections on Investment in Man", *Journal of Political Economy* **60** (5 Pt 2: Investment in Human Beings), 1–8.
Schultz, T.W. (1975), "The Value of the Ability to Deal with Disequilibria", *Journal of Economic Literature* **13**, 872–6.
Weiss, A. (1995), "Human Capital vs Signalling Explanations of Wages", *Journal of Economic Perspectives* **9**, 133–54.
Wooden, M. (1991), "The Experience of Refugees in the Australian Labour Market", *International Migration Review* **25**, 514–35.

8 Language skills and earnings among legalized aliens[1]

1. Introduction

Language skills are a key aspect of the immigrant labor market adjustment process that has dominated analyses of immigrant well-being by economists. It has been shown using decennial census data that immigrants in the United States who are proficient in English earn about fifteen to twenty percent more than immigrants who have not mastered the English language (see Chiswick and Miller 1992, 1997).[2] The immigrants who are more likely to master the English language are readily identifiable. Among other characteristics they arrive in the United States when they are young, are well educated, have lived in the United States for a relatively long period of time, live away from concentrations of individuals with whom they share a non-English second language, have a lower probability of emigrating back to the origin, have an origin language linguistically closer to English, and come from a country that is further from the United States (see, for example, Chiswick and Miller 1992, 1998). Moreover, the major findings carry over to analyses conducted for immigrants from individual countries or regions, including immigrants from Europe, Mexico, other parts of Latin America, and Asia. Analyses for countries such as Australia, Canada, Israel and Germany show that immigrants in these countries have many experiences in common with immigrants in the United States (see, for example, Chiswick and Miller 1995; Chiswick 1998; Chiswick and Repetto 1998; Dustmann 1994; and Kee 1993). Thus, the model developed to explain dominant language proficiency appears to be robust across countries of origin, countries of destination, and time periods when applied to samples representative of all immigrants. It has also been shown that the model can be used to account for language skills among apprehended illegal aliens in a specific geographical region (see Chiswick 1991).[3]

This chapter has four main purposes. First, the applicability of the model of language fluency, developed and tested largely for the large, random samples of immigrant populations in census data, is examined for a random sample of illegal aliens who attained "temporary legal status" under the Immigration Reform and Control Act of 1986.[4] The data are from the *Legalized Population Survey* (*LPS*) conducted in 1989. These immigrants are less educated than the typical immigrant, have been in the United States for a shorter period of time, had been in an illegal status, and are more likely to be from Spanish-speaking countries than immigrants in general.

Second, the survey of legalized aliens that is analysed in the current study contains details not collected in alternative data sets, such as the 1990 Census, that may influence language outcomes. For example, information was collected on the number of family members living in the same household in the United States at the time the individual applied for amnesty, and on the number of times the individual had entered the United States prior to applying for temporary legal status. Examination of the links between these variables, among others, and language skills will further develop the model that has been successfully applied to a range of countries, time periods and situations (see Chiswick and Miller 1995).[5]

Third, information was collected on English-reading proficiency as well as English-speaking proficiency. This permits examination of the robustness of the model to a wider range of language skills than has generally been possible to date. Chiswick (1991) shows that similar processes enhance speaking and reading skills among a sample of apprehended illegal aliens in the Los Angeles region. The analyses reported below examine whether this finding characterizes a broader sample of illegal aliens.[6]

Fourth, the links between English language proficiency (reading and speaking skills) and earnings are examined in detail. The research question is whether speaking or reading skills alone will lead to higher earnings, or whether it is necessary to develop both skills in order to capture the income rewards that are associated with English language proficiency in the labor market.[7]

The structure of the chapter is as follows. Section 2 describes the data set and presents some basic descriptive material on the sample, including information on English-reading skills and English-speaking skills. Section 3 applies to these data the model of English language proficiency developed by Chiswick and Miller (1992, 1995, 1998). Both proficiency in speaking English and reading English are examined, and several extensions are considered to the model previously estimated using 1980 and 1990 U.S. Census data. In Section 4 the relationship between language skills (reading and speaking) and earnings among the legalized population is examined. Section 5 presents a summary and conclusion.

2. The Legalized Population Survey

The analyses reported in this paper are based on the *Legalized Population Survey* (*LPS*) Public Use Sample made available by the Immigration and Naturalization Service (INS). This data set contains information on 6,193 aliens who attained "temporary legal status" (amnesty) under Section 245(A) of the Immigration Reform and Control Act of 1986. To qualify under the provisions of this Act aliens must have lived in the United States "continuously" since before January 1, 1982, and they had to apply to the Immigration and Naturalization Service for the temporary legal status which would, after one year, result in permanent legal (resident alien) status. Approximately 1.8 million persons qualified for temporary legal status under these provisions.[8] Most applications were filed between May 1987 and May 1988. The sample drawn from this population was collected using a two-stage stratified cluster design.[9] Only individuals 18 years or older were interviewed. Interviews were conducted in the first half of 1989.

The sample contains data on demographic characteristics, language proficiency, immigration details (number of times entered the United States, year of first entry, reasons for staying in the United States, reasons for leaving the United States, country of citizenship), state of residence in the United States, employment prior to entering the United States, and in the United States, family composition, health status, use of social services, education, income, etc.[10]

The *Legalized Population Survey* (*LPS*) contains a range of information on language skills. Table 8.1 presents summary information on the language questions and responses.[11] Fully, 58% of respondents indicated that they could read and understand a newspaper written in English, and a slightly higher percentage, 65%, indicated that they could read and understand instructions or recipes written in English. The percentages indicating an ability to speak English in specific situations are slightly higher than those indicating English-reading ability. These data, therefore, indicate that about forty percent of the sample experience difficulties with the English language.

By comparison, data from the 1990 U.S. Census show that, among the foreign born with 6 to 10 years of residence in the U.S., 34% have poor English-speaking skills (*i.e.*, they self-report their English-speaking skills as "Not Well" or "Not at All") (see Chiswick and Miller 1996b). Hence, at this

Table 8.1 Summary information on responses to language questions in the Legalized Population Survey, 1989

Question[a]	*% responding "Yes"*[b]
Can you read and understand a newspaper written in English?	57.5
Can you read and understand a magazine written in English?	57.1
Can you read and understand instructions or recipes written in English?	64.5
Can you speak to a sales clerk in English?	65.5
Can you speak to a doctor, nurse or teacher in English?	63.0
Can you speak on the telephone in English?	60.5

Notes

a The form but not the substance of the survey question differs from that used in the table.

b Percentages derived from weighted data, computed using only valid responses to each question.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice, 1989 very broad level, it appears that respondents to the *LPS* have somewhat poorer English-speaking skills than the general immigrant population of short-to-medium duration in the United States.

However, due to differences in the nature of the questions on language skills in the two information sources, and the different populations (legalized aliens versus immigrants in general), further inquiry is needed to determine if the processes determining English-language skills among illegal aliens who applied for amnesty are similar to those of the immigrant population in general. To this end, the model of Chiswick and Miller (1998) is applied to the data. A number of methodological issues associated with the application of this model need to be considered.

Given that the focus of analysis is on the acquisition of language skills, Chiswick and Miller (1998) restrict their analysis to immigrants from non-English-speaking countries. The analyses of the *LPS* presented below will also be restricted to aliens from non-English-speaking countries for whom the modelling of language skills in the United States is more relevant. There are, however, several differences from the modelling strategy employed by Chiswick and Miller (1998).

First, two variables which are used in the study of the 1990 U.S. Census, namely resident of a rural area and veteran status, are not available in the *LPS*. As these have a minor impact in the models estimated by Chiswick and Miller (1998), and it is unlikely that any of the legalized aliens were veterans of the U.S. Armed Forces, their omission from the current study is not of major concern.

Second, Chiswick and Miller (1998) employ a set of seventeen birthplace dummy variables to capture differences in English proficiency across countries of birth. Many of these are highly significant. In the current analysis, however, for two inter-related reasons a restricted set of birthplace categories is used. One is that the sample of aliens is dominated by individuals from Mexico (71.5% of the sample for males, 69.4% of the sample for females) and the Spanish-speaking regions of Central and South America (16.9% of the sample for males, 20.6% of the sample for females), which is dominated by individuals from El Salvador. The second is that after removing individuals with missing values for relevant variables the sample size is reduced to 3,183 unweighted observations for males and 2,425 unweighted observations for females.[12] As a result, the sample size among those not of Mexican or Central and South American origin is quite small. Therefore, only three birthplace categories are identified in the analysis: Mexico, the Spanish-speaking regions of Central and South America, and a residual category of all other birthplace regions.

Third, because of the predominance in the sample of aliens from just two birthplace regions, both of which are Spanish-speaking, several variables that were constructed for use in the analysis of the general immigrant population in Chiswick and Miller (1998) have limited variability and hence have not been included in the models presented here. These are a measure of the linguistic distance between the individual's mother tongue and English, and a measure of the physical distance between the individual's country of origin and the U.S.[13]

Table 8.2 lists means and standard deviations for legalized aliens in the *LPS*. The left-hand columns refer to males and the right-hand columns refer to females. For comparison, means and standard deviations for immigrants from non-English speaking regions from the 1990 Census are also presented.

The typical alien in the *LPS* is around 34 years of age, has 8 years of education and has resided in the United States for 12 years.[14] By comparison, the typical immigrant in the 1990 Census was 40 years of age, had 11 years

Table 8.2 Means and standard deviations, selected variables, Legalized Population Survey and 1990 U.S. Census[a]

Variable	*Males*		***Females***	
	LPS[b]	*1990 Census*	*LPS*[b]	*1990 Census*
Age	33.49 (8.94)	39.79 (10.63)	34.21 (9.33)	41.10 (10.91)
Educational attainment	8.00 (4.34)	11.63 (4.99)	7.55 (4.13)	11.20 (4.68)
Years since migration	12.01 (5.09)	15.21 (11.08)	11.40 (3.99)	16.15 (11.53)
Minority language concentration	16.02 (8.16)	7.39 (8.92)	16.74 (7.75)	6.81 (8.70)
Resident of Southern State	0.208 (0.406)	0.237 (0.425)	0.198 (0.399)	0.246 (0.431)
Married	0.531 (0.499)	0.655 (0.475)	0.606 (0.489)	0.689 (0.463)
Born in Mexico	0.715 (0.451)	0.259 (0.425)	0.694 (0.461)	0.234 (0.423)
Born in Cent. & So. America	0.169 (0.375)	0.109 (0.312)	0.206 (0.405)	0.132 (0.338)
English-speaking proficiency	0.667 (0.471)	0.730 (0.444)	0.557 (0.497)	0.696 (0.460)
English-reading proficiency	0.601 (0.490)	[c]	0.517 (0.500)	[c]
Sample size[d]	3,183	237,770	2,425	243,496

Notes

a Sample restricted to aliens/immigrants from non-English-speaking countries; standard deviations in parentheses.

b *LPS = Legalized Population Survey*.

c Variable not available.

d Weighted data with sample size scaled to actual number of observations.

Sources: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice, 1989. 1990 U.S. Census of Population and Housing, Public Use Microdata Sample, 5% sample.

of education and had lived in the United States for 15 years. Hence, the respondents to the *LPS* are younger, less-well educated and have lived in the United States for a shorter period of time than the typical immigrant. They are also more likey to live in regions that have concentrations of individuals with whom they share a mother tongue, with the means of the minority language concentration variable (constructed from the 1990 Census), being more than twice as great in the *LPS* as in the 1990 Census data, around 16% in the *LPS* compared to 7% in the 1990 Census. This is due to the very large representation of Spanish speakers in the *LPS* and their disproportionate concentration in areas with other Spanish speakers.

Two-thirds of males in the *LPS* and 56% of females were proficient in speaking English, defined here as being able to speak to a doctor, nurse or teacher in English (see Appendix 8A). About 60% and 52% of males and females, respectively, were proficient in reading English, defined here as being able to read and understand a newspaper written in English (see Appendix 8A). It is important for a range of policy issues to understand why some of this group of aliens are able to speak and/or read English while others lack these skills. Analyses of the determinants of English-speaking and reading proficiency are reported in the following section.

3. The language model and estimates

The model

The model of language fluency that is estimated in this section is based on the model developed and tested by Chiswick and Miller (1992, 1995, 1998). This model is structured around three main sets of conceptual variables: economic incentives, exposure and efficiency. The empirical measures of these conceptual variables are discussed in detail in these studies, and only brief comments need be provided here.

The schema of the language model may be described as:

$$\text{LANG} = \text{f(economic incentives, exposure, efficiency)},$$

where LANG is a measure of the immigrant's proficiency in the dominant language. The empirical counterpart of this conceptual equation used in the study of the *LPS* is a modification on the model outlined in Chiswick and Miller (1998), namely:

$$\begin{aligned}\text{LANG} = \text{f}\,[&\text{duration in destination }(+)\text{, marital status }(?)\text{, age at migration}\\ &(-)\text{, education }(+)\text{, minority language concentration }(-)\text{, rela-}\\ &\text{tives in U.S. }(-)\text{, children }(?)\text{, location }(?)\text{, country of origin}(?)],\end{aligned}$$

where the expected partial effects (positive, negative or ambiguous) are indicated in parentheses.

The length of time that an immigrant has resided in the United States provides a measure of the immigrant's exposure to English. Many studies of immigrant adjustment have shown that as length of residence in the United States increases, immigrants adjust to the specific conditions of the United States economy and society. Part of the adjustment process appears to involve the learning of English. This is presumably due to the interactions that come about as part of every-day life in the country of destination, as well as specific investments that are made in language training, such as "English as a second language" programs. It is expected that a similar process is relevant for newly legalized aliens and, therefore, that language skills will improve with duration of residence in the United States, even when they were in an illegal status (Chiswick 1991). To capture the non-linear effect, that is, that duration has its largest impact on language skills in the first few years and its marginal effect diminishes with duration, duration of residence is entered in the model in quadratic form.

Age and educational attainment are expected to impact on the individual's efficiency in learning English. It is well established in the linguistics literature that the very young have a superior ability to acquire language skills.[15] Such an ability may also reside in those who are better educated. The greater efficiency in language acquisition of the more educated may arise because they have a greater mastery of their mother tongue and are therefore more efficient in learning new concepts and new terminology. Furthermore, those with schooling in the destination would be expected to be more fluent in the destination language as fluency may be a prerequisite for school enrollment and the destination schooling itself would enhance fluency. Unfortunately, it is not possible to differentiate between years of schooling in the country of origin and the United States in these data. Finally, those with greater ability (or some other unmeasured characteristics) may make greater investments in all forms of human capital, including both formal schooling and language skills.

The minority language concentration variable is included in the estimating equation to capture an aspect of the intensity of exposure to English per unit of time in the destination, or rather, the ease of avoiding exposure to English. It has been measured in several previous studies by the extent to which the individual's mother tongue is spoken in the area in which the respondent lives, whether by immigrants or natives. It is hypothesized that the greater the extent of this phenomenon, the easier it is to avoid using English and hence the poorer the English language skills. The concentration variable used in this study is computed from the very large sample in the 1990 Census (see Appendix 8A).

Marital status is defined at the time of application for temporary legal status. Where the person was married prior to entry into the United States the person would typically share a mother tongue with the spouse.[16] In this situation, opportunities for conversation in that mother tongue within the home substitute for conversations in English, and thus both reduce the need to learn English and limit the learning by doing that may otherwise take place. When the individual married after arrival in the United States, however, it is less likely than pre-immigration marriage to be to a person with the same mother tongue. This would tend to enhance the individual's proficiency in English through greater use of English in the home. Unfortunately, it is not possible to identify in the *LPS* when the marriage of those currently married occured.

The intensity of exposure per unit of time in the United States will also depend on the presence of young children in the household. Several hypotheses have been presented for the effect of young children on parental language skills. One hypothesis is that parents may teach the origin language to their children. This may be motivated by desires for return migration, or by a drive to maintain the culture of the country of origin. This appears to be an important feature of the language attainment process in Australia (see Chiswick and Miller 1996a). It is a factor which might be particularly intense among aliens for whom there would always be a possibility of apprehension and deportation, or for whom there may be a higher rate of voluntary return migration. Both considerations may characterize illegal aliens in the United States.

The second way in which children may affect the language outcomes of their parents is through a child-to-parent transmission mechanism. Under this scenario, parents learn English from their children. It is well known in the second-language acquisition literature that children learn languages more rapidly than older persons. They do this in part because they are more efficient at learning languages (see Long 1990; Service and Craik 1993). They also have a more intensive exposure to English through the school system and through interaction with other children.

Third, children may serve as parental interpreters to the world outside the ethnic language enclave. To the extent that children perform this function, they will reduce the incentive and opportunity for parents to learn English. This effect is more likely to be relevant in the home sector and for consumption activities than in the labor market. This implies that children have a smaller positive effect or a larger negative effect on the English language proficiency of their mothers than of their fathers.

Finally, children have a depressing effect on the labor supply of their mothers, but not of their fathers. To the extent that investments in English language skills are made because of their value in the labor market and to the extent that labor force activity provides greater exposure to English than does engaging in home production, a lower labor force participation rate would lower English-language proficiency. Hence, through this mechanism, children would have a greater depressing effect on their mother's rather than their father's English language proficiency.

Consistent with this discussion, data on the presence and ages of children have been included in many empirical applications of models of language attainment, including this study.

The estimating equations

The multiple regression analyses of English language proficiency are presented in Tables 8.3 to 8.7. Tables 8.3 and 8.4 refer to English-speaking proficiency, Table 8.5 refers to English-reading proficiency, while Tables 8.6 and 8.7 contain extensions of both models of English-speaking and reading proficiency. The discussion will first focus on Tables 8.3 and 8.4. The results in Table 8.3 are for males while those in Table 8.4 are for females. In each case results are presented from similar models estimated using the *LPS* and the 1990 U.S. Census.

The most striking feature of Table 8.3 is the similarity of the results for males obtained from the *LPS* and the 1990 U.S. Census. In the analysis of each data set, English-speaking proficiency increases with educational attainment and decreases with age at migration. It also increases at a decreasing rate with the

Table 8.3 Regression estimates of model of English-speaking proficiency, males, Legalized Population Survey and 1990 U.S. Census[a]

Variable	*Legalized Population Survey* (i)	(ii)	*1990 U.S. Census* (iii)	(iv)
Constant	–0.266 (5.30)	0.442 (7.66)	0.257 (60.39)	0.457 (97.99)
Age	–0.008 (8.39)	–0.009 (9.53)	–0.006 (66.74)	–0.007 (84.46)
Educational attainment	0.042 (25.85)	0.037 (19.92)	0.039 (230.85)	0.031 (162.83)
Years since migration (YSM)	0.035 (7.07)	0.035 (6.90)	0.021 (98.26)	0.023 (110.47)
$YSM^2/100$	–0.068 (5.22)	–0.064 (4.86)	–0.022 (50.67)	–0.026 (60.92)
Married	0.075 (4.75)	0.086 (5.43)	0.033 (18.70)	0.029 (16.70)
South	–0.021 (1.07)	–0.020 (1.03)	–0.010 (5.55)	0.012 (6.57)
Minority language concentration	[b]	–0.006 (6.99)	[b]	–0.010 (89.73)
$\bar{R}^2$	0.1762	0.1864	0.2781	0.3070
Sample size[c]	3,183	3,183	237,770	237,770

Notes

a Sample restricted to aliens/immigrants from non-English-speaking countries; figures in parentheses are White's (1980) heteroskedasticity-consistent '*t*' statistics.

b Variable not entered.

c Weighted data with sample size scaled to actual number of observations.

Sources: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice, 1989. 1990 U.S. Census of Population and Housing, Public Use Microdata Sample, 5% sample.

number of years the immigrant has lived in the United States. Language skills are higher among married men than among their non-married counterparts, and are lower in regions with a concentration of individuals speaking the same non-English second language as the immigrant.

According to the estimates obtained from the *LPS*, each additional year of age at migration is associated with a reduction in the probability of English-speaking proficiency of almost one percentage point. Although comparisons across data sets may be problematic, it appears that this impact is stronger in the *LPS* than in the 1990 U.S. Census. English-speaking skills improve with educational attainment, at around 4 percentage points per year of schooling, an impact which is remarkably similar to that estimated from the 1990 U.S. Census. The improvement in English-speaking skills with duration of residence in the United States is estimated to be 2.1 percentage points per year when evaluated at 10 years of residence. By comparison, the improvement in English-language skills with duration of residence estimated from the 1990 U.S. Census is 1.7 percentage points when evaluated at 10 years of residence.

The positive effect on language skills for men associated with marriage is almost double in the *LPS* compared to the 1990 U.S. Census (7.5 percentage points compared to 3.3 percentage points). Finally, the effect on English-speaking proficiency of residence among a concentration of individuals who speak the same minority language as the individual is slightly weaker in the *LPS* than in the 1990 U.S. Census. The estimated impacts in the two data sets are –0.006 and –0.010, respectively.

It is also noted that the results in Table 8.3 are broadly similar to those reported in Chiswick (1991) for a sample of apprehended illegal aliens who were young and had a short average duration in the United States. The main exception is that age at migration is negative but statistically insignificant in Chiswick's (1991) analysis, whereas it is highly significant in the current study. It is probable that this difference is associated with the fact that the mean age of Chiswick's (1991) sample was quite young (23 years compared to 33 years in the *LPS*), and age had a smaller standard deviation (7.3 years compared to 8.9 years in the *LPS*).

The results for females in Table 8.4 are similar to those described for males in terms of both the direction of effects and the comparison of the results from the *LPS* and the 1990 Census. Two features of the findings should, however, be noted. First, the impact of duration of residence among women is stronger among the legalized aliens in the *LPS* than among the general sample of immigrants in the 1990 Census. The partial effect at 10 years of residence is 3.4% for the legalized aliens in the *LPS* but only 1.8% among immigrants in the 1990 U.S. Census. Second, the impact of being married is statistically insignificant in both specifications listed for the *LPS*. While the marital status effect is significant in the equations estimated from the Census, it varies in direction of impact across the specifications.

The model for English-reading proficiency in Table 8.5 is the same as that estimated for English-speaking proficiency. With one exception, the basic

Table 8.4 Regression estimates of model of English-speaking proficiency, females, Legalized Population Survey and 1990 U.S. Census[a]

Variable	*Legalized Population Survey*		*1990 U.S. Census*	
	(i)	*(ii)*	*(iii)*	*(iv)*
Constant	–0.193 (3.00)	–0.036 (0.51)	–0.137 (30.05)	–0.381 (76.89)
Age	–0.005 (5.15)	–0.006 (5.88)	–0.005 (56.44)	–0.007 (80.14)
Educational attainment	0.054 (28.40)	0.050 (23.36)	0.043 (241.20)	0.035 (176.86)
Years since migration (YSM)	0.063 (8.98)	0.063 (8.92)	0.022 (105.31)	0.024 (121.70)
$YSM^2/100$	–0.143 (6.81)	–0.145 (6.68)	–0.022 (51.82)	–0.027 (65.82)
Married	–0.009 (0.53)	–0.007 (0.37)	–0.009 (5.09)	–0.010 (6.02)
South	0.047 (2.11)	0.049 (2.21)	–0.003 (1.73)	0.023 (13.15)
Minority language concentration	[b]	–0.006 (5.72)	[b]	–0.012 (112.86)
$\bar{R}^2$	0.2487	0.2568	0.3037	0.3461
Sample size[c]	2,425	2,425	243,496	243,496

Notes

a Sample restricted to aliens/immigrants from non-English-speaking countries; figures in parentheses are White's (1980) heteroskedasticity-consistent '*t*' statistics.

b Variable not entered.

c Weighted data with sample size scaled to acutal number of observations.

Sources: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice, 1989. 1990 U.S. Census of Population and Housing, Public Use Microdata Sample, 5% sample.

pattern of results is the same for the two skills. Thus, the similarities are that English-reading proficiency decreases with age at migration, increases with educational attainment, and increases at a decreasing rate with years since migration.[17] With respect to age at migration, each additional year is associated with a 0.8 percentage point reduction in English-speaking proficiency and a 0.7 percentage point reduction in English-reading proficiency. The partial effect of educational attainment on English-reading proficiency is slightly greater than that on English-speaking proficiency, while the improvement of English-reading proficiency with years since migration occurs at a less rapid pace than does English-speaking proficiency. The latter finding is also a characteristic of Chiswick's (1991) analysis of speaking and reading skills.

The major difference between the findings for the determinants of English-

Table 8.5 Regression estimates of model of English-reading proficiency, male and female legalized aliens, Legalized Population Survey[a]

Variable	*Males*		*Females*	
	(i)	*(ii)*	*(iii)*	*(iv)*
Constant	0.122 (2.53)	0.174 (3.10)	–0.226 (3.52)	–0.133 (1.89)
Age	–0.007 (7.06)	–0.007 (7.32)	–0.006 (5.87)	–0.006 (6.28)
Educational attainment	0.055 (36.96)	0.054 (31.16)	0.057 (30.14)	0.054 (25.97)
Years since migration (SYM)	0.028 (5.75)	0.027 (5.69)	0.059 (8.26)	0.059 (8.22)
$YSM^2/100$	–0.053 (4.37)	–0.052 (4.25)	–0.132 (6.19)	–0.133 (6.13)
Married	0.030 (1.90)	0.033 (2.09)	0.001 (0.06)	0.003 (0.15)
South	0.023 (1.20)	0.023 (1.23)	0.128 (5.90)	0.129 (5.98)
Minority language concentration	[b]	–0.002 (2.04)	[b]	–0.004 (3.34)
$\bar{R}^2$	0.2550	0.2556	0.2706	0.2732
Sample size[c]	3,183	3,183	2,425	2,425

Notes

a Sample restricted to aliens from non-English-speaking countries; figures in parentheses are White's (1980) heteroskedasticity-consistent '*t*' statistics.
b Variable not entered.
c Weighted data with sample size scaled to actual number of observations.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice, 1989.

speaking proficiency and English-reading proficiency occurs in relation to the minority language concentration variable. In the case of English-speaking skills, this variable is highly significant, and has a coefficient of –0.006 for both males and females. For English-reading proficiency, however, the minority language concentration variable is only one-third the magnitude in the model of English-speaking skills for males, and two-thirds the impact in the case of females. In other words, while a concentration of individuals speaking the same non-English language as the respondent is associated with a marked reduction in English-speaking skills, there is less of an impact on English-reading skills. Presumably this reflects the fact that reading does not require a second party for this skill to be acquired and retained.[18]

Table 8.6 considers several extensions to the basic model. A variable for whether the person had entered the United States more than once prior to the

Table 8.6 Regression estimates of extended model of English-speaking and reading proficiency, male and female legalized aliens, Legalized Population Survey[a]

Variable	*Males*		*Females*	
	Speaking	*Reading*	*Speaking*	*Reading*
Constant	0.509	0.207	0.039	–0.115
	(7.94)	(3.13)	(0.51)	(1.51)
Age	–0.011	–0.008	–0.007	–0.007
	(9.81)	(7.49)	(6.49)	(6.07)
Educational attainment	0.034	0.052	0.047	0.053
	(16.28)	(26.49)	(19.50)	(22.94)
Years since migration	0.039	0.030	0.070	0.064
	(7.52)	(6.12)	(9.78)	(8.86)
$YSM^2/100$	–0.069	–0.055	–0.152	–0.137
	(5.27)	(4.57)	(7.15)	(6.43)
Married	0.025	–0.031	–0.010	–0.001
	(1.14)	(1.41)	(0.51)	(0.01)
South	–0.022	0.023	0.048	0.130
	(1.14)	(1.21)	(2.16)	(6.00)
Minority language concentration	–0.006	–0.002	–0.004	–0.003
	(4.74)	(1.50)	(2.69)	(1.80)
Multiple entry into U.S.	–0.040	–0.035	–0.055	–0.074
	(2.29)	(2.04)	(2.47)	(3.42)
Born in Mexico	–0.061	–0.019	–0.125	–0.055
	(2.00)	(0.62)	(3.39)	(1.43)
Born in Cent. and South America	–0.007	0.026	–0.040	–0.010
	(0.25)	(0.90)	(1.11)	(0.28)
Number of children (KIDS)	0.055	0.056	0.006	0.007
	(4.15)	(4.41)	(0.62)	(0.80)
KIDS squared	–0.007	–0.007	–0.001	–0.001
	(3.83)	(4.25)	(0.26)	(1.28)
$\bar{R}^2$	0.1935	0.2608	0.2630	0.2770
Sample size[b]	3,183	3,183	2,425	2,425

Notes

a Sample restricted to aliens from non-English-speaking countries; figures in parentheses are White's (1980) heteroskedasticity-consistent '*t*' statistics.

b Weighted data with sample size scaled to actual number of observations.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice, 1989.

application for temporary legal status is added to the model. This may reflect the effect of less time in the United States among those with multiple entries compared to a person with the same year of initial entry who never left the country. Moreover, it has been argued that individuals who plan to return to their country of origin invest less in U.S.-specific skills, including language, and hence have poorer English skills (see, for example, Chiswick and Miller 1998). This influence had been measured in Chiswick and Miller (1998) by average emigration rates for nationality groups. Such an approach is attributing to each individual a group propensity as a proxy for the expected emigration rate. The variable for multiple entries considered in the current study is a direct measure of the individual's previous international mobility. English language skills are expected to be inferior where the individual entered the United States more than once prior to applying for temporary legal status, other variables being the same.

Moreover, variables for country or region of birth are included in the estimating equation. As noted previously, the concentration of the sample among individuals from Mexico and Central and South America has dictated the use of just three birthplace groups: Mexico, Central and South America, and a Remainder group which serves as the benchmark. These country of birth variables will capture the effects of unmeasured variables that vary systematically by country of birth.

Finally, a variable for the number of children living in the household is included in the model. Chiswick and Miller (1992, 1994) report that in the United States and Australia children tend to have a more positive (or less negative) effect on the langauge skills of their fathers than of their mothers. The variable for the number of children in the household is entered in the model in quadratic form.[19]

The results in Table 8.6 have several main features. They show that legalized aliens who entered the United States more than once prior to applying for temporary legal status have lower levels of English-speaking and reading proficiency than those individuals who entered the United States only once. This effect is stronger for females than for males, especially in the case of English-reading skills.

In addition, legalized aliens from Mexico have lower levels of English-speaking skills than other groups, even after controlling for multiple entries. This differential is particularly noticeable among females where aliens from Mexico have a predicted level of English-speaking proficiency that is 13 percentage points lower than other groups (the equivalent of 3 years of schooling).

The variables for the number of children are significant for males, but not for females. Among men, proficiency in English increases at a decreasing rate for up to about four children, and the effect of children on fluency is positive for up to 8 children. Thus, children in the household are associated with greater English language proficiency for men but not for women.

The inclusion of the variables for the number of children in the estimating equation is associated with a reduction to statistical insignificance of the coefficient on the marital status variable. This finding is of importance for those studies such as Chiswick and Miller (1998) where, due to data limitations, the number of children cannot be held constant in the analysis conducted for males.

The sample is restricted to aliens born in Mexico in Table 8.7.[20] This table lists results for both males (left-hand columns) and females (right-hand columns). For each gender there are results for speaking and reading skills and, in the case of speaking skills, findings from the *LPS* and the 1990 Census are reported.

Table 8.7 Regression estimates of model of English-speaking and reading proficiency, males and females from Mexico

Variable	*Males*			*Females*		
	Speaking		*Reading*	*Speaking*		*Reading*
	LPS[b]	*1990 Census*	*LPS*[b]	*LPS*[b]	*1990 Census*	*LPS*[b]
Constant	0.422 (4.94)	0.427 (42.30)	0.108 (1.32)	–0.179 (1.67)	0.387 (34.40)	–0.261 (2.47)
Age	–0.011 (6.28)	–0.008 (39.10)	–0.006 (3.78)	–0.006 (3.76)	–0.007 (32.48)	–0.006 (3.79)
Educational attainment	0.034 (10.64)	0.028 (74.60)	0.055 (18.61)	0.050 (14.17)	0.032 (75.92)	0.054 (16.22)
Years since migration	0.041 (5.62)	0.028 (56.04)	0.030 (4.45)	0.077 (6.82)	0.026 (52.86)	0.071 (6.27)
$YSM^2/100$	–0.074 (4.15)	–0.029 (26.69)	–0.059 (3.61)	–0.002 (4.99)	–0.026 (23.03)	–0.164 (4.58)
Married	0.104 (4.33)	0.058 (15.67)	0.041 (1.76)	–0.011 (0.43)	–0.022 (5.72)	0.001 (0.02)
South	–0.007 (0.24)	0.031 (7.63)	0.041 (1.35)	0.096 (2.83)	0.042 (9.94)	0.187 (5.75)
Minority language concentration	–0.007 (3.96)	–0.010 (41.22)	–0.003 (1.65)	–0.006 (2.73)	–0.013 (48.47)	–0.004 (1.91)
$\bar{R}^2$	0.1372	0.2033	0.2004	0.2279	0.2470	0.2521
Sample size[a]	1,588	68,512	1,588	1,175	57,044	1,175

Notes
a Weighted data with sample size scaled to actual number of observations; figures in parentheses are White's (1980) heteroskedasticity-consistent '*t*' statistics.
b *LPS* = *Legalized Population Survey*.

Sources: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice, 1989. 1990 U.S. Census of Population and Housing, Public Use Microdata Sample, 5% sample.

Table 8.7 demonstrates that the model is quite robust when applied to an individual country, and that the basic patterns reported above apply to legalized aliens from Mexico. The latter finding is not surprising given the predominance of aliens from Mexico in this data set.

Hence, in summary, aliens who applied for temporary legal status under

Section 245(A) of the Immigration Reform and Control Act of 1986 appear to have relatively low levels of English-language skills. The processes determining their English fluency, however, seem to be governed by the same sets of factors that apply in the broader immigrant population, as revealed from the parallel analyses of the *LPS* and the 1990 Census. Specifically, greater English skills are possessed by the more educated, by those who entered the United States at a younger age, and those who have lived in the United States for a longer period of time. Living in a region where a high proportion of the residents speak the same non-English language as the individual is associated with poorer English-speaking ability, but not necessarily with inferior English-reading ability. Multiple entries into the United States prior to applying for temporary legal status are also associated with relatively inferior English language skills. The presence of children in the household has a positive effect on the fluency of males but no effect on females. The impact of multiple entries on language skills and the findings linking children to language skills among males offer important insights for the analysis of language skills among the broader immigrant population where these variables cannot generally be measured.

4. Language and earnings

The model

While many individuals see the main benefit of the acquisition of English-speaking and reading skills as the promotion of social cohesion and political integration, research in the economics literature has tended to concentrate on the more tangible benefits of the higher incomes that appear to be associated with proficiency in English. English language skills have been put on equal standing with formal schooling and on-the-job-training, and viewed as a form of human capital that has been acquired at current cost in the expectation of future returns. This literature then has set about assessing the expected economic return associated with proficiency in English. This has been achieved through the estimation of the human capital earnings equation (Mincer 1974), modified for the study of immigrant earnings (Chiswick 1978), and augmented with variables measuring the individual's proficiency in the English language.

Thus, the basic estimating equation is:

$$\text{LnY} = \text{f(schooling (+), experience (+), experience squared (–), duration of residence (+), duration of residence squared (–), language (+))},$$

where the dependent variable LnY is the natural logarithm of weekly earnings, "schooling" is the number of years of formal education the individual has received, "experience" in the number of years of potential labor market work, "duration of residence" is the number of years the individual has lived in the United States, and "language" is a measure of the language proficiency of the individual. Hypothesized signs are in parentheses.

In this specification, the experience variable effectively records the impact on earnings of experience prior to entering the United States. The duration of residence variable captures the premium to years of work experience in the United States compared to work experience in the country of origin. Other variables may be added to this equation to capture particular aspects of the earnings generation process (see Chiswick and Miller (1997) for an extensive discussion of the augmented earnings equation). In this study the standardizing variables are the natural logarithm of hours worked per week, marital status, residence in a Southern state, and birthplace variables that distinguish aliens born in Mexico and the Spanish-speaking countries of Central and South America from other countries of birth. See Appendix 8A for further details on the definition and construction of these variables.

There are several features of the data that impact upon the way the analysis is conducted. The earnings and hours worked data refer to the main job held by the respondent at the time of application for temporary residence (typically May 1987 to May 1988).[21] The earnings data are for usual weekly income from work, including overtime, commissions, and tips from all jobs, before taxes and deductions, at the time of application for legal status. Because it is usual weekly earnings at the time of application it is less subject to the recall error inherent in census and survey data that ask for earnings in the prior year. Since approval of "temporary legal status" was not contingent on earnings level, there was no particular reason for the applicant to provide false information, especially since a fraudulant application might have negative consequences.

The only information available on labor market experience is that constructed using the standard algorithm developed by Mincer (1974) whereby labor market experience is approximated by the difference between age and the school leaving age. This approximation relies heavily upon there being continuous attachment to the labor market after the completion of schooling. This approximation is generally held to be reasonably accurate for males, but will contain considerable measurement error in the case of females. For this reason the analysis of earnings in this study follows prior research and focuses on males. Some results for females are cited in footnotes.

As noted in the previous sections, data are available on both English-speaking skills and English-reading skills. This information is used in a number of ways. Variables for proficiency in speaking English and proficiency in reading English are included in separate earnings equations. Variables measuring each skill are also included in a single model of earnings. Finally, the data on English-speaking proficiency and English-reading proficiency analyzed in the previous section are used to construct four new variables: (i) proficient in reading and speaking English (56% of the sample); (ii) proficient in speaking English but not at reading English (11% of the sample); (iii) proficient in reading English but not at speaking English (5% of the sample); (iv) proficient at neither reading nor speaking English (28% of the sample). This alternative categorization permits an assessment of whether an individual is rewarded in the labor market for possession of just one language skill or whether both language skills are required to obtain higher earnings.

The partial effects of the language variables on earnings may reflect the economic returns (measured by earnings) to proficiency in spoken and written English. It is not possible to identify the extent to which accent discrimination or unobserved dimensions of intelligence or ambition that are correlated with English language proficiency may bias this coefficient.

The estimating equations

Table 8.8 reports estimates for a number of specifications of the earnings equations for males. All equations in this table refer to the full sample of male legalized aliens from non-English-speaking countries. A parallel set of analyses for legalized aliens from Mexico is presented in Table 8.9. Table 8.10 presents comparison equations from the 1990 U.S. Census for immigrants from non-English-speaking countries (columns (i) and (ii)) and for immigrants from Mexico (columns (iii) and (iv)).

The estimates of the earnings equations are obtained using Ordinary Least Squares. To the extent that the language variables are endogenous, these estimates may be biased. However, with the multiple measures of English proficiency used in the preferred specification, there are insufficient identifying instruments to pursue an instrumental variables approach. Some analyses that use an instrumental variables approach with a more restrictive specification of the estimating equation are presented in Chiswick and Miller (1995). These suggest that, if anything, the estimates presented below will provide an underestimate of the true return to English language skills.

The results in Table 8.8 column (i) are from a standard specification of the immigrant earnings equation that does not contain variables for English language skills. These results show that each additional year of schooling among legalized aliens is associated with around 3% higher earnings, *ceteris paribus*. This partial effect is only one-half that for all immigrants from non-English-speaking countries (Table 8.10).

Earnings increase at a decreasing rate with years of pre-immigration labor market experience. Evaluated at 10 years of experience, each additional year of experience is associated with 1.6% higher earnings. There is a sizeable earnings premium associated with duration in the United States. An additional year in the United States yields an additional 2% higher earnings (evaluated at 10 years of residence), to give a total earnings increase of 3.6% (1.6% plus 2.0%). These earnings gains associated with years of labor market activity are similar to the increase received by all immigrants from non-English-speaking countries (Table 8.10).

Weekly earnings are shown to increase with hours worked per week, although the elasticity at around 0.8 is far below unity. This is about the same

Table 8.8 Regression estimates of weekly earnings, male legalized aliens from non-English-speaking countries, 1987–1988

Variable	*(i)*	*(ii)*	*(iii)*	*(iv)*	*(v)*	*(vi)*
Constant	1.761 (8.06)	1.738 (7.88)	1.956 (8.78)	1.974 (8.86)	1.955 (8.76)	1.955 (8.79)
Education	0.032 (12.39)	0.030 (11.22)	0.022 (7.82)	0.021 (7.29)	0.021 (7.32)	0.021 (7.42)
Experience (Exp)	0.024 (7.95)	0.024 (8.07)	0.020 (6.69)	0.020 (6.54)	0.020 (6.67)	0.021 (6.86)
$Exp^2/100$	–0.042 (7.96)	–0.041 (7.86)	–0.038 (7.28)	–0.038 (7.21)	–0.038 (7.23)	–0.039 (7.39)
Year since migration (YSM)	0.033 (6.76)	0.030 (6.15)	0.034 (6.87)	0.036 (7.11)	0.034 (6.82)	0.034 (6.76)
$YSM^2/100$	–0.067 (5.53)	–0.061 (5.12)	–0.067 (5.59)	–0.069 (5.76)	–0.066 (5.57)	–0.066 (5.53)
Log hours worked	0.785 (13.79)	0.787 (13.71)	0.783 (13.63)	0.784 (13.62)	0.784 (13.60)	0.785 (13.66)
Married	0.111 (7.29)	0.105 (6.93)	0.115 (7.49)	0.120 (7.76)	0.116 (7.51)	0.116 (7.53)
South	–0.080 (4.51)	–0.078 (4.45)	–0.084 (4.75)	–0.087 (4.91)	–0.085 (4.80)	–0.084 (4.77)
Proficient in speaking English	[a]	0.081 (5.00)	0.073 (4.53)	[a]	0.055 (2.93)	[a]
Proficient in reading English	[a]	[a]	[a]	0.063 (3.96)	0.033 (1.76)	[a]
Proficient in reading and speaking English	[a]	[a]	[a]	[a]	[a]	0.083 (4.45)
Proficient in speaking but not reading English	[a]	[a]	[a]	[a]	[a]	0.010 (0.45)
Proficient in reading but not speaking English	[a]	[a]	[a]	[a]	[a]	–0.053 (1.58)
Birthplace						
Mexico	[a]	[a]	–0.149 (4.40)	–0.157 (4.64)	–0.150 (4.45)	–0.143 (4.24)
C. & S. America (Spanish)	[a]	[a]	–0.101 (2.80)	–0.106 (2.94)	–0.102 (2.84)	–0.098 (2.72)
$\bar{R}^2$	0.2725	0.2778	0.2839	0.2828	0.2843	0.2863
Sample size	2,936	2,936	2,936	2,936	2,936	2,936

Note

a Variable not entered. '*t*' statistics in parentheses computed using White's (1980) heteroscedasticity-consistent covariance matrix estimator.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice, 1989.

as the elasticity of 0.7 between weekly earnings and hours worked per week in Chiswick's (1991) study of illegal aliens. In other words, a 1% increase in the number of hours worked per week results in a less than 1% increase in weekly earnings.[22] This may be a reflection of the different standard number of hours of work in different jobs, with, for example, poorly paid work in the agricultural and manufacturing sectors involving a greater number of hours than unionized work in construction. Alternatively, it may be due to substantial measurement errors in the hours worked data.

Finally, in terms of the conventional variables, it is noted that both the marital status and South variables are highly significant. The results show that the married men have earnings around 11% higher than their non-married counterparts, while residents of the Southern states have earnings around 8% lower than residents of other states. Both impacts are of the same sign but are smaller in magnitude than the estimates for immigrants obtained from the 1990 U.S. Census.

The specification in Table 8.8 column (ii) includes a variable for English-speaking proficiency. Inclusion of this variable in the estimating equation has a minor impact on the other coefficients. The estimated coefficient reveals that legalized aliens who are able to converse in English have earnings 8% higher than aliens who lack this skill. This partial effect on earnings of English-speaking skills among the sample of legalized aliens is far lower than that estimated in the general population (see the 19% premium recorded in Table 8.10). The estimated premium of 8% is, however, considerably greater than the statistically significant impact of only 2 to 3% reported in Chiswick's (1991) study of illegal aliens.

Table 8.8, column (iii) augments the estimating equation with variables for birthplace. These results show that legalized aliens from Mexico have weekly earnings around 15% lower than the benchmark group, while legalized aliens from Central and South America have earnings 10% lower than the benchmark group from other regions. The addition of the birthplace variables in the estimating equation has a marked impact on the returns to education. The estimated payoff to an additional year of schooling falls by one-third, from around 3% to 2%. There is also a modest (one percentage point) reduction in the returns associated with English-speaking skills.

The fourth column of Table 8.8 replaces the variable for English-speaking skills with a variable for English-reading skills. This reading variable is highly significant and reveals that aliens who possess English-reading skills receive 6% higher earnings than aliens who lack this skill. The coefficient for English-reading skills is marginally less than the coefficient for English-speaking skills in column (iii), suggesting the English-speaking skills may be marginally more important.

In column (v) both the variables for English-speaking ability and English-reading ability are entered into the estimating equation at the same time. Both are statistically significant, although the coefficient for English-reading ability is only significant at the 10% level. The coefficients show that English-speaking ability is associated with 5% higher earnings, and English-reading ability with 3% higher earnings. These results contrast with those reported by Chiswick (1991) where English reading ability but not English-speaking ability affected earnings when both were entered into the equation.[23]

Finally, the results in Table 8.8 column (vi) provide further insights into the determinants of the economic returns to language skills. Among male legalized aliens, it is only those who possess both reading and speaking skills who are better off in the labor market. Holding constant country of birth and other determinants of earnings, aliens who can both read and speak English have 8% higher earnings than other language groups. The ability to speak English in the absence of the ability to read English, or the ability to read English in the absence of the ability to speak English, is not beneficial.[24] Remedial programs should therefore be addressed at all English language skills, not just spoken English.

Table 8.9 reports the estimates of models of earnings for legalized aliens from Mexico. It is quite clear that the basic results still hold. However, the returns from schooling for aliens from Mexico, at between 1.5 and 2.0%, is far less than that for all aliens. Chiswick (1991) reports a similar pattern in his analysis of earnings among illegal aliens, and the same differential effects of schooling on earnings for immigrants from Mexico and immigrants from all non-English-speaking countries is evident in the analysis of the 1990 Census presented in Table 8.10.

The partial effects of language skills on the earnings of legalized aliens from Mexico are similar to the effects reported for all aliens from non-English-speaking countries. Earnings are positively related to English-speaking and English-reading skills, although the evidence suggests that while speaking proficiency may be marginally more important than reading proficiency, neither skill by itself leads to higher earnings. Rather, it is necessary for aliens from Mexico to be able to both read and speak English in order to be paid more in the United States labor market. Where an alien from Mexico can both read and speak English, earnings are 9% higher than for those who possess only one or neither of these language skills.[25]

The estimated effect of the language variables on earnings may be reflecting a pure return to language skills in the labor market. They may, however, also be capturing the effects of unmeasured variables that are correlated with English language proficiency. Thus, those with less proficiency in English, other variables the same, may have less innate ability, less motivation, or a shorter expected duration in the United States. This would imply an upward bias, that is, the true effect of language is smaller than what is measured here. On the other hand, purely random measurement errors would bias the coefficient of the language variable downward rather than upward. In other studies an instrumental variables approach has been used to address this issue (see, for example, Chiswick and Miller 1995). When this is done the measured

Table 8.9 Regression estimates of weekly earnings, male legalized aliens from Mexico, 1987–1988

Variable	*(i)*	*(ii)*	*(iii)*	*(iv)*	*(v)*
Constant	2.043 (6.14)	2.022 (6.01)	2.035 (6.04)	2.023 (5.99)	2.023 (6.01)
Education	0.019 (5.05)	0.017 (4.57)	0.015 (4.07)	0.015 (4.15)	0.016 (4.29)
Experience (Exp)	0.020 (4.73)	0.020 (4.86)	0.019 (4.69)	0.020 (4.81)	0.020 (4.99)
$Exp^2/100$	–0.039 (5.39)	–0.038 (5.32)	–0.038 (5.25)	–0.038 (5.27)	–0.039 (5.41)
Year since migration (YSM)	0.033 (5.32)	0.030 (4.79)	0.031 (5.01)	0.030 (4.76)	0.029 (4.70)
$YSM^2/100$	–0.063 (4.26)	–0.057 (3.90)	–0.060 (4.06)	–0.057 (3.89)	–0.056 (3.85)
Log hours worked	0.745 (8.56)	0.747 (8.50)	0.748 (8.48)	0.748 (8.47)	0.750 (8.53)
Married	0.131 (6.41)	0.123 (6.05)	0.129 (6.33)	0.124 (6.08)	0.125 (6.12)
South	–0.095 (3.83)	–0.093 (3.80)	–0.097 (3.92)	–0.095 (3.85)	–0.094 (3.82)
Proficient in speaking English	a	0.077 (3.68)	a	0.058 (2.33)	a
Proficient in reading English	a	a	0.068 (3.32)	0.036 (1.50)	a
Proficient in reading and speaking English	a	a	a	a	0.089 (3.74)
Proficient in speaking but not reading English	a	a	a	a	0.010 (0.35)
Proficient in reading but not speaking English	a	a	a	a	–0.054 (1.20)
$\bar{R}^2$	0.2564	0.2627	0.2610	0.2632	0.2660
Sample size	1,469	1,469	1,469	1,469	1,469

Note
a Variable not entered. '*t*' statistics in parentheses computed using White's (1980) heteroscedasticity-consistent covariance matrix estimator.

Source: Legalized Population Survey, Immigration and Naturalization Service, Department of Justice, 1989.

partial effect of language proficiency on earnings is considerably larger than the estimated OLS effect. Because of the problem of selecting identifying instruments in the *LPS* data, the instrumental variables technique is not used here.

Table 8.10 Regression estimates of annual earnings, male immigrants from non-English-speaking countries, 1990 Census

Variable	*Non-English-speaking countries*		*Mexico only*	
	(i)	*(ii)*	*(iii)*	*(iv)*
Constant	4.631 (169.61)	4.603 (169.12)	5.184 (114.93)	5.171 (114.99)
Education	0.062 (142.72)	0.056 (120.67)	0.029 (31.53)	0.025 (28.01)
Experience (Exp)	0.024 (35.70)	0.025 (38.56)	0.015 (12.63)	0.016 (13.71)
$Exp^2/100$	–0.034 (28.30)	–0.035 (29.66)	–0.027 (13.26)	–0.027 (13.33)
Year since migration (YSM)	0.029 (52.76)	0.025 (45.19)	0.029 (29.99)	0.025 (25.61)
$YSM^2/100$	–0.036 (28.83)	–0.032 (25.38)	–0.036 (15.05)	–0.032 (13.37)
Log weeks worked	0.980 (137.02)	0.973 (136.16)	0.920 (82.36)	0.914 (81.87)
Married	0.226 (57.97)	0.219 (56.39)	0.206 (33.37)	0.198 (32.17)
South	–0.133 (32.51)	–0.131 (32.04)	–0.186 (27.26)	–0.187 (27.57)
Proficient in speaking English	[a]	0.193 (41.73)	[a]	0.152 (24.52)
$\bar{R}^2$	0.4002	0.4050	0.4064	0.4121
Sample size	212,384	212,384	61,700	61,700

Note

a Variable not entered. '*t*' statistics in parentheses computed using White's (1980) heteroscedasticity-consistent covariance matrix estimator.

Source: 1990 U.S. Census of Population and Housing, Public Use Microdata Sample, 5% sample.

5. Summary and conclusions

This chapter uses the data on adult males and females from the 1989 *Legalized Population Survey* to study the determinants of English language proficiency (speaking and reading skills) and earnings among those who applied for legal status under one of the provisions of the Immigration Reform and Control Act of 1986. In principle the applicants had lived in the United States continuously in an illegal status since before January 1, 1982 when they applied for legal status in 1987/1988.

The analysis of English-language proficiency is based on a model that relates destination language proficiency to economic incentives, exposure to the destination language, and efficiency in acquiring languages skills. The empirical analysis reveals that the variables affecting language skills in the general immigrant population also affect language skills in this legalized alien population.

English language proficiency, whether fluency or literacy, increases with the level of education and duration of residence, and is lower for those who immigrated at an older age or who live in an area where many others speak the same non-English language. Those who had multiple entries since their initial entry are less fluent, perhaps because of more time outside the United States or less investment in English language skills while in the United States. Children in the household have a positive effect on the English language skills of adult men, but have no effect on the language skills of adult women. This suggests that their greater role as translators for their mothers and their adverse effect on their mother's labor force participation may offset other positive effects they have on both their mother's and father's language skills. Even when other variables are the same, Mexican legalized aliens have a lower level of English-speaking skills than those from Central and South America and the rest of the world.

The analysis of usual weekly earnings at the time of application for legalization is reported in detail for men, with highlights reported for women. Among men, earnings increase with schooling (coefficient 0.03), labor market experience, and duration in the United States, and are higher for married men (cofficient 0.11). The effects of schooling and marriage are smaller among legalized aliens than among immigrant men in general. Other things the same, Mexican legalized aliens have the lowest earnings (approximately 15% lower), followed by those from Central and South America (approximately 10% lower) compared to legalized aliens from all other countries.

Earnings increase with both speaking and reading skills, whether analyzed separately or jointly, among both males and females. Earnings are higher by about 8% for males and by 17% for females proficient in both speaking and reading, compared to those lacking both skills. Possessing one but not the other skill has no significant effect on earnings. Among men, it appears that these effects on earnings are much smaller than what is found among immigrants in the decennial census, other variables the same.

In summary, the analyses demonstrate that the model of destination language acquisition is robust for the study of illegal aliens. Their language skills are responsive to the explanatory variables in a manner similar to that of legal immigrants. Moreover, destination language skills are shown to be important determinants of earnings even among illegal aliens.

Appendix 8A: Definitions of variables

The definitions of all variables are given below. All of the variables except CONC are from the *Legalized Population Survey*.

Age: Information on year of birth (A1YEAR) is used to determine the age of all individuals aged 20–64 years used in the statistical analysis. Valid information is available for all individuals. The experience variable (EXP) is computed as age minus schooling minus five.

Gender: Gender is determined from the codes to the variable on sex (A3). All respondents have valid information on this variable.

Speak/Read: Information was collected in the survey on a range of aspects of English proficiency, namely:

- ability to read and understand a newspaper written in English (B2A);
- ability to read and understand a magazine written in English (B2B);
- ability to read and understand recipes written in English (B2C);
- ability to speak to a sales clerk in English (B3A);
- ability to speak to a doctor, nurse or teacher in English (B3B);
- ability to speak on the telephone in English (B3C).

The percent of respondents giving affirmative answers to these questions are (among valid responses): 62.0, 61.9, 68.3, 69.7, 67.2, 65.4%, respectively. There is a high degree of overlap between affirmative responses. For example, on the questions on English-speaking skills the overlaps are in excess of 90%. There should, therefore, be little difference in the statistical results if the English-language proficiency variable is constructed using one specific question rather than an alternative question. In the analyses presented in the text, the ability to speak to a doctor, nurse or teacher in English is used as the measure of English-speaking proficiency and the ability to read and understand a newspaper written in English is adopted as a measure of English-reading proficiency.

Second language: The survey contains a set of questions on the language a person speaks best. This information is used to construct a variable on the second language spoken. The sequence of questions allows Spanish and one other non-English language to be recorded, and in the small number of cases (less than one-half of one percent) where this happens Spanish is assumed to be the main non-English language spoken. (B1A, B1CSPEC)

Years since migration (YSM): Respondents were asked questions on the time they entered the United States prior to their application for temporary residence. Among individuals who entered the United States only once, the year of entry (C3YEAR) is used as the reference point in the construction of the years since migration variable. For individuals who entered the United States more than once prior to applying for temporary residence, the year in which they first entered the United States is used in the construction of the years since migration variable (C5YEAR). For eight respondents with invalid information on the year in which they first entered the United States but with valid information on the year of last entry into the United States (C6YEAR), the latter information was used. Consequently, only 26 respondents have invalid information on this variable (10 who entered the United States only once and 16 who entered more than once or did not specify the number of times they entered the United States).

Multiple entry: Respondents who had entered the United States more than once prior to their application for temporary legal status are distinguished from all others on the basis of information in the survey question C2. 42 individuals in the sample of 6,193 had invalid information on this characteristic.

Marital status: The individual's marital status at the time of application for temporary residence is used in the statistical analysis. Individuals who were married and living with their spouses in the same household or were living with a partner as if married were classified as married in the analysis (E1).

Education: The education variable is defined as the highest grade of regular school at the date of the survey (H1A). 21 individuals have incomplete information on this characteristic.

Birthplace: The birthplace information refers to country or region of citizenship (COC). All individuals have reported a valid place of birth.

Earnings: The earnings variable is the usual weekly income from work at the time of application for temporary residence before taxes or deductions but including overtime pay, commissions and tips usually received (D37). It is used in the earnings equation in natural logarithmic form.

Hours: The variable "Hours" is the number of hours per week the individual worked in the main job at the time of application for temporary residence (D36). It is entered in the earnings equation in natural logarithmic form.

South: The variable for residence in a Southern state is constructed from the survey information on state of residence at the time of interview. 26 states are represented in the data (STATE), and the states categorized as "South" are: Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia. Valid information on this characteristic is available for each individual.

Relatives/Children: These variables are defined as the number of relatives (grandparents, parents, siblings from either side of the family) and children (children and grandchildren from either side of the family) living in the same household at the time of application for temporary residence (E2, E3, E4, E5, E6). Four respondents have missing information on these characteristics.

Minority group concentration (CONC): Each respondent is assigned a measure equal to the percentage of the population aged eighteen to sixty-four in the state in which he/she lives, who report the same non-English language as the respondent. These percentages are computed from the microdata files from the 1990 Census of Population. In the construction of this variable, only the twenty-four largest nationwide language groups are considered. In descending order there are: Spanish; French; German; Italian; Chinese;

Tagalog; Polish; Korean; Vietnamese; Japanese; Portugese; Greek; Arabic; Hindi; Russian; Yiddish; Thai; Persian; French Creole; Armenian; Hebrew; Dutch; Hungarian; Mon-Khmer. These constitute 94% of all responses to the question concerning the use of a language other than English in the home. Representation in the other language groups (6% in total) is so small numerically that the proportions are approximately zero, and this value is assigned. Those who reported only English are also assigned a zero value. For further information, see Chiswick and Miller (1998).

For further information on the *Legalized Population Survey*, see *Documentation for the Legalized Population Survey Public Use Tape*, produced by the Statistics Division, Office of Strategic Planning, Immigration and Naturalization Service, United States Department of Justice.

Notes

1 The research for this chapter was funded, in part, by the Bureau of International Labor Affairs (ILAB), U.S. Department of Labor. We appreciate the comments from Shirley Smith, the ILAB Project Officer. We wish to thank Richard Hockey, Department of Public Health, The University of Western Australia, for assistance with the SAS package. Helpful comments were received from the American Economics Association Annual Meeting, New Orleans, January 1997, The Midwest Economics Association Annual Meeting, Chicago, March 1998, and the Human Resources Workshop, University of Illinois at Chicago, 1997.

2 See McManus et al. (1983) for one of the earliest studies of the effect of English language skills on the earnings of an immigrant group, Hispanic men.

3 Chiswick's (1991) sample was of illegal aliens apprehended in the Los Angeles area who were interviewed in the Immigration and Naturalization Service (INS) Los Angeles detention center.

4 Temporary legal status, the term established in the legislation, is a misnomer, as "transitional" legal status is a more accurate term. Nearly all recipients of this transitional status received permanent resident alien ("green card") status after the required waiting period. For an economic analysis of the provisions of the 1986 Act see Chiswick (1988).

5 Chiswick and Miller (1996a) show that for immigrants in Australia, having relatives in Australia other than spouse and children lowers English language proficiency. Chiswick and Miller (1994, 1995) show that having a spouse with a foreign origin tends to lower destination language proficiency and that children have a more negative or less positive effect on language fluency for their mothers than for their fathers.

6 The U.S. Census in 1980 and 1990 asked about fluency but not about literacy. A similar data problem exists for the Australian and Canadian censuses. While Chiswick and Miller (1996a) examine English reading, writing and speaking skills in their analysis of immigrant language skills in the Multicultural Survey in Australia, the skip-patterns in this survey mean that the information in the various dependent variables was not independent. While reading and speaking skills are naturally linked, there are no survey-imposed dependencies in the skip patterns in the *Legalized Population Survey*. Analyses of the fluency and literacy of immigrants in the dominant language have been conducted using survey data for Germany (Dustmann 1994 and Israel (Beenstock 1993) and census data for Israel (Chiswick and Repetto 1998 [2001]).

7 Other studies that have addressed these issues include Beenstock (1993), Chiswick (1991) (Rivera-Batiz 1996) and Dustmann (1994) using survey data and Chiswick and Repetto (1998) using census data.

8 Approximately 1.3 million other individuals applied for legal status under the seasonal farm worker (SAW) provisions of the 1986 Act. We are not aware of any systematic survey of this population.

9 In the first stage, 40 legalization offices were selected. In the second stage, subsamples of applicants within sampled legalization offices were selected.

10 Approximately two-thirds of the sample were followed up in 1992. However, as the initial set of 6,193 observations is a relatively small sample for the detailed statistical analyses considered below, in order to maximize sample size only the information from the initial survey in 1989 has been used in this analysis. Moreover, changes to the wording of certain key questions preclude viewing the two panels as a longitudinal data set (see Chiswick and Miller 1999).

11 For a comparison of the language questions in the *LPS* with the language questions in a 1992 follow-up survey of this population (*LPS2*), as well as the 1990 Census see Chiswick and Miller (1999).

12 Around 600 observations from the initial sample of 6,193 aliens are not used in the statistical analysis due to the focus on 20–64-year olds, or due to missing information on variables included in the estimating equation. Appendix 8A describes the variables used and provides information on the extent of missing information.

13 Experiments with both variables revealed that they were statistically insignificant in the analysis of the *Legalized Population Survey*.

14 This sample is therefore quite different from the sample of young recently arrived illegal aliens studied by Chiswick (1991), where the mean age was 23 years, the mean duration in the United States was 1.5 years, and the mean educational attainment was around 7 years.

15 See, for example, Long (1990), Service and Craik (1993), and the references therein. In the *LPS* data and in general a quadratic term for the age variable adds little to the explanatory power, so a linear specification is a good approximation, and is used here.

16 The data available preclude identifying the birthplace of the spouse. Since the respondents in this survey were in an illegal status when they applied for amnesty, it is highly unlikely that their spouse was a U.S. citizen at the time of application.

17 This result differs from that reported by Chiswick (1991) where English reading skills are greater among those who migrated at an older age.

18 Following Chiswick (1991), equations were estimated where a variable that records English-speaking proficiency was included in the model of English-reading proficiency. As with Chiswick (1991), this modification of the specification is associated with reductions in the estimated partial impacts of educational attainment and duration of residence. The estimated partial effect of age at migration is also diluted. However, as with Chiswick's (1991) analysis, even after controlling for English-speaking proficiency, those aliens with more schooling and who have been in the United States for a longer period of time are more likely to be proficient at reading English.

19 A variable for the number of other relatives living in the household at the time of application for legal status was also considered, but was found to be statistically insignificant for both males and females.

20 To increase statistical efficiency, legalized aliens not of Mexican origin were oversampled in the *LPS*. Thus the sample sizes for legalized aliens born in Mexico are smaller in Table 8.7 than implied by the weighted proportion in the full sample.

21 Tienda and Singer (1995) and Singer (1996) analyze earnings and employment using the *LPS* data, but do not consider the effects of language proficiency. They

argue that there were no changes in the unmeasured dimensions in cohort quality over time in the *LPS* sample.

22 It is noted that the estimated impact of weeks worked on annual earnings in the 1990 U.S. Census (Table 8.10) is much closer to unity (0.98) than that the estimated elasticity coefficient of 0.8 between weekly earnings and hours worked per week.

23 Comparable equations were estimated for females, and the results for language skills parallel those reported for males. Hence, specifications identical to those in columns (iii), (iv) and (v) yield the following results for females: speaking skills entered separately (10.9 percent higher earnings, '*t*' = 4.46); reading skills entered separately (10.6% higher earnings, '*t*' = 4.22); speaking and reading skills entered simultaneously (for speak, 7.5% higher earnings, '*t*' = 2.79; for read, 6.8% higher earnings, '*t*' = 2.46).

24 Among female legalized aliens, 14% higher earnings are received by those who can both read and speak English compared to those who can do neither. There is no return to possessing only one of the two language skills analyzed in this study.

25 Among female legalized aliens from Mexico, those who can both speak and read English have earnings 17% higher than those who possess neither skill or only one of these language skills. The larger partial effect of English language skills on earnings among females may reflect the occupational segregation of females into service occupations where English language skills have a larger pay-off.

References

Beenstock M (1993) Learning Hebrew and Finding a Job: Econometric Analysis of Immigrant Absorption in Israel. Paper presented at the Conference on the Economics of International Migration: Econometric Evidence, Konstanz, Germany, February

Chiswick BR (1978) The Effect of Americanization on the Earnings of Foreign-born Men. *Journal of Political Economy* 86(5):897–921

Chiswick BR (1984) Illegal Aliens in the United States Labor Market: Analysis of Occupational Attainment and Earnings. *International Migration Review* 18(3):714–732

Chiswick BR (1988) Illegal Immigration and Immigration Control. *Journal of Economic Perspectives* 2(3):101–111

Chiswick BR (1991) Speaking, Reading and Earnings Among Low-Skilled Immigrants. *Journal of Labor Economics* 9(2):149–170

Chiswick BR (1998) Hebrew Language Usage: Determinants and Effects on Earnings Among Immigrants in Israel. *Journal of Population Economics* 11:253–271

Chiswick BR, Miller PW (1992) Language in the Immigrant Labor Market. In: Chiswick BR (ed) *Immigration, Language and Ethnicity: Canada and the United States*. American Enterprise Institute, Washington, 229–296

Chiswick BR, Miller PW (1994) Language and Labor Supply: The Role of Gender Among Immigrants in Australia. In: Slottje DJ (ed) Research on Economic Inequality. JAI Press, Greenwich, CT 153–190

Chiswick BR, Miller PW (1995) The Endogeneity Between Language and Earnings: International Analyses. *Journal of Labor Economics* 13(2):246–288

Chiswick BR, Miller PW (1996a) Ethnic Networks and Language Proficiency Among Immigrants. *Journal of Population Economics* 9:16–25

Chiswick BR, Miller PW (1996b) The Languages of the United States: What is Spoken and What it Means. READ Perspectives 3(2):5–41

Chiswick BR, Miller PW (1997) Language and Earnings Among Immigrants. Department of Economics, University of Illinois at Chicago, xerox

Chiswick BR, Miller PW (1998) English Language Fluency Among Immigrants in the United States. *Research in Labor Economics* 17:151–200

Chiswick BR, Miller PW (1999) Language Skill Definition: A Study of Legalized Aliens. *International Migration Review* 32(4):877–900

Chiswick BR, Repetto G (1998 [2001]) Immigrant Adjustment in Israel: Literacy and Fluency in Hebrew and Earnings. Department of Economics, University of Illinois at Chicago, (Republished as: Immigrant Adjustment in Israel: The Determinants of Literacy and Fluency in Hebrew and the Effects on Earnings, in Slobodan Djajic (ed.), *International Migration: Trends, Policy and Economic Impact*, New York: Routledge, 2001, pp. 204–228)

Dustmann C (1994) Speaking Fluency, Writing Fluency and Earnings of Migrants. *Journal of Population Economics* 7(2):226–236

Immigration and Naturalization Service (undated). *Documentation for the Legalized Population Survey Public Use Tape*. Statistics Division, Office of Strategic Planning, United States Department of Justice, Washington, D.C.

Kee P (1993) *The Economic Status of Male Immigrants in the Netherlands*. University of Amsterdam, Amsterdam

Long MH (1990) Maturational Constraints on Language Development. *Studies in Second Language Acquisition* 12(3):251–285

McManus W, Gould W, Welch F (1983) Earnings of Hispanic Men: The Role of English Language Proficiency. *Journal of Labor Economics* 1(2):101–130

Mincer J (1974) *Schooling, Experience and Earnings*. National Bureau of Economic Research, New York

Rivera-Batiz FL (1996) English Language Proficiency, Quantitative Skills, and the Economic Progress of Immigrants. In: Duleep HO, Wunnava PV (eds) *Immigrants and Immigration Policy: Individual Skills, Family Ties, and Group Identities*. JAI Press, Greenwich, CT, 57–77

Service E, Craik FIM (1993) Differences Between Young and Older Adults in Learning a Foreign Vocabulary. *Journal of Memory and Language* 32:608–623

Singer A (1996) Changes in the Employment and Earnings of the Legalized Population. Working Paper, Division of Immigration Policy and Research, Bureau of International Labor Affairs, U.S. Department of Labor

Tienda M, Singer A (1995) Wage Mobility of Undocumented Workers in the United States. *International Migration Review* 29(1):112–138

White H (1980) A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48:817–838

9 Language and labor supply

The role of gender among immigrants in Australia

I. Introduction

The development of dominant language fluency is generally viewed as an important step in the economic and social adjustment of immigrants in the destination country. In Australia, immigration policy was historically based on an "assimilation model" whereby immigrants were expected to "embrace wholeheartedly the Australian way of life and deny and forget their origins" (Secretariat to the Committee to Advise on Australia's Immigration Policies 1987, p. 14). Learning English was an integral part of the assimilation process. More recently, a "multiculturalism model," fostered by the federal government, has placed greater emphasis on origin-language maintenance, though with immigrants still expected to adapt to a core of Australian institutions and values, English language fluency presumably retains a major role in the adjustment process. However, despite the substantial immigrant presence (over one-fifth of the Australian population was born overseas) and the emphasis on language skills in government policy, there is a paucity of research into the determinants and consequences of learning the English language among immigrants in Australia. In this way the Australian literature mirrors the slim literature for the other major immigrant-receiving countries (e.g., the United States, Canada, and Israel).

The studies of the determinants of dominant language skills among adult *male* immigrants in Australia, Canada, Israel, and the United States (Chiswick and Miller 1992a, 1995) reveal a robust process both within and across these three countries. Dominant language skill is related to efficiency in learning the language, exposure to the language, and economic incentives for acquiring fluency. It is shown to be related to particular efficiency variables, that is, positively related to the immigrant's educational attainment and negatively related to age at arrival. It is also positively related to exposure variables, such as whether the language is spoken in the origin, duration of residence in the destination, and the extent to which the minority (origin) language is spoken in the area of residence (minority language concentration) and in the home (language of the spouse and presence of children). These studies also demonstrate that dominant language skills are rewarded in the labor market (as measured by earnings) and that these rewards influence the decision to acquire dominant language fluency. That is, dominant language skills are greatest among those for whom the expected increases in earnings are greatest. One exception, however, is that the nexus between labor market rewards and dominant language skills is weaker in Australia.

This raises an interesting issue: if the expectations of labor market rewards encourage dominant language fluency, there should be a difference in English language fluency between those who anticipate full-year, full-time labor market attachment for their entire working life, and those with expectations of intermittent, part-year, or part-time involvement in the labor market.[1] Thus, there is expected to be a difference in Australia between males and females in the acquisition of English language skills. Is it possible to reconcile this near equality in English language fluency and the divergence in participation rates?

Section II presents a brief statement of the model of dominant language fluency. In Section III the dataset utilized, the 1986 Australian Census of Population and Housing, is introduced and the empirical results pertaining to the models of dominant language fluency are presented and discussed. Analyses are undertaken for both males and females. The role of language skills in labor supply decisions is presented in Section IV. Section V contains a summary and conclusion.

II. Dominant language fluency among immigrants

Table 9.1 lists the major features of three studies of English language fluency using Australian data: Evans (1987), McAllister (1987), and Chiswick and Miller (1995).[2] These studies adopt techniques and specifications of the estimating equation that are representative of the North American literature [see Chiswick and Miller (1992a) and the references therein].

The determinants of dominant language fluency can be divided into two categories: personal characteristics and environmental (or background) influences.[3] The former includes educational attainment, age at arrival, and duration of residence. The latter includes country of origin, whether the person married prior to migrating, current family circumstances, and the language characteristics of the area in which the individual resides. The empirical findings relating to these variables are clear from Table 9.1, and these results accord with the effects hypothesized in the various studies. A brief outline of the bases for the relationships expected between dominant language fluency and a number of personal characteristics and environmental factors is provided below. This outline draws upon Chiswick and Miller (1992a, 1995).

Educational attainment is expected to have a positive relationship with dominant language fluency for a number of reasons. First, individuals who are more able are expected to receive higher benefits per dollar of investment, whether in formal education (see Becker 1975) or in language skills. For any given marginal cost of financing human capital investment, therefore, the

Table 9.1 Summary of major studies of English language fluency among immigrants in Australia

Study/Primary datasets	*Populations studied/ Primary methodology*	*Dependent variable*	*Explanatory variables*	*Major findings*
Evans (1987) Census of Population and Housing, 1981	Males aged 25–64 working full time by birthplace region; least squares regression	Categorical data on English fluency converted to a continuous measure	Education, experience, duration of residence	Language proficiency positively related to educational attainment and negatively related to age at arrival
McAllister (1987) General Survey, 1983	Individuals aged 15–30; least squares regression	Dichotomous measure of origin language maintenance Continuous measure of English proficiency formed from categorical data on English speaking, understanding, reading, and writing skills	Age, gender, duration of residence, father's occupational status, citizenship, social network variable	English language proficiency positively related to duration of residence and education, and negatively related to age at migration; Reverse effects on origin language maintenance
Chiswick & Miller (1995), Census of Population & Housing, 1981, 1986	Employed males aged 25– 64; least squares regression	Dichotomous measure of English proficiency constructed from census categorical data	Age at arrival, education, duration of residence, location, marriage, children, language concentration, birthplace	Language proficiency increases with duration of residence and educational attainment and negatively related to foreign marriage, age at arrival, and minority language concentration Important interactions between minority language concentration and age at arrival, education, and duration of residence

more able will invest larger amounts in both schooling and language skills than their less able counterparts. Second, there may be complementarities between formal education and other means of acquiring human capital, including language skills. Hence the better educated may simply be more efficient at acquiring language capital. Third, those with more schooling are more likely to have been exposed to English in their preimmigration schooling. English is introduced in many curriculums in non-English-speaking countries in upper grades. Finally, those with more schooling in Australia are likely to be more proficient because English language skills may be a prerequisite for admission, and Australian schooling would further enhance these skills. On the other hand, the Australian point system for issuance of visas for independent immigrants, which awards points for both schooling and English fluency, could impart a negative correlation.

A positive correlation is also expected between duration of residence and dominant language fluency. Other variables the same, this arises from exposure, or "learning by doing," in an environment. Language fluency may also rise with duration of residence if there is selective emigration, that is, if those with greater difficulty learning English are more likely to remigrate. To the extent that labor market work provides both greater exposure to English and greater rewards from fluency in English compared to home production, it would be expected that English language skills would increase less sharply with duration of residence for women than for men.

A personal characteristic that would be expected to impact negatively upon dominant language fluency is age at arrival. Youths have a greater facility for learning languages than older people (Long 1990), and are more likely to be placed in circumstances that render the learning of the dominant language easier, for example, association with native-born children and exposure to the dominant language in schools. An additional consideration is that individuals who migrate at a later age have disincentives because of a higher opportunity cost of time, and at still later ages may take the view that they have too little time left to reap the rewards from investments in learning the language of the country of destination.

A range of environmental/background factors has been included in the various studies. Country of origin is of obvious importance, with the major consideration being the extent to which the dominant language of the destination country was used in the country of origin, either as the sole or primary language (e.g., English in the United States) or as a *lingua franca* (e.g., English in India). Immigrants exposed to the dominant language of the destination country prior to migration are expected to arrive with greater fluency and to become more fully fluent more readily than other immigrant groups. Thus, immigrants from English-speaking regions (e.g., United Kingdom, United States, New Zealand, Canada) would normally be expected to have high levels of dominant language fluency in Australia. Immigrants from regions that had significant U.S. or British colonial experience (e.g., Philippines, India) are more likely to have had some exposure to English prior to migration, and thus may arrive with greater proficiency and reach a high level of proficiency sooner than immigrants without prior exposure to English. Refugees (e.g., the Vietnamese) would be less fluent in English for two reasons. One is that fluency is a less explicit criterion for admission.[4] Furthermore, refugees were less likely to have planned for the move and are moving for reasons other than the easy transferability of their skills, and hence include a large number of individuals with poor fluency. The Chiswick and Miller (1995, 1992a) studies of adult men in Australia, Canada, Israel, and the United States show that the pattern of fluency rates by country of origin is consistent with these arguments.

Another background factor that captures preimmigration characteristics is foreign marriage. Foreign marriage is, ceteris paribus, more likely to be marriage to a person with the same non-English mother tongue. It is therefore likely to result in the immigrant receiving less exposure in the home to the dominant language and thus facing a higher cost of acquiring dominant language fluency. In addition, as non-dominant-language-speaking immigrants would be able to converse in their origin language in the home, there would be fewer benefits from learning the dominant language.

Within the family context, the presence of children may also exercise an influence on parents' language skills, although the net direction of influence cannot be determined a priori. As noted earlier, children appear to have a greater facility for acquiring language skills, and their greater exposure to the dominant language through the school system makes it highly likely that they will develop fluency at an early age. These skills may then be passed on to parents at home, meaning that parents may learn the dominant language from their children. It is also possible that children can act as interpreters for their parents, thereby lessening the benefits for adults from acquiring dominant language fluency. This is more likely to be relevant in consumption and home activities than in the labor market. Thus, the "children as interpreters" role is more compelling the smaller the extent to which English language fluency is determined by (or is in response to) labor market outcomes. This point would be more compelling for women, particularly nonworking women, than for men. This implies a less positive or more negative effect of children on the language fluency for women than for men.

A final background factor, included in some analyses, is the language composition of the area in which the individual lives. The presence of a sizable segment of the population speaking the same nondominant language as the individual would be expected to retard the development of dominant language skills.[5] This effect could evolve from the substitution of conversations in the origin language for conversations in the dominant language, and the development of nondominant language enclave markets, both labor markets and markets for consumer goods.[6] Note that the greater the importance of minority language enclaves in the labor market, the smaller the economic penalty from not being fluent and the weaker the incentives for fluency. This effect is expected to be greater the greater the person's labor market attachment, whether measured by hours worked per year, or number of years worked. Thus, the negative effect of minority language concentration on fluency is hypothesized to be larger (i.e., more negative) for men than for women.

In summary, the model outlined above points to a positive relationship between dominant language fluency and both education and period of residence, and a negative relationship between dominant language fluency and age at arrival, foreign marriage, and the minority language concentration within the region of residence. Furthermore, it is hypothesized that the positive effect of duration of residence is smaller for women, the effect of minority language concentration is less negative for women, and the effect of children is less positive or more negative for women than for men. These hypotheses are tested in the econometric analysis.

III. Data and empirical results for the model of English language fluency

The statistical analyses in this paper are based on the 1 percent public use sample drawn from the Census of Population and Housing conducted in Australia on June 30, 1986.[7] This dataset has a number of features that affect the analysis. Those who reported the use of a language other than (or in addition to) English in the home were asked to identify this language, something that was not done in the 1981 Census. Twelve languages other than English are identified in the 1986 data. This information allows the construction of a minority language concentration variable similar to what was used in analyses for men in Australia, Canada, Israel, and the United States (Chiswick and Miller 1995, 1992a). Offsetting this advantage are four weaknesses: data on age, period of residence in Australia, and duration of marriage are available only in broad intervals, and the detail on country of birth in the public use sample is considerably less than for the 1981 Census (only 40 foreign birthplaces in the 1986 data compared to 100 in the 1981 data). Comparisons of results from the 1986 data with results derived from the 1981 data (where the age and period of residence variables were available in individual years) suggest that the different degree of detail on year of birth and year of arrival are, at most, of minor consequence. However, the limitations of the 1986 data effectively prevent the construction of a foreign marriage variable.[8] An additional important feature, however, of the 1986 data is the greater number of non-European immigrants and their longer average duration of residence.

Dominant language fluency in these analyses is measured from the self-reported responses to the question ‘How well do you speak English?’ asked of those who speak a language at home other than or in addition to English. Responses were coded into four mutually exclusive groups: very well, well, not well, and not at all. Individuals who spoke only English at home and those who spoke English “very well” are classified as fluent in English for the purpose of these analyses.[9]

On this basis, the fluency rates were 73.9 percent for males and 71.0 percent for females, a difference of 2.9 percentage points. Considering the substantial differences in labor force participation rates between males and females, the similarity of the mean fluency rates is surprising. It may be indicative of a situation for women where language skill acquisition is governed more by basic socioeconomic and demographic variables (efficiency and exposure variables) than by direct labor market factors. Further insights in this regard are obtained from the estimates of the model of dominant language fluency presented in Table 9.2.

Table 9.2 Regression estimates of dominant language fluency model among immigrants, by gender, Australia 1986 (dependent variable: GOODENG)[a]

	Males			*Females*		
	OLS		*Logit*	*OLS*		*Logit*
	(i)	*(ii)*	*(iii)*	*(iv)*	*(v)*	*(vi)*
Constant	0.773 (30.63)	0.774 (31.31)	4.018 (11.34)	0.748 (26.04)	0.745 (26.23)	3.870 (9.99)
Education	0.024 (18.16)	0.022 (16.88)	0.204 (14.74)	0.030 (20.01)	0.029 (19.36)	0.250 (15.47)
Age	–0.005 (13.42)	–0.005 (12.40)	–0.057 (13.04)	–0.005 (12.85)	–0.005 (12.26)	–0.061 (13.07)
Small urban location	0.024 (2.55)	–0.002 (0.24)	–0.049 (0.39)	0.016 (1.66)	–0.001 (0.15)	–0.015 (0.11)
Rural Location	0.047 (4.18)	0.018 (1.68)	0.308 (2.08)	0.027 (2.45)	0.007 (0.63)	0.219 (1.39)
Married	–0.021 (2.10)	–0.018 (1.88)	–0.191 (1.86)	–0.013 (1.49)	–0.010 (1.21)	–0.212 (2.22)
Child <6	–0.004 (0.30)	–0.003 (0.23)	0.042 (0.28)	–0.015 (1.02)	–0.015 (0.99)	0.025 (0.15)
Child <6 and >5 and <15	–0.004 (0.27)	–0.004 (0.30)	–0.032 (0.25)	–0.005 (0.32)	–0.004 (0.28)	–0.005 (0.04)
Child >5 and <15	–0.014 (1.69)	–0.012 (1.39)	–0.061 (0.73)	–0.019 (2.22)	–0.017 (2.00)	–0.172 (1.90)
Duration of residence	0.009 (23.82)	0.009 (24.78)	0.111 (23.33)	0.008 (20.22)	0.008 (21.44)	0.098 (20.08)
Minority language concentration	[b]	–0.073 (14.97)	–0.364 (13.25)	[b]	–0.050 (9.42)	–0.269 (8.82)
Southern Europe	–0.550 (52.62)	–0.413 (29.34)	–5.672 (22.22)	–0.601 (53.15)	–0.501 (31.20)	–5.805 (22.85)
Northern Europe	–0.164 (13.00)	–0.141 (11.43)	–4.062 (14.89)	–0.200 (13.26)	–0.181 (12.21)	–4.259 (15.96)
Other Europe	–0.374 (20.78)	–0.360 (20.14)	–5.328 (20.58)	–0.466 (23.33)	–0.455 (22.83)	–5.610 (21.84)

(Continued Overleaf)

Table 9.2 Continued

	Males			*Females*		
	OLS		*Logit*	*OLS*		*Logit*
	(i)	*(ii)*	*(iii)*	*(iv)*	*(v)*	*(vi)*
Southwestern Asia	–0.528 (20.95)	–0.471 (18.47)	–5.614 (20.60)	–0.617 (23.97)	–0.573 (21.89)	–6.002 (21.34)
Vietnam	–0.815 (35.38)	–0.745 (32.37)	–7.288 (18.20)	–0.843 (35.60)	–0.791 (32.82)	–7.613 (17.66)
Other Asia	–0.314 (18.98)	–0.279 (17.31)	–4.639 (17.89)	–0.343 (21.14)	–0.319 (20.10)	–4.728 (18.76)
South America	–0.586 (13.44)	–0.543 (12.63)	–5.829 (17.32)	–0.667 (15.84)	–0.629 (14.64)	–6.162 (18.01)
North America	–0.050 (3.49)	–0.041 (2.95)	–1.909 (3.36)	–0.060 (4.56)	–0.054 (4.15)	–1.479 (1.96)
Africa	–0.148 (6.49)	–0.111 (4.99)	–3.604 (11.86)	–0.230 (9.24)	–0.205 (8.42)	–4.176 (14.41)
$\bar{R}^2$	0.4240	0.4492		0.4833	0.4936	
Log Likelihood			2641.7			2356.7
Sample size	8961	8961	8961	8114	8114	8114

Notes
a *t*-statistics in parentheses.
b Variable not entered.

Source: 1986 Census of Australia public use sample.

The Table 9.2 results are derived using a sample of all adult immigrants, compared with the sample of employed male immigrants used in Evans (1987) and Chiswick and Miller (1995).[10] The first three columns relate to males, and the final three columns to females. For each group the first specification omits the minority language concentration variable. The second and third specifications include this variable. Estimates presented in the first two columns are obtained using ordinary least squares (OLS), while those in the final column for each group are obtained from a logit model of dominant language fluency. In terms of the signs and statistical significance of the coefficients, the OLS and logit equations are nearly identical, so the discussion will focus on the OLS results.

When males and females are pooled (regression not shown) with a gender variable added to the equation, women are 2.7 percentage points less likely to be fluent than are men ($t = 5.41$).[11] There is, therefore, a small but statistically significant English language fluency advantage for males, other measured variables the same. A Chow test determined that the male and female samples should not be pooled: the test statistic was 5.47, which exceeded the critical

$F_{20,17035}$ = 1.57. Hence the presentations that follow are based on equations estimated separately for men and women.

Focusing first on Table 9.2, columns (i) and (iv), it is seen that most regressors are statistically significant determinants of dominant language fluency. According to the OLS results, dominant language fluency increases by 2.4 percentage points with each additional year of education for males, and by a larger amount (3.0 percentage points) for females. The gender difference of 0.6 percentage points is statistically significant (t = 3.00). The negative impact of age at arrival on dominant language fluency is 0.535 percentage points for males and 0.544 percentage points for females. Although these partial effects of age at arrival are highly significant, the gender difference is not statistically significant (t=0.16).

The change in language fluency with duration of residence does differ between males and females. There is a 0.921 percentage point improvement in English language fluency with each extra year of residence in Australia among males but only a 0.777 percentage point improvement for females, and the difference is statistically significant (t = 2.62). The steeper effect of duration on English language fluency for men than for women may be a consequence of greater exposure to English in the labor market, or may reflect greater incentives to invest in English language skills among those who expect a stronger labor market attachment.

There are some differences between the locality coefficients in the language models. Among males, there is a clear differentiation between the language skills of residents of the metropolitan area reference group and residents of both small urban (coefficient = 0.024, t=2.55) and rural (coefficient = 0.047, t = 4.18) areas. Among females, however, while both locality variables have positive coefficients, only that for residence of a rural area (0.027, t = 2.45) is significant. Thus, the English language skills among immigrants in Australia are better the lower the population density of the area.

The environmental/background variables included in the model tell a compelling story. Consider the birthplace variables, where United Kingdom and Ireland is the excluded benchmark category. The ranking of language proficiency differentials between immigrants from predominantly English-speaking countries and other immigrant groups are similar for men and women, and the differentials are only slightly wider for women. The birthplace variables can be viewed as falling into three clear categories. Among those from North America the negative effect is about 5 percentage points.[12] The birthplace groups with estimated negative partial effects of 15 to 23 percentage points are Northern Europe and Africa (primarily white South Africans). All other birthplace groups are associated with negative partial effects of 30 percentage points or more. The Vietnamese immigrants, nearly all of whom are refugees, have the lowest English language proficiency, even after controlling for their recent arrival, among other variables.[13]

The specifications presented in Table 9.2, columns (ii) and (v), attempt to capture influences on the language skills decision of the wider environment in which the individual lives. Thus, a variable for the minority language concentration of the area in which the individual resides is included in the estimating equation. This variable is measured as the percentage of the population in the region of residence (regardless of country of birth) that speaks the same non-English language as the respondent. (Additional details on the method of construction are contained in the Data Appendix.)

As outlined in Section II, it is expected that the presence of a large population speaking the same second language as the respondent will provide shelter against having to learn English. This expectation is borne out by the empirical results: the coefficient on the minority language concentration variable is negative, and highly statistically significant for both males and females. Moreover, the negative influence is stronger for males than for females. The estimated coefficient of –0.073 for males implies, for example, that a Greek male immigrant who settled in an area where five percent of the population spoke Greek would have an English language fluency rate 35 percentage points lower than a similar immigrant who settled in a region where he was the only Greek speaker. Among females, a five percent concentration of Greek speakers is associated with a 25 percentage point reduction in English language proficiency. While the five percent figure used in this illustration is extreme in terms of the concentration measures computed in this analysis, it would not be extreme at a less aggregated level.[14]

The greater sensitivity for men than for women of own language skills to minority language usage outside one's own home but in the region of residence may be a consequence of labor market factors. The greater the extent to which workers can find employment within a language enclave, the smaller will be their improvement in English. This effect is less relevant for women given a lower and more intermittent labor force participation.

The children's variables in Table 9.2 are not statistically significant, except for the significant negative effect of the presence of at least one child aged 5 to 15 years in the female equation. Thus, having school age children in the home appears to have a negative effect on female fluency but not on male fluency. This difference may arise because children reduced female labor supply, and language skills are either learned in the labor market or in anticipation of labor market work. It may also mean that the "children as interpreters" explanation is more compelling for women than for men.

In summary, the striking feature of this study of the 1986 Australian Census is the overall similarity of the processes determining dominant language fluency for males and females. However, interesting differences in the pattern by gender are found. These include the larger negative effects on fluency of minority language concentration for men and children for women, and the larger positive effects of duration for men. These findings suggest that English language skills are positively related to the extent of employment in the English-speaking labor market.

The decomposition technique outlined by Blinder (1973) may be used to assess the relative importance of the determinants of the linguistic differentials.

This technique provides a means of allocating the difference in mean fluency rates between the two groups under consideration to (i) differences between males and females in mean values of the regressors included in the model, and (ii) differences in estimated coefficients between the male and female equations. This decomposition can be computed using the formula:

$$\overline{\text{GOODENG}}_m - \overline{\text{GOODENG}}_f = \hat{\beta}_m(\bar{X}_m - \bar{X}_f) + (\hat{\beta}_m - \hat{\beta}_f)\bar{X}_f, \tag{1}$$

where ($\overline{\text{GOODENG}}_m - \overline{\text{GOODENG}}_f$) is the mean fluency rate differential of 2.9 percentage points in favor of males, $\hat{\beta}_i$ is the vector of estimated coefficients for group i ($i = m$ for males, $i = f$ for females), and $\bar{X}_i$ denotes the mean level of regressors for the ith group.[15] According to this decomposition, only six percent of the differential of 2.9 percentage points is attributable to differences in mean values of regressors, and 94 percent (i.e., 2.7 percentage points) is due to differences in the estimated coefficients. Hence, the male advantage in English language fluency is due mainly to differences in the estimated coefficients in the models of dominant language fluency, particularly the duration, minority concentration, and children variables.

The role of minority language concentration may be developed further by including a series of interaction terms between minority language concentration, and the variables age, education, and duration of residence.[16] The partial derivative of dominant language fluency (GOODENG) with respect to the minority language concentration variable (CONC) is given as:

$$\partial\text{GOODENG}/\partial\text{CONC} = -0.075 + 0.004{*}\text{Education} - 0.003{*}\text{Age} + 0.005{*}\text{PER}$$

(which equals –0.077 when evaluated at the means of Education, Age, and PER) for males and as:

$$\partial\text{GOODENG}/\partial\text{CONC} = -0.030 + 0.004{*}\text{Education} - 0.003{*}\text{Age} + 0.004{*}\text{PER}$$

(=–0.053 at mean values of variables) for females. Thus, for both groups, evaluated at means, there is a negative effect of concentration on language fluency and, as expected, the negative effect is larger for males than for females. However, minority language concentration has a less negative effect the higher the level of education (because of greater own human capital), and the longer the duration (which reflects the acquisition of skills regardless of where the individual lives), but the adverse effect is greater the older the age at immigration.

Whether the shelter from having to learn English provided by the presence of a sizable minority concentration is productive or counterproductive, however, cannot be determined from these data. For example, it could be argued that being able to converse in one's origin language is a useful step in the immigrant adjustment process, enabling the immigrant to become familiar with the infrastructure of the destination country; thus the significant positive sign on the CONC term in the partial derivative ∂GOODENG/∂PER = 0.007 + 0.005*CONC (= 0.010 at the mean of CONC) for males, and ∂GOODENG/∂PER = 0.006 + 0.004*CONC (= 0.009 at the mean) for females. However, the fact that ∂GOODENG/∂CONC is negative for both groups in the interactive model should not be overlooked when pursuing alternative explanations.

The extent to which the findings reported above can be the result of complex country-of-origin interactions is studied by separate analyses for each of the six large country groups with 200 or more observations (Table 9.3).[17] Part A of Table 9.3 refers to males, Part B to females. Part C analyzes the decomposition of the male/female English fluency differential. A striking feature of the results is that within each of the 12 equations education and duration have highly significant positive effects on English language fluency, while age (i.e., age at arrival) and minority language concentration have highly significant negative effects. This suggests a remarkable degree of robustness to the model.[18]

The gender difference in English language fluency rates in favor of males varies between four percentage points (Northern Europe, Other Asia) and 12 percentage points (Southwestern Asia). Among those born in Asia (both Southwestern Asia and Other Asia) differences in variables account for a sizable segment of the language fluency differential. For example, for immigrants from Southwestern Asia, the male/female differential in English fluency is 12.1 percentage points, and 5.8 percentage points of this is attributable to differences in the variables.[19] For all other immigrant groups differences in mean values of variables account for negligible portions of the fluency differential.[20] As in the aggregate analyses, for five of the six birthplaces, the major part of the small English language fluency differentials between male and female immigrants in Australia is due to differences in the estimated coefficients of the models of language fluency.

IV. English language skills and labor supply decisions

The analyses presented in Section III suggest that there are few differences between males and females in Australia in terms of overall levels of language fluency, but there are differences in the processes determining these levels of fluency, which relate to different expectations of labor supply. This section examines the relationship between female labor supply and English language fluency.

Table 9.4 presents summary information on eight major studies of female labor supply in Australia. It is apparent from this table that few Australian studies pay much attention to the overseas born, even though they account for 30 percent of the female labor force. Four of the eight most commonly cited studies of female labor supply in Australia (Haig and Wood 1976;

Table 9.3 Selected coefficients from models of dominant language fluency by birthplace group and gender, OLS, 1986 Australian Census[a]

	So. Eur.	*No. Eur.*	*Other Eur.*	*SW Asia*	*Other Asia*	*Africa*
A. *Males*						
Education	0.029	0.019	0.027	0.040	0.039	0.022
	(9.24)	(3.50)	(4.03)	(5.75)	(10.36)	(2.60)
Age	−0.014	−0.004	−0.009	−0.006	−0.002	−0.007
	(12.00)	(2.98)	(4.77)	(1.98)	(1.01)	(2.64)
Duration of residence	0.023	0.008	0.020	0.019	0.014	0.006
	(19.92)	(4.28)	(10.45)	(5.93)	(8.94)	(2.63)
Minority language concentration	−0.061	−0.259	−0.114	−0.041	−0.176	−0.071
	(11.20)	(7.54)	(2.19)	(1.44)	(7.11)	(2.47)
$\bar{R}^2$	.2697	.2049	.1918	.1563	.2754	.0816
Sample size	2262	687	648	345	703	225
GOODENG_m (mean)	.4202	.8719	.6204	.4087	.6700	.8431
B. *Females*						
Education	0.040	0.029	0.032	0.024	0.035	0.024
	(13.19)	(4.79)	(4.05)	(3.49)	(9.38)	(2.11)
Age	−0.013	−0.005	−0.018	−0.011	−0.003	−0.010
	(9.56)	(2.70)	(8.57)	(3.30)	(2.32)	(3.67)
Duration of residence	0.017	0.007	0.021	0.013	0.014	0.007
	(13.42)	(3.04)	(11.13)	(4.15)	(9.82)	(3.25)
Minority language concentration	−0.034	−0.253	−0.055	−0.015	−0.194	−0.089
	(5.41)	(4.91)	(1.50)	(0.48)	(7.66)	(2.97)
$\bar{R}^2$	.2588	.1721	.2106	.1247	.2889	.1635
Sample size	1846	601	557	288	776	252
GOODENG_f (mean)	.3386	.8286	.5260	.2882	.6276	.7817
C. *Decomposition*						
$\overline{\text{GOODENG}}_m - \overline{\text{GOODENG}}_f$	.0819	.0433	.0943	.1205	.0424	.0614
Due to Regressors $\hat{\beta}_m(\bar{X}_m - \bar{X}_f)$	.0279	.0207	.0165	.0584	.0630	−.0034
Coefficients[b] $(\hat{\beta}_m - \hat{\beta}_f)\bar{X}_f$	.0539	.0226	.0778	.0621	−.0206	.0648

Notes

a Estimating equations also include variables for location, marital status, and children; *t*-statistics in parentheses.

b Components may not sum to total due to rounding.

Source: 1986 Census of Australia, Public Use Sample.

Eyland et al. 1982; Gregory et al. 1985; Ross 1986) do not examine the relative position of the foreign born. In Miller and Volker (1983) and Miller (1985) the foreign born are treated as a homogeneous group, and thus distinguished from other groups by the use of a single dichotomous variable. Both studies find that the foreign born participate in the labor market to a higher degree than the native born. Evans (1984) and Brooks and Volker

Table 9.4 Summary of major studies of female labor supply in Australia

	Haig and Wood (1976)	*Eyland, Mason and Lapsley (1982)*	*Miller and Volker (1983)*	*Evans (1984)*
Decision analyzed, i.e., dependent variable	Participation (numbers)	Participation, hours of work	Participation	Participation within birthplace groups
Type of data	Grouped time-series	Individual cross-section	Grouped cross-section	Individual cross-section
Source of data	ABS Labor Force Survey 1962(1) to 1972(2)	1977 Survey of women in suburb of Eastwood	1976 Census	1981 Census
Overseas born	Overseas born not discussed	Eastwood primarily of Anglo-Saxon origin	% of area overseas born	Detailed analysis of overseas born
Other independent variables	Population industry structure, unemployment, rate, seasonal dummies	Education, age youngest child, marital status, age	Education, location, % with > 1 children, husband's income, unempl. rate, own wage, birthplace	Education, location, no. children, marital status, English skill, years since migration, age
Major findings	Increased participation due to higher demand for labor	Presence and age of children main determinants of labor supply	Overseas born has a large positive effect	English skill not signif. for any birthplace group

	Gregory, McMahon and Whittingham (1985)	*Brooks and Volker (1985)*	*Miller (1985)*	*Ross (1986)*
Decision analyzed, i.e., dependent variable	Participation	Participation, hours worked	Hours of work	Participation, hours of work
Type of Data	Grouped time-series	Individual cross-section	Individual cross-section	Individual cross-section
Source of data	ABS Labor Force Survey 1966(4) to 1980(4)	1981 Census	1973 Social Mobility in Australia survey	1980 Sydney Survey of Market Work Patterns of Married Women
Overseas born	Overseas born not discussed	Four overseas birthplace dummy variables	Overseas born dummy variable	Overseas born not discussed
Other independent variables	Divorce rate, unempl. rate, children <5 per woman, time trend, seasonal dummies, wages	Education, location, age youngest child, no. of dependents, husband's income, marital status, birthplace, language, duration of residence, age	Number/age of children, husband's income, own wage, birthplace	Education, experience, number/age of children, husband's wage, other income, age
Major findings	Aggregate time-series supply equations unsatisfactory	Birthplace, language skills significant, children and husband's income important	Foreign-born have higher participation	Specification of children variables important

Source: Adapted from Johnson (1988).

(1985) provide a more detailed analysis of the foreign born, but use different methods of analysis and arrive at different conclusions. Brooks and Volker include dummy variables for four foreign birthplace regions in their probability models (the benchmark group is the native born) and report that three of the four coefficients (for Anglo-Saxon, Southern Europe, and other birthplace groups) are positive and statistically significant. Only the Northern Europe birthplace variable was negative, though it was statistically insignificant. Evans (1984), however, conducts separate analyses for a number of major birthplace regions. Regression-adjusted labor force participation rates were computed. Brooks and Volker and Evans differ somewhat in their ranking of countries of origin in terms of predicted female labor force participation rates.

Only two studies (Evans 1984; Brooks and Volker 1985) examine whether language skills affect participation decisions, and their conclusions also contrast sharply.[21] Brooks and Volker find that language skills exercise an important influence on female labor supply decisions, whereas Evans reports that "English skill has no significant effect on the labor force participation of any group" (1984, p. 1076).

The study by Evans also draws attention to differences in coefficients in labor supply response across the various birthplace groups. For example, it is reported that family formation has little effect on the labor force involvement of Mediterranean and Eastern European immigrants (1984, p. 1074).[22] Evans also finds that marital status does not affect the labor supply of female immigrants from Eastern European and Mediterranean countries, and educational attainment was not a statistically significant determinant of the labor force participation of immigrants from the Mediterranean countries or from Third World countries.

The model of female labor force participation adopted here is a standard reduced-form equation. The participation decision is related to the respondent's age, educational attainment, location of current residence, marital status, presence and age structure of children, husband's income, birthplace, duration of residence in Australia, citizenship, and English language proficiency.

Table 9.5 presents estimates of models of labor force participation for three groups. The first three columns are for the sample of foreign-born females aged 25–64 years. The second set of two columns lists results for the foreign born from non-English-speaking countries (58 percent of the sample), while the third set lists results for the native born. In each instance estimates for both an OLS model and a logit model are presented. The comparable OLS and logit equations are open to similar interpretations and only the OLS estimates will be discussed. Furthermore, the estimates for immigrants from non-English-speaking countries are virtually identical to the results for the overseas born as a whole, so only the latter results are discussed in the text.

Seven dummy variables are entered in the estimating equation for the broad five-year age groups available in the Census file (Table 9.5, column 1). The 40–44-year age bracket is the reference group. There are no statistically

Table 9.5 Estimates of models of labor force participation, females aged 25–64 years, 1986 Australia Census[a]

	Total overseas born			*Non-English speaking*		*Native born*	
	(1)*OLS*	(2)*OLS*	(3)*LOGIT*	(4)*OLS*	(5)*LOGIT*	(6)*OLS*	(7)*LOGIT*
Constant	0.498	0.468	–0.335	0.446	–0.389	0.271	–1.377
	(16.46)	(14.35)	(1.98)	(11.98)	(2.08)	(13.37)	(12.43)
Age (40–44)							
25–29	–0.006	–0.009	–0.016	0.015	0.086	0.043	0.272
	(0.31)	(0.43)	(0.15)	(0.52)	(0.59)	(3.40)	(4.19)
30–34	0.019	0.018	0.104	0.029	0.148	0.038	0.212
	(0.91)	(0.86)	(1.02)	(1.05)	(1.10)	(2.99)	(3.49)
35–39	0.028	0.027	0.132	0.009	0.038	0.034	0.174
	(1.56)	(1.50)	(1.47)	(0.38)	(0.33)	(2.77)	(2.98)
45–49	–0.100	–0.098	–0.477	–0.099	–0.458	–0.062	–0.319
	(4.95)	(4.85)	(4.96)	(3.68)	(3.69)	(4.42)	(4.87)
50–54	–0.246	–0.243	–1.114	–0.228	–1.014	–0.215	–1.009
	(10.84)	(10.66)	(10.51)	(7.73)	(7.45)	(14.28)	(14.64)
55–59	–0.406	–0.404	–1.844	–0.379	–1.728	–0.400	–1.830
	(17.50)	(17.38)	(15.91)	(12.70)	(11.60)	(26.67)	(25.06)
60–64	–0.567	–0.563	–2.906	–0.515	–2.712	–0.558	–2.789
	(26.14)	(25.83)	(20.35)	(17.87)	(14.50)	(39.36)	(32.24)
Education	0.025	0.024	0.130	0.020	0.103	0.044	0.239
	(13.04)	(12.22)	(11.61)	(8.54)	(8.15)	(30.61)	(26.91)
Location (metropolitan area)							
Small	–0.047	–0.048	–0.243	–0.018	–0.091	–0.037	–0.188
Urban	(3.07)	(3.10)	(3.11)	(0.75)	(0.75)	(4.99)	(5.11)
Rural	–0.043	–0.044	–0.216	–0.014	–0.063	0.025	0.112
	(2.26)	(2.32)	(2.40)	(0.48)	(0.46)	(2.74)	(2.53)
Marital status (single)							
Married	–0.060	–0.058	–0.309	–0.095	–0.511	–0.006	0.017
	(3.47)	(3.35)	(3.23)	(4.39)	(4.20)	(0.51)	(0.29)
Partner is unemployed	–0.062	–0.060	–0.273	–0.016	–0.050	–0.129	–0.605
	(2.20)	(2.14)	(2.08)	(0.46)	(0.31)	(6.05)	(5.95)
Age youngest child (no children)							
<2	–0.442	–0.441	–2.077	–0.381	–1.765	–0.491	–2.399
	(12.73)	(12.69)	(12.26)	(8.16)	(7.98)	(25.72)	(24.73)
2–5	–0.245	–0.245	–1.193	–0.205	–0.966	–0.333	–1.648
	(9.73)	(9.73)	(9.89)	(6.15)	(6.18)	(22.11)	(22.25)
5–15	–0.060	–0.059	–0.329	–0.019	–0.112	–0.115	–0.638
	(2.91)	(2.87)	(3.27)	(0.72)	(0.86)	(8.85)	(10.10)
>15	0.031	0.033	0.143	0.074	0.341	0.034	0.108
	(1.38)	(1.48)	(1.27)	(2.53)	(2.42)	(2.12)	(1.38)
Number of dependents	–0.029	–0.028	–0.137	–0.029	–0.137	–0.019	–0.084
	(3.61)	(3.59)	(3.65)	(2.89)	(2.90)	(3.95)	(3.79)
Husband's Income (<$9,000)							
9,000–15,000	0.132	0.132	0.689	0.180	0.925	0.028	0.145
	(6.79)	(6.76)	(7.03)	(7.59)	(7.79)	(2.20)	(2.21)
15,000–22,000	0.099	0.098	0.507	0.127	0.659	0.012	0.068
	(5.51)	(5.44)	(5.55)	(5.57)	(5.77)	(1.06)	(1.14)

(*Continued Overleaf*)

Table 9.5 Continued

	Total overseas born			Non-English speaking		Native born	
	(1)*OLS*	(2)*OLS*	(3)*LOGIT*	(4)*OLS*	(5)*LOGIT*	(6)*OLS*	(7)*LOGIT*
22,000–32,000	0.072	0.068	0.353	0.108	0.553	–0.013	–0.059
	(3.65)	(3.44)	(3.51)	(4.12)	(4.22)	(1.03)	(0.94)
>32,000	0.054	0.050	0.256	0.075	0.391	–0.018	–0.096
	(2.28)	(2.10)	(2.20)	(2.29)	(2.52)	(1.24)	(1.35)
Birthplace							
South Europe	–0.057	–0.033	–0.151	[b]	[b]	[b]	[b]
	(3.75)	(1.90)	(1.71)				
North Europe	–0.060	–0.052	–0.259	–0.014	–0.070	[b]	[b]
	(2.93)	(2.52)	(2.50)	(0.60)	(0.61)		
Other Europe	–0.045	–0.027	–0.138	0.021	0.113	[b]	[b]
	(2.15)	(1.20)	(1.19)	(0.91)	(0.94)		
South West Asia	–0.109	–0.086	–0.414	–0.049	–0.231	[b]	[b]
	(3.65)	(2.75)	(2.66)	(1.58)	(1.53)		
Other Asia	–0.006	0.007	0.020	0.068	0.322	[b]	[b]
	(0.34)	(0.35)	(0.19)	(2.91)	(2.74)		
Vietnam	0.083	0.115	0.609	0.171	0.857	[b]	[b]
	(1.88)	(2.47)	(2.61)	(3.66)	(3.70)		
South America	–0.098	–0.072	–0.351	–0.011	–0.059	[b]	[b]
	(2.07)	(1.48)	(1.50)	(0.21)	(0.25)		
North America	0.077	0.080	0.427	0.147	0.746	[b]	[b]
	(2.06)	(2.13)	(2.02)	(3.60)	(3.29)		
Africa	–0.047	–0.038	–0.197	0.021	0.092	[b]	[b]
	(1.59)	(1.27)	(1.30)	(0.65)	(0.57)		
Oceania	0.032	0.038	0.204	[b]	[b]	[b]	[b]
	(0.64)	(0.77)	(0.78)				
New Zealand	0.074	0.073	0.384	[b]	[b]	[b]	[b]
	(3.31)	(3.26)	(2.98)				
Duration of residence (20+ years)							
0–4	–0.063	–0.053	–0.294	–0.078	–0.406	[b]	[b]
	(3.11)	(2.60)	(2.78)	(2.63)	(2.75)		
5–9	0.019	0.026	0.123	–0.015	–0.091	[b]	[b]
	(1.06)	(1.43)	(1.27)	(0.58)	(0.69)		
10–14	0.021	0.025	0.122	–0.006	–0.037	[b]	[b]
	(1.18)	(1.42)	(1.35)	(0.23)	(0.29)		
15–19	–0.008	–0.005	–0.024	–0.021	–0.103	[b]	[b]
	(0.57)	(0.35)	(0.34)	(1.04)	(1.03)		
Citizen	0.024	0.024	0.124	0.033	0.166	[b]	[b]
	(2.12)	(2.09)	(2.13)	(1.88)	(1.96)		
GOODENG	[b]	0.039	0.194	0.040	0.186	[b]	[b]
		(2.55)	(2.51)	(2.46)	(2.31)		
$\bar{R}_2$	0.2037	0.2043		0.1911		0.1969	
$L - L^c$			–4663.2		–2738.3		–12350.0
Sample size	8114	8114	8114	4692	4692	21231	21231

Notes
a *t*-statistics in parentheses.
b Variable not entered.
c Log likelihood function.

Source: 1986 Census of Australia, public use sample significant differences between the participation rates of individuals 40–45 years of age and those younger, other variables the same. As these age brackets represent the child-bearing and -rearing years, the lack of statistically significant age effects is reassuring in the sense that it indicates that fertility considerations among the foreign born are being adequately modeled through the variables for age of the youngest child and total number of dependents. At ages greater than 45 years, however, there is a steady decline in the attachment to the labor force, by 10 percentage points among 45–49-year-olds, by 25 percentage points among 50–54-year-olds, 41 percentage points among 55–59-year-olds, and a massive 57 percentage points for 60–64-year-olds.[23] In other words, the propensity to choose leisure or home production (perhaps because of fewer investments relevant for the labor market when they were young) rather than working rises rapidly with age in these cohorts.

There is a highly significant, positive relationship between labor force participation and educational attainment. The estimated coefficient of 0.025 indicates, for example, that the participation rate, ceteris paribus, of a female possessing a university degree would be about 13 percentage points higher than for a female who left school following the completion of year 10. While this effect is sizable in absolute terms, it is quite minor relative to the effect of age among the 40 years and over group.

Labor force participation differs across locations, with residents of both small urban and rural regions having participation rates about four percentage points lower, ceteris paribus, than residents of the metropolitan area control group. These estimated effects presumably reflect differences by degree of urbanization and population density in demand-side considerations (job vacancies, wage offers) in Australia.[24]

The pattern of effects across the family-based variables accords with expectations. The negative effect of a child on labor force participation becomes less intense as the age of the youngest child rises. The presence of a young child (<2 years of age) reduces the participation rate by 44 percentage points, presumably because the children are extremely intensive in mother's time. If the youngest child is 2–5-years-of-age, participation is reduced by 25 percentage points, and if the youngest child is 6–15 years then participation is reduced by only six percentage points. Where the youngest child is 16 years or older (compared to those with no children), there is a positive, but not always significant impact on labor force participation. Thus, in spite of the lower past labor market experience of women with grown children, compared to similarly situated women who never had children the former have a higher participation rate. This may reflect the greater "goods-intensity" of adult children. This suggests an optimal allocation of mothers' time between home production and the labor market, depending on whether their children are time- or goods-intensive (see Chiswick 1986).

Added to these effects is the negative effect operating through the "total number of dependents" variable, a proxy measure for consumption per capita when husband's income is held constant. This effect, however, is relatively minor in importance, indicating a reduction in labor force participation of 2.9 percentage points for each additional dependent child. Thus, the major impact on labor force participation appears to be determined by the age of the youngest child, indicating considerable merit to the modeling approach employed by Brooks and Volker (1985).

There is an inverted U-shaped relationship between female participation and husband's income. Participation rates rise from the lowest income category (<$9,000) to the $9,000–$15,000 group. Thereafter, however, participation rates decline as the income of the husband rises.[25]

There is also considerable variation in labor supply across marital states, and this effect is not neutral with respect to the unemployment status of the spouse. Marriage to an employed man is associated with a participation rate 6 percentage points lower than that for single persons, other things the same. However, marriage to an unemployed man is associated with a participation rate 12 percentage points (= –0.060–0.062) lower than that for single persons, ceteris paribus. This suggests that in terms of female labor supply decisions, the discouraged-worker effect (i.e., employment opportunities in the area) is of greater numerical importance than the added-worker effect (i.e., current income status of the family).

The effect of birthplace on participation varies sharply even when other variables are the same (Table 9.5, column 1). The high participation group is from Vietnam, North America, and New Zealand, while the low participation group is from Southern Europe, Northern Europe, Other Europe, Southwestern Asia, and South America. Immigrants from Other Asia, United Kingdom, Africa, and Oceania comprise the medium participation group. Generally, the estimated birthplace coefficients are quite small, though the range of 19 percentage points between the participation rates of the groups with the greatest (Vietnamese) and least (South America and Southwestern Asia) attachment to the labor force is still quite impressive.

Participation rates also vary with duration of residence in Australia, with recent arrivals having participation rates 6 percentage points lower than other groups. As the other Australian duration variables have small coefficients, and are not statistically significant, it can be concluded that there is a rapid adjustment to the circumstances of the labor market of the destination country.[26]

The variable for English language fluency is included in the equation presented in Table 9.5, column 2. Inclusion of this variable in the estimating equation is associated with only minor changes in most coefficients, although it should be noted that the coefficients for South America, Southern Europe, Southwestern Asia, and Vietnam rise by 2–3 percentage points and the coefficient for South America becomes statistically insignificant. Immigrants possessing English language fluency have participation rates 3.9 percentage points higher than other groups (t-ratio of 2.6). This effect is the equivalent of that of 1.5 years of schooling. Thus, language skills are associated with a greater attachment to the labor market, and the effect is statistically significant.

Finally, results from equations estimated for the native born are presented in Table 9.5, columns 6 and 7. The distinguishing features of these results are (a) the presence of statistically significant (though numerically small) age effects among 25–39-year-olds, (b) the larger impact of educational attainment on the participation decision (coefficient equal to 0.044 compared to 0.025 for the total overseas born and 0.020 for immigrants from non-English-speaking countries), and (c) the larger effect of the age of the youngest child on the participation decision.[27] This last difference is most apparent in the case of preschool children. Among the native born, for example, the effect of children aged 2–5 is –0.333, compared to –0.245 for the sample of all immigrants and –0.205 for the sample of immigrants from non-English-speaking countries.

Participation rate equations were estimated (see Table 9.6) for each of the six birthplace group where the sample size was at least 200. The effects of age and marital status are fairly uniform across the six birthplace groups. The effects on participation of most other variables, however, differ appreciably. Education, for example, has a sizable impact (of about 3 percentage points) on participation among immigrants from Africa, Northern Europe, and Other Europe, and a smaller impact (1–2 percentage points) for immigrants from Southern Europe, Other Asia, and Southwestern Asia. The minor impact of education on participation among immigrants from Asia is consistent with the findings reported in Evans (1984), but that relating to Southern Europe is not. This may reflect the richer specification of the participation equation in this study; the coefficients of determination in Tables 9.5 and 9.6 are typically twice the magnitude of those reported by Evans (1984). Duration of residence has an impact on participation decisions only among the Asian-born women, particularly immigrants from Southwestern Asia. There is a lack of statistical significance of the duration of residence variables for the other birthplace groups.

The number of dependents has a statistically significant negative impact on the participation decision only among immigrants from Southern Europe and Other Asia. However, the variables recording the age of the youngest child are important to the participation decision of immigrants from all countries other than Other Europe and Southwestern Asia. Where the youngest child is of preschool age, participation is reduced, by between 19 percentage points for immigrants from Southern Europe or Other Asia and 58 percentage points for immigrants from Africa with a child under age 2. Where the youngest child is 5–15 years of age, however, there is no statistically significant participation rate effect. And where the youngest child is older than 15 years, there tends to be a positive effect on participation, particularly in the case of immigrants from Southern Europe, Other Europe, and Africa. The latter finding for the Southern Europeans is consistent with the finding in the educational attainment literature that Southern European immigrants attach greater weight to the educational achievements of their children than other immigrant groups (see, for example, Williams et al. 1987).

Table 9.6 Estimates of models of labor force participation, by birthplace group, females aged 25–64 years, 1986 Australian Census

	South Europe	*North Europe*	*Other Europe*	*South West Asia*	*Other Asia*	*Africa*
Constant	0.365 (6.20)	0.382 (2.99)	0.198 (1.67)	0.889 (6.26)	0.591 (6.11)	0.462 (2.02)
Age (40–44)						
25–29	0.050 (0.95)	–0.025 (0.22)	0.039 (0.40)	–0.049 (0.45)	0.025 (0.41)	0.130 (1.12)
30–34	–0.004 (0.07)	–0.036 (0.50)	0.059 (0.67)	–0.059 (0.54)	0.075 (1.25)	0.182 (1.63)
35–39	–0.004 (0.10)	0.046 (0.76)	–0.045 (0.54)	–0.003 (0.03)	0.053 (0.92)	0.001 (0.00)
45–49	–0.115 (2.88)	–0.047 (0.62)	–0.022 (0.25)	–0.094 (0.81)	0.013 (0.18)	–0.210 (1.74)
50–54	–0.212 (4.83)	–0.330 (4.11)	–0.008 (0.09)	–0.334 (2.82)	–0.184 (2.30)	–0.383 (3.18)
55–59	–0.345 (7.63)	–0.398 (5.34)	–0.372 (4.34)	–0.571 (4.79)	–0.288 (3.33)	–0.302 (1.97)
60–64	–0.435 (9.42)	–0.563 (7.66)	–0.471 (5.95)	–0.711 (7.08)	–0.545 (7.28)	–0.525 (3.84)
Education	0.020 (5.53)	0.028 (3.10)	0.039 (5.83)	0.012 (1.76)	0.009 (1.87)	0.026 (1.87)
Location (metropolitan area)						
Small urban	–0.076 (1.69)	–0.051 (1.06)	0.069 (1.14)	0.134 (0.62)	0.063 (1.08)	0.129 (1.40)
Rural	0.116 (2.31)	–0.118 (2.27)	–0.121 (1.38)	0.060 (0.48)	–0.152 (1.71)	0.126 (1.14)
Marital status (single)						
Married	–0.065 (1.83)	–0.145 (2.53)	–0.151 (3.22)	–0.140 (1.58)	–0.127 (2.16)	–0.241 (2.26)
Partner is unemployed	–0.046 (0.69)	–0.014 (0.15)	0.049 (0.49)	–0.137 (1.63)	0.130 (1.43)	0.167 (0.94)
Age youngest child (no children)						
<2	–0.337 (3.33)	–0.535 (3.66)	–0.269 (1.58)	–0.148 (1.00)	–0.418 (5.00)	–0.579 (3.29)
2–5	–0.172 (2.59)	–0.287 (2.71)	–0.167 (1.51)	–0.241 (2.06)	–0.162 (2.40)	–0.270 (2.13)
5–15	0.013 (0.27)	–0.081 (0.98)	–0.013 (0.14)	–0.071 (0.72)	0.048 (0.73)	–0.026 (0.23)
>15	0.093 (2.10)	0.036 (0.41)	0.181 (1.86)	0.316 (2.93)	–0.036 (0.40)	0.198 (1.71)
Number of dependents	–0.036 (1.80)	0.015 (0.47)	–0.034 (0.91)	–0.021 (0.91)	–0.050 (2.13)	–0.032 (0.74)

Husband's income (<$9,000)						
9,000–15,000	0.226	0.301	0.223	0.049	0.129	0.123
	(6.55)	(4.34)	(3.18)	(0.58)	(1.90)	(1.14)
15,000–22,000	0.161	0.182	0.129	–0.019	0.084	0.196
	(4.79)	(2.80)	(2.24)	(0.21)	(1.33)	(1.89)
22,000–32,000	0.013	0.107	0.163	0.150	0.127	0.158
	(2.34)	(1.55)	(2.36)	(1.37)	(1.97)	(1.47)
>32,000	0.103	0.119	0.158	–0.104	0.060	0.141
	(1.40)	(1.40)	(1.66)	(0.59)	(0.84)	(1.26)
Duration of residence (20+ years)						
0–4	0.036	–0.098	0.013	–0.280	–0.101	0.109
	(0.41)	(0.97)	(0.16)	(2.24)	(1.71)	(0.20)
5–9	–0.061	–0.090	–0.014	–0.125	0.043	0.079
	(1.01)	(0.94)	(0.20)	(1.43)	(0.80)	(0.87)
10–14	–0.053	–0.121	–0.037	–0.111	0.054	0.049
	(1.03)	(1.10)	(0.42)	(1.29)	(0.99)	(0.57)
15–19	–0.025	0.041	0.007	–0.132	0.059	–0.141
	(0.87)	(0.51)	(0.10)	(1.73)	(1.03)	(1.86)
Citizen	0.046	–0.051	0.049	–0.173	0.080	0.096
	(1.71)	(1.05)	(0.89)	(2.36)	(1.86)	(1.18)
	0.039	0.112	0.021	0.058	0.021	–0.058
GOODENG	(1.45)	(2.38)	(0.52)	(0.90)	(0.56)	(0.78)
R^2	0.1591	0.2167	0.2762	0.1685	0.1927	0.2013
Sample size	1846	601	557	288	776	252

Note
a *t*-statistics in parentheses.

Source: 1986 Census of Australia, public use sample.

Fluency in the dominant language typically has a positive impact on participation, but it is significant only in the case of immigrants from Northern Europe (Table 9.6). This provides a reconciliation of the Evans (1984) and Brooks and Volker (1985) results. On the basis of equations estimated for separate birthplace groups as in Table 9.6, Evans (1984) finds that English language proficiency does not affect the participation decision of immigrant women. Brooks and Volker (1985), however, on the basis of a participation rate equation estimated for all foreign-born women as in Table 9.5, report that dominant language proficiency does affect the participation decisions of immigrant women. It would appear, therefore, that the effect of dominant language fluency on participation is sufficiently small that, while significant for immigrant women as a whole, it is not significant when the data are disaggregated by birthplace categories and the sample sizes are reduced.

Thus, it appears that English language deficiency has a small retarding effect but does not represent a major obstacle to labor force participation among immigrant women in Australia. It is perhaps for this reason that groups with different degrees of attachment to the paid labor force, such as males and females, exhibit similar levels of dominant language fluency.

V. Conclusion

According to the 1986 Australian Census of Population and Housing, 74 percent of male immigrants and 71 percent of female immigrants are fluent in English. The decision of immigrants to acquire English-speaking skills was modeled in this study as a function of personal characteristics and environmental factors. It was hypothesized that there would be a positive relationship between English language fluency and both educational attainment and period of residence, and a negative relationship with age at arrival, foreign marriage, and the proportion of the population within the region of residence speaking the same nondominant language as the individual. The relationship between children and language fluency was argued to be a priori ambiguous. Furthermore, it was argued that because of gender differences in labor force activity the negative effect of minority language concentration would be larger in absolute value for men than for women, that the positive effect of duration of residence on fluency would be larger for men than for women, and the effect of children is more positive or less negative for men than for women.

The empirical results were consistent with these hypotheses. Thus, for both males and females, English language fluency increases by 2–3 percentage points for each additional year of education and declines by 0.5 percentage points with an older age at arrival. There is a 0.92 percentage point improvement in English language fluency with each extra year of residence in Australia among male immigrants but only a 0.78 percentage point improvement for females. The difference is statistically significant. Having children in the family has a negative effect on female fluency, but not on male fluency. Finally, the minority language concentration of the area in which the individual resides is an important determinant of language fluency: the greater the minority language concentration the lower the rate of English fluency, and the diminution effect is larger for men than for women.

In summary, the differences in the processes (partial effects of explanatory variables) determining male and female English language fluency among immigrants in Australia appear to relate directly to anticipated labor supply. Moreover, English language fluency among immigrant women is associated with a four percentage point increase in the female labor force participation rate. English language fluency rates, however, are roughly similar for men and women in spite of the large gender difference in labor force participation rates.

Appendix: Description of variables

The analyses of the 1986 Australian Census presented in the paper are based on the 1986 Household Sample File (Section of State). This is a one-in-a-hundred sample of the population (see Table 9A.1).

Definition of population: Foreign-born men and women aged 25–64. The

Table 9A.1 Means and standard deviations of selected variables, 1986 Australian Census

Variable	*Immigrants*				*Native-Born*	
	Males		*Females*		*Females*	
	Mean	*Standard deviation*	*Mean*	*Standard deviation*	*Mean*	*Standard deviation*
GOODENG	0.739	0.439	0.710	0.454	*a*	*a*
Education	11.029	2.979	10.625	2.876	10.955	2.093
Age	43.820	10.621	42.669	10.761	41.428	11.311
Duration of residence	21.229	10.133	20.583	10.435	*a*	*a*
Small urban location	0.124	0.329	0.124	0.329	0.254	0.435
Rural location	0.085	0.278	0.088	0.284	0.165	0.371
Married	0.815	0.388	0.784	0.412	0.741	0.438
Southern Europe	0.252	0.434	0.228	0.419	*a*	*a*
Northern Europe	0.077	0.266	0.074	0.262	*a*	*a*
Other Europe	0.072	0.259	0.069	0.253	*a*	*a*
Western Asia	0.039	0.192	0.035	0.185	*a*	*a*
Other Asia	0.078	0.269	0.096	0.293	*a*	*a*
Vietnam	0.017	0.127	0.017	0.128	*a*	*a*
South America	0.011	0.103	0.012	0.109	*a*	*a*
Other America	0.018	0.131	0.017	0.130	*a*	*a*
Africa	0.028	0.166	0.031	0.173	*a*	*a*
Oceania	0.009	0.095	0.010	0.100	*a*	*a*
New Zealand	0.042	0.201	0.050	0.217	*a*	*a*
Minority language concentration	0.672	1.273	0.676	1.270	*a*	*a*
Sample size	8,961		8,114		21,231	

Note
a Variable not relevant.

Source: 1986 Census of Australia, public use sample.

analysis is restricted to individuals living in private dwellings and who were members of the primary family in such dwellings (i.e., all single-family private dwellings and the primary family in multifamily households).

Years of education (EDUC): This variable records the total years of full-time education. It has been created from the census Age Left School and Qualifications variables. Years of education is calculated as Age Left School—5 for those without tertiary qualifications (diplomas, degrees, graduate diplomas, higher degrees). Individuals who stated a school-leaving age of 19 or more years and who did not possess a tertiary qualification were assigned 13 years of education. Individuals who possess a diploma have been assumed to have the equivalent of 15 years of full-time education, individuals who possess a bachelor degree the equivalent of 16 years of full-time education, individuals who possess a graduate diploma the equivalent of 17 years of full-time education, and individuals who have a higher degree (masters, Ph.D) have been coded as having 19 years of education.

Years since migration: The 1986 Census years since migration data are released in categorical form, and only five broad categories are distinguished: 0–4 years, 5–9 years, 10–14 years, 15–19 years, and 20 years and over. A "continuous" duration of residence variable is created using the midpoints of the closed intervals, and a value of 30 years for the open-ended upper interval. The continuous measure is used in the language models while dummy variables for the categories in the original data are used in the models of labor force participation.

Birthplace: Eleven broad birthplace regions are recognized in the study. They are United Kingdom and Ireland, Southern Europe (Albania, Greece, Italy, Malta, Portugal, Spain, Yugoslavia), Northern Europe (The Netherlands, Germany), Other Europe (all other countries of Europe, including USSR), Western Asia (Bahrain, Jordan, Kuwait, Oman, Qatar, Saudi Arabia, Yemen, Cyprus, Iraq, Israel, Lebanon, Syria, Turkey), Vietnam, Other Asia, South America (Argentina, Brazil, Chile, Columbia, Ecuador, Peru, Uruguay, Other South America), North America (Canada, Caribbean, El Salvador, Mexico, United States, Other American), New Zealand, Other Oceania.

Citizenship: The citizenship variable is set to one for immigrants who are citizens of Australia and to zero for all other immigrants.

English language proficiency (GOODENG): GOODENG is set to one for individuals who speak only English at home, or if a language other than or in addition to English is spoken in the home, speak English "very well." The GOODENG variable is set to zero where a language other than English is spoken in the home and the respondent speaks English either "well," "not well," or "not at all."

Marital status (MARRIED): This is a binary variable, defined to equal one for individuals who are married (spouse present), and defined to equal zero for all other marital states. Information on whether the individual is married is derived from the census question on marital status. Information on whether the spouse is present is derived from the family structure of the public use sample file.

Husband unemployed: This variable is defined only for the married and records whether the husband is unemployed.

Husband's income: Husband's income is an annual amount inferred from the usual weekly income. It includes wage, salary, self-employment income, and unearned income (interest, rent, government transfers). Five categories are distinguished for income: <\$9,000, \$9,000–\$15,000, \$15,000–\$22,000, \$22,000–\$32,000, and >\$32,000. The lowest-income group (<\$9,000) is used as the benchmark group. This variable is defined only for the married and in this sense is equivalent to interaction terms between husband's income and marital status.

Children: In the model of dominant language fluency, three variables were constructed from the family structure of the public use sample file to parallel the variables included in analyses for the U.S. census. The first of these records whether one or more children aged less than six years were living in the family and there were no older children. The second records whether one or more children aged between six and 17 years inclusive were living in the family, and there were no younger children. The third variable records the presence of children aged less than 6 years and between six and 17 years. For the labor supply models, dichotomous variables are included in the estimating equation for the age of the youngest child (<2, 2–5, 5–15, <15) and a continuous variable is included for the number of dependents. This specification is based on Brooks and Volker (1985).

Location: The only geographic information contained in the Household Sample File distinguishes individuals living in major urban areas, other urban areas, and rural areas. On this basis, two dichotomous variables were formed, the first for individuals living in "other urban areas" and the second for individuals living in "rural areas." Individuals residing in major urban areas comprise the control group. No data are available on state of residence.

Age: The age data are presented in five-year intervals: 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64. The midpoints of these intervals have been used to create the "continuous" age measure used in some analyses.

Minority languages: Twelve specific minority languages are coded in the Household Sample File: Arabic/Lebanese, Chinese, Dutch, French, German, Greek, Italian, Maltese, Polish, Serbian and Croatian, Spanish, Vietnamese. They comprise 74 percent of the population that reports speaking a language other than English at home. The minority language concentration variable (CONC) is constructed from these data as follows: Each respondent is assigned a value equal to the percentage of the population aged 15–64 in the region (defined broadly using information on location) in which he or she lives that reports the same second (minority) language as the respondent. The CONC variable is assigned a value of zero for those who speak only English, the 26 percent of respondents who speak a minority language at home that is not among those listed or who did not report the specific language.

Note: The excluded categories for the categorical variables define the reference group for the analyses as follows:

Language proficiency: Nonmarried residents of metropolitan areas, who do not have children and were born in an English-speaking country.

Participation: 40–44-year-old nonmarried immigrants from an English-speaking country with a duration of residence in Australia in excess of 19 years, who do not have children and do not hold Australia citizenship.

Notes

1 The English language fluency rates in Australia among adult immigrants are 74 percent for men and 71 percent for women, in contrast to labor force participation rates of 89 percent and 52 percent, respectively.

2 A number of general issues concerning language needs and difficulties faced by major immigrant groups are outlined in Cox (1975).

3 Alternatively, they may be categorized as exposure variables (language in origin, duration in the destination, minority language concentration, spouse of the same language) and efficiency variables (age at immigration, schooling).

4 Immigration to Australia for persons who are not refugees or the immediate relatives of an Australian resident is rationed on the basis of a set of skill-based criteria, one of which is English language proficiency.

5 The same effects will arise when the birthplace concentration of the region of residence, a proxy measure of minority language usage, is considered (see Chiswick & Miller 1995).

6 The latter are discussed in Evans (1989) and Kossoudji (1988).

7 The Data Appendix provides a description of all variables used in statistical analyses.

8 The strengths and weaknesses suggest that there is merit to conducting parallel analyses using the 1981 Census. Results from such a study are similar to those presented below (based on the 1986 data) and thus are not reported here. In the 1981 analysis, foreign marriage has a significant negative effect on English language fluency among both men and women.

9 This classification is based on Chiswick and Miller (1995). They base the division on the results of earnings equations estimated on a sample of immigrants from non-English-speaking countries.

10 Nonparticipants are included in the analysis because of the interest in studying gender differences. Furthermore, the analysis does not require any labor market variables that would be defined only for the employment segment of the population.

11 The independent variables explain only 6 percent of the small observed difference (2.9 percentage points) between men and women.

12 The birthplace group "North America" includes the United States, Canada, Caribbean, El Salvador, Mexico and Other American. Most are from the United States and Canada.

13 It is noted that analysis of the 1981 Census data shows that foreign marriage reduces dominant language fluency by almost 4 percentage points for males and by 3 percentage points for females. Marriage after arrival in Australia, compared to being unmarried, does not affect dominant language proficiency.

14 The minority language concentration effect appears to be considerably stronger in Australia than in either Canada or the United States (Chiswick and Miller 1995).

15 An alternative decomposition is provided by

$$\overline{\mathrm{GOODENG}}_m - \overline{\mathrm{GOODENG}}_f = \hat{\beta}_f(\bar{X}_m - \bar{X}_f) + (\hat{\beta}_m - \hat{\beta}_f)\bar{X}_m.$$

Basing the analysis on this alternative does not affect the conclusions of this study in any material way.

16 Selected coefficients and *t*-statistics for males and females are as follows:

Variable	*Males*	*Females*
EDUC	0.018	0.023
	(12.62)	(13.62)
AGE	–0.003	–0.003
	(7.91)	(7.46)
PER	0.007	0.006
	(19.64)	(16.43)
CON	–0.075	–0.030
	(3.12)	(1.29)
CON*EDUC	0.004	0.004
	(3.37)	(3.57)
CON*AGE	–0.003	–0.003
	(8.89)	(9.53)
CON*PER	0.005	0.004
	(9.55)	(8.61)

17 These six regions include 92 percent of the observations from non-English-speaking countries.

18 The statistically insignificant effect of minority language concentration for Southwest Asia (i.e., Middle East) is the only exception.

19 This is mainly due to a male advantages in years of education (around one extra year for each group) and duration of residence (two years for Other Asia, one year for Southwestern Asia).

20 Only in Other Asia does the difference in regression coefficients result in a slight advantage for females. This is primarily due to the positive effect of marriage for men and the negative effect for women. Although both partial effects are insignificant at conventional levels, it is the point estimates that matter in the Blinder (1973) decomposition.

21 Some of the differences between the studies by Brooks and Volker (1985) and Evans (1984) may reflect differences in the modeling strategies employed. Evans treats the 1981 Census unit record sample as an "individual-based" file, whereas Brooks and Volker use it as a "family-based" file. This affects the specification of the estimating equation: Brooks and Volker, for example, include variables for husband's income and the age of the youngest child. The Brooks and Volker data requirements on immediate relatives sharply reduces their sample size and may also cause sample selection bias. Brooks and Volker (p. 49) write, "We have, however, excluded from our data set all individuals who failed to answer (*or whose immediate relatives failed to answer*) all the relevant questions" (emphasis added). While each of these approaches to modeling has merit, the considerable importance to the female labor supply decision of the family-based variables included in the study by Brooks and Volker has influenced the decision to use the 1986 Census data as a "family-based" file in the analysis of labor supply in this paper.

22 Eyland et al. (1984) note, "The data suggested that women from Mediterranean countries worked at an earlier age, stayed in the workforce while children were young but retired earlier than Australians." Further analysis of these patterns was not attempted due to the small size of the relevant sample.

23 The negative age—participation profile could be associated with a cohort effect.

Because of changes in societal attitudes toward working women and the decline in the gender differential in relative wages in recent decades, older age groups will possess a lifetime participation profile that is everywhere below that of younger age groups. A negative relationship between age and participation may therefore result from the transition in the cross section from younger to older cohorts.

24 Cardwell and Rosenzweig (1980) attributed differences in the wage differential between men and married women across metropolitan areas in the United States to demand-side influences, including industrial structure and monopsony power.

25 Kidd and Viney (1991) report a similar pattern, in particular, participation rises with husband's income until $300/week (about the median income), and declines thereafter. They suggest that part of the reason for this pattern may be the high effective tax rate on some forms of government assistance among low-income groups.

26 In a study of employment of immigrant men in the United States, Chiswick (1982) found a similar pattern; lower participation in the first five years and essentially no difference among the other cohorts with a longer duration of residence. The lower participation in the early period may be due in part to a greater enrollment in schooling and other skill-enhancing training (see Hashmi 1987; Chiswick & Miller 1992b).

27 The smaller effect of schooling on labor supply for immigrants than for natives is consistent with the finding in other studies of a smaller partial effect of schooling on earnings among immigrants.

References

Becker, G. S. (1975). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, 2nd ed. Chicago: University of Chicago Press.

Blinder, A. S. (1973). "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8, 436–465.

Brooks, C., and P. A. Volker. (1985). "Labor Market Success and Failure: An Analysis of the Factors Leading to the Workforce Destinations of the Australian Population." In Paul A. Volker (ed.), *The Structure and Duration of Unemployment in Australia: Proceedings of a Conference*, Bureau of Labor Market Research Monograph Series No. 6, Canberra: Australian Government Publishing Service, pp. 43–71.

Cardwell, L. A. and M. R. Rosenzweig. (1980). "Economic Mobility, Monopsonistic Discrimination and Sex Differences in Wages." *Southern Economic Journal* 46(4), 1102–1117.

Chiswick, B. R. (1982). *The Employment of Immigrants in the United States*. Washington: American Enterprise Institute.

Chiswick, B. R. (1986). "Labor Supply and Investment in Child Quality: A Study of Jewish and Non-Jewish Women." *Review of Economics and Statistics* 68, 700–703.

Chiswick, B. R., and P. W. Miller. (1992a). "Language in the Immigrant Labor Market." In Barry R. Chiswick (ed.), *Immigration, Language and Ethnicity: Canada and the United States*. Washington: American Enterprise Institute, pp. 229–296.

Chiswick, B. R., and P. W. Miller. (1992b). *Post-Immigration Qualifications in Australia: Determinants and Consequences*. Canberra: Australian Government Publishing Service.

Chiswick, B. R., and P. W. Miller. (1995). "The Endogeneity Between Language and Earnings: An International Analysis." *Journal of Labor Economics 12*(2).

Cox, D. (1975). "The Role of Ethnic Groups in Migrant Welfare." In Ronald F.

Henderson (ed.), *Commission of Inquiry into Poverty: Welfare of Migrants*. Canberra: Australian Government Publishing Service, pp. 1–150.

Evans, M. D. R. (1984). "Immigrant Women in Australia: Resources, Family and Work." *International Migration Review* 18(4), 1063–1090.

Evans, M. D. R. (1987). "Language Skill, Language Usage and Opportunity: Immigrants in the Australian Labor Market." *Sociology* 21(2), 253–274.

Evans, M. D. R. (1989). "Immigrant Entrepreneurship: Effects of Ethnic Market Size and Isolated Labor Pool." *American Sociological Review* 54, 950–962.

Eyland, E. A., C. A. Mason, and H. M. Lapsley. (1982). "The Determinants of Female Employment." *Economic Record* 58(160), 11–17.

Gregory, R. G., P. McMahon, and B. Whittingham. (1985). "Women in the Australian Labor Force: Trends, Causes and Consequences." *Journal of Labor Economics* 3(1, part 2), S293–S309.

Haig, B. D., and M. P. Wood. (1976). "A Simulation Study of Married Women in the Australian Work Force 1961–1972." *Australian Economic Papers* 15(27), 171–185.

Hashmi, A. (1987). *Post-Migration Investment in Education by Immigrants in the United States*. Unpublished Ph.D. dissertation, The University of Illinois at Chicago.

Johnson, P. (1988). *An Analysis of Female Labor Supply in Australia*. Unpublished Honours dissertation, Department of Economics, The University of Western Australia.

Kidd, M. P., and R. Vinney. (1991). "Sex Discrimination and non-Random Sampling in the Australian Labor Market." *Australian Economic Papers* 30(56), 28–49.

Kossoudji, S. A. (1988). "English Language Ability and the Labor Market Opportunities of Hispanic and East-Asian Immigrant Men." *Journal of Labor Economics* 6(2), 205–228.

Long, M. H. (1990). "Maturational Constraints on Language Development." *Studies in Language Acquisition* 12(3), 251–285.

McAllister, I. (1987). "Speaking the Language: Language Maintenance and English Proficiency Among Immigrant Youth in Australia." *Ethnic and Racial Studies* 9(1), 24–42.

Miller, P. W. (1985). "Female Labor Supply in Australia: Another Example of a Backward-Bending Labor Supply Curve." *Economic Letters* 19, 287–290.

Miller, P. W., and P. A. Volker. (1983). "A Cross-Section Analysis of the Labor Force Participation of Married Women in Australia." *Economic Record* 59(164), 28–42.

Ross, R. T. (1986). "Analysis of the 1980 Sydney Survey of Work Patterns of Married Women: Further Results." *Economic Record* 62(178), 325–337.

Secretariat to the Committee to Advise on Australia's Immigration Policies. (1987). *Understanding Immigration*. Canberra: Australian Government Publishing Service.

Williams, T., Y. Harsel, J. Clancy, H. Miller, and C. Greenwood. (1987). *Participation in Education*, Australian Council for Educational Research, Research Monograph No. 30, Hawthorn.

10 Immigrant adjustment in Israel

The determinants of literacy and fluency in Hebrew and the effects on earnings

With Gaston L. Repetto

Introduction

This chapter is concerned with an econometric analysis of the determinants of Hebrew language proficiency among adult male immigrants in Israel and the effect of this proficiency on the labor market earnings of these immigrants. The analysis is based on the 1972 Census of Israel and parallels an analysis performed for the 1983 Census (Chiswick, 1998). It is not possible to perform a similar analysis for the 1995 Census of Israel as this census did not include any questions on Hebrew language proficiency.

This study differs from analyses of language and earnings of immigrants in Israel performed using the various immigrant absorption surveys (see, for example, Beenstock 1993, 1996a, 1996b; Beenstock and Ben-Menahen, 1997; Eckstein and Shachar, 1995; Eckstein and Weiss, 1998; Neuman, 1998, and the references therein). The census contains a much larger sample size, and includes immigrants over a wide range of durations of residence and from all countries of origin, in contrast to the absorption surveys which have smaller samples, with limited duration in Israel (usually three or fewer years), and sometimes limited to specific origins (e.g. Jews from the former Soviet Union).

The motivation for this study is twofold. One is to develop even further and to sharpen the tests for the robustness of models for the acquisition by immigrants of the destination language and the effects of destination language skills on their earnings. Most such studies have been performed for the US, Canada and Australia, three highly developed, English-speaking countries of overseas settlement. English is an international language which may have value in the labor market even in the non-English speaking countries of origin. Are the models of language acquisition and impact also useful for a less well developed economy (Israel in 1972) in which the destination language is not English? Moreover, unlike the US, Canadian and Australian censuses, which ask only about speaking ability, the 1972 Census of Israel includes information on literacy in Hebrew, through a question on the ability to write in Hebrew, as well as on speaking Hebrew.[1]

Another motivation is to learn more about the immigrant absorption process in Israel. Israel expends considerable resources on Hebrew language training for new immigrants. What are the basic determinants of Hebrew language proficiency and what are the consequences for labor market earnings in the Israeli economy? The answers to these questions will provide insights that can guide immigration policy in countries that ration immigration visas on kinship, skill or refugee criteria, and can guide absorption (adjustment) policy in Israel and elsewhere.

The chapter begins with a thumbnail sketch of the language and earnings models. The details of these models are presented elsewhere. There follows a description of the variables in the 1972 Census of Israel that form the basis for this study. The empirical results are then presented. The chapter closes with a summary and conclusion.

The models—immigrant language acquisition and earnings

The theoretical models of immigrant acquisition of the destination language and immigrant earnings adjustment are presented here in thumbnail fashion as they have been developed elsewhere in detail. For the basic model development for earnings see Chiswick (1978), and for language see Chiswick and Miller (1992, 1995), with an application of these approaches to Israel presented in Chiswick (1998).

Hebrew language proficiency

The model of destination language proficiency among immigrants is based on a human capital framework. Language skills among immigrants are expected to be productive in the labor market and in consumption activities, are acquired at a sacrifice of time and out-of-pocket (direct) expenditures by the immigrant and those financing the immigrant's language acquisition, and these skills are embodied in the person. Therefore, language skills satisfy the three components of the definition of human capital.

The model of acquisition of dominant language proficiency is based on three conceptual explanatory variables: exposure, efficiency and economic factors. These three conceptual variables are discussed in turn with the development of variables to measure their influences.

Proficiency in Hebrew among immigrants in Israel is expected to be greater the more they are exposed to Hebrew. Exposure can be thought of as having three components: exposure prior to immigration, exposure measured in units of time in Israel, and exposure per unit of time in Israel.

The Census does not include any information on exposure to Hebrew prior to immigration.[2] Although country of birth is known, there is no country other than Israel in which Hebrew is a dominant language and most immigrants to Israel do not arrive with a working knowledge of the language. Some absorption surveys, however, do include information on pre-immigration knowledge of or study of Hebrew, and find that it enhances proficiency after immigration (see, for example, Beenstock 1996a).

Exposure in units of time is usually measured by duration of residence in the destination. It is typically measured as the number of years since the person first came to the destination as a permanent migrant. For most immigrants to Israel this is a one time event, although among North American immigrants there is a greater propensity for return migration that sometimes results in re-immigration. It is to be expected that the effect of duration on language skills is initially large and that the incremental effect on proficiency diminishes with duration of residence, suggesting a quadratic specification.

The intensity of exposure per unit of time in the destination is measured by several proxy variables. One is the ability to avoid using Hebrew, which is proxied by the extent to which others in the region within Israel in which the respondent lives speak the same non-Hebrew language as the respondent. The ability to avoid using Hebrew is greater if one speaks a langauge other than Hebrew that is common in the area (e.g. English) compared to a language that is rare (e.g. Greek).

The ability to avoid using Hebrew is also greater if one immigrates with a spouse who speaks the same origin language. This suggests that linguistic interactions within the household are important, and that proficiency would be lower among those whose current marriage was prior to immigration (married overseas). There is no particular hypothesis for the effect of being married after immigration in contrast to remaining single.

Children, especially children born in Israel, may have partially offsetting impacts on parental proficiency. Parents may learn Hebrew from their Israeli-born or Israeli-educated children. On the other hand, children would detract from parental acquisition of language skills if they serve as translators for their parents. Children as translators is likely to be more relevant for consumption activities than for labor market activities.[3] Children also detract from parental destination language proficiency if the parents speak the origin language to preserve it among their children, or among women if children have an adverse effect on female labor supply. Previous research suggests that children born in the destination have a positive effect on the destination language proficiency of their fathers, but that the effect of children is less positive or more negative on their mother's language skills.

Efficiency in language acquisition refers to the process by which exposure is converted into destination language proficiency (human capital). Developmentally children are more efficient in language acquisition than are adults (Long, 1990). It is expected that proficiency would fall with a rise in age at immigration. It is also expected that those with a higher level of schooling would be more proficient.[4] Persons with more schooling may be more efficient (more able) learners; they would also have greater proficiency in their origin language and may have a greater understanding of the structure of languages.

Another factor relevant for the acquisition of destination language skills is the "linguistic distance" between that language and the origin languages. The closer are origin and destination languages, the greater the expected proficiency in the new destination language. Although a measure of linguistic distance between English and other languages has been developed and used successfully for the study of immigrants in the US and Canada (Chiswick and Miller, 1998a), no comparable index has been developed for Hebrew. Yet some languages, such as Arabic, are linguistically closer to Hebrew than are other languages, such as English. Indeed, of all the major immigrant languages, Arabic is linguistically closest to Hebrew.

Labor market factors also affect the incentives to acquire destination language proficiency. Those who expect to receive higher wages if they were to become proficient have a greater incentive to become proficient. This is difficult to model empirically, and while sample selectivity tests have been performed on this proposition for English-speaking immigrant-receiving countries, this will not be done here, in part because of the lack of identifying instruments (Chiswick and Miller, 1992, 1995). It has also been shown elsewhere that immigrants with a higher level of schooling receive a larger increase in earnings from proficiency in the host language (Chiswick and Miller, 1995). That is, it appears that schooling and language skills are complementary inputs in the generation of earnings. As a result, the education variable in the language equation will, in part, reflect the effect of greater economic benefits to the more educated to becoming proficient.

Immigrants to Israel from some countries have higher rates of emigrating, either to return to their origin or to go a third country, than from others. The higher the probability of an immigrant leaving Israel, the shorter is the expected duration of residence, and given the country-specific nature of Hebrew, the weaker the incentive to invest in the language. Indeed, to the extent that immigrants to Israel from the US and Canada have high propensities for return migration and immigrants from the Arab countries of North Africa and the Middle East have virtually no return migration, one would expect lesser fluency in Hebrew among the former than among the latter, other things being the same (Beenstock, 1996b, Blejer and Goldberg, 1980).

As a results of this discussion the analysis of Hebrew language proficiency for adult males is based on the following equation:

$$\text{LANG} = f(\underset{+}{\text{YSM}}, \underset{-}{\text{YSMSQ}}, \underset{-}{\text{AGE}}, \underset{+}{\text{EDUC}}, \underset{?}{\text{MARR}}, \underset{-}{\text{MARROVER}}, \underset{-}{\text{CONPR}},$$
$$\underset{?}{\text{Children}}, \underset{+}{\text{Children born in Israel}}, \underset{?}{\text{Region of Residence}}, \underset{?}{\text{Country of Birth}}),$$

where LANG is a measure of proficiency in Hebrew, YSM, YSMSQ, AGE and EDUC denote years since migration and its square, age and educational attainment, respectively, while MARR and MARROVER are dichotomous variables that are unity for those currently married and those whose marriage occurred prior to immigration, respectively. CONPR is a language concentration measure. The actual measures of these variables available from the 1972 Census are discussed below. The hypothesized signs of the partial effects are indicated below the variables.

Earnings

The modeling of the effect of Hebrew language skills on earnings is much more straightforward. The approach uses the "human capital earnings function" which relates the natural logarithm of earnings to human capital (e.g. schooling and years of labor market experience) and demographic variables (e.g. gender, marital status, region of residence, etc.). It has been expanded to incorporate immigrant-related variables, including duration of residence, citizenship, country of origin, and of special interest here, destination language skills (Chiswick, 1978; Chiswick and Miller, 1995).

The earnings equation may be written as:

$$\text{LNY}=f(\underset{+}{\text{EDUC}}, \underset{+}{\text{EXP}}, \underset{-}{\text{EXPSQ}}, \underset{+}{\text{YSM}}, \underset{-}{\text{YSMSQ}}, \underset{+}{\text{MARR}}, \underset{+}{\text{LANG}}, \underset{+}{\text{CITIZ}},$$
$$\underset{?}{\text{Region of residence}}, \underset{?}{\text{Country of Birth}}),$$

where LNY is the natural logarithm of earnings, EXP and EXPSQ are years of potential labor market experience and its square, and CITIZ is a dichotomous variable for citizenship.[5] The hypothesized signs are indicated below the explanatory variables.

The data—1972 Census of Israel

The data under study are the microdata sample created by the Central Bureau of Statistics from the 1972 Census of Israel. This is a 20 percent (one-in-five) simple random sample of the Israeli population. The data drawn from this sample for this study are limited to foreign-born Jewish men age 25 to 64 years in 1972 who were not enrolled in a Yeshiva in 1972. The age limits include the prime labor force years and are beyond the usual age of compulsory military service. Those enrolled in a Yeshiva are not labor market participants. The analysis is limited to males at this exploratory stage.

The language questions in the 1972 Census are questions (11) and (12). Translated into English they are:

11) Do you know how to write (at least a simple letter)?

a) Do you know how to write in Hebrew? Yes, No

b) Do you know how to write in Arabic? Yes, No

c) Do you know how to write in another language or languages? Yes, No

Respondents were not asked to specify the other languages in which they could write.

12) What is the language (or languages) that you speak every day?

____ Record the sole language or the primary language.

____ Record the second language.

____ Record the third language.

(Do not record languages that you know but do not speak every day).

There is no information on the degree of literacy or the degree of fluency in spoken Hebrew. Literacy is expressed in this study as a dichotomous variable taking the value of unity for those who can write at least a simple letter in Hebrew, and zero otherwise (HEBWRIT). For speaking fluency four categories are considered: only speaks Hebrew (HEBONLY), speaks Hebrew as a primary but not only language (HEBPRIM), speaks it as a secondary or tertiary language (HEBSECTH), and does not speak Hebrew on a daily basis (HEBNONE). In some analyses the Hebrew fluency variable is dichotomous, where HEBSOP equals unity for those for whom it is the only or primary language spoken daily, and it is zero otherwise.

The earnings variable is the gross annual salary for male wage and salary workers in Israeli Liras, the currency used in Israel at that time. The earnings analysis is performed only for those with positive earnings.

Most of the explanatory variables used in the analysis are fairly straightforward. Age (AGE) is measured in years since birth. Education (EDUC) is measured as years of schooling attended, with a top coding of 22 years. The maximum potential labor market experience is measured as age minus schooling minus five (EXP=AGE–EDUC–5), and is defined to equal zero for any negative values. Duration in Israel (YSM) is measured as the current year (1972) minus the year of "aliya" (immigration) to Israel.

Marital status (MARR) is unity for those who are currently married and zero otherwise. Married overseas (MARROVER) is unity for those in their first marriage if this took place prior to immigration (year of first marriage equal to or earlier than year of immigration), otherwise it is zero. The two children variables are a dichotomous variable equal to unity if there are children under age 20 living in the household (CHILDREN) and a dichotomous variable that equals unity if any of the married, widowed or divorced women living in the household has a child born in Israel (HCHILBIS).

The region of residence dichotomous variables are defined for Tel Aviv and Jerusalem, with the rest of the country as the benchmark. The various countries of birth are combined into seven regions: Asia (nearly all coming from the Asian countries of the Middle East), North Africa (from Morocco to Somalia), English-speaking countries (US, Canada, UK, Ireland, Australia, New Zealand, South Africa), Western Europe (other than the UK and Ireland), USSR and Latin America, with Eastern Europe (i.e. the post-war Communist bloc countries of Europe other than the USSR) as the benchmark.

The language concentration measure (CONPR) is constructed in the following manner. The foreign-born adult male Jewish population is divided into the "natural regions" of the country.[6] Within each region (i), the percentage of the group speaking each of the 12 most frequently reported only or primary languages other than Hebrew (j) is computed.[7] This percentage is the CONPR for each respondent in the region (i) reporting language (j) as their only or primary language. Thus, for a French speaker in Jerusalem the CONPR is the percent of foreign-born adult Jewish men in Jerusalem who speak French as their only or primary language. For those who speak only Hebrew, CONPR is defined to be zero since CONPR refers to the concentrations speaking languages other than Hebrew. It is also defined to equal zero for those reporting a primary language not in the top 12 languages other than Hebrew because the population density of these language speakers is so low.

Empirical analysis

This section reports the statistical analysis of spoken Hebrew language usage and ability to write in Hebrew, followed by the statistical analysis of annual earnings among wage and salaried workers.

Hebrew language proficiency

The distribution of language skills among adult foreign-born Jewish men in Israel is reported in Table 10.1. Hebrew is the only language spoken on a daily basis among 24 percent of the men, and for another 52 percent it is the primary but not the only language, making a total of 75 percent for whom it is the only or primary language. All told, including second and third languages reported, 89 percent report Hebrew as spoken daily.

The second most frequently spoken language is Arabic, which is spoken primarily by North African and Middle Eastern immigrants. It is spoken by 5 percent of the Jewish immigrants as an only or primary language, but by 23 percent if second and tertiary languages are included. Yiddish, the trad-

Table 10.1 Frequency distribution of languages spoken in Israel, 1972 (foreign-born Jewish men, age 25 to 64)

Language	*Only and primary*[(a)]	*Only, primary and second*	*Only, primary, second and third*
Hebrew	75.4	86.5	88.9
Arabic	4.8	20.0	22.9
Yiddish	4.0	14.3	17.7
Romanian	3.4	8.4	9.9
French	1.9	7.7	10.7
Spanish-Ladino	1.6	4.9	6.0
English	1.3	5.4	8.6
Hungarian	1.1	3.3	4.1
Kurdish	0.9	1.8	2.1
Russian	0.8	2.0	3.0
German	0.8	3.9	5.9
Persian	0.7	2.2	2.3
Polish	0.7	3.2	5.2
Other	2.9	2.3	2.2
Total	100.0	(b)	(b)

Notes: Sample size: 92,797.
(a) Hebrew is the only language spoken by 23.7 percent.
(b) Column adds to more than 100 percent due to dual and triple language fluency.

Source: 1972 Census of Population and Housing, Israel, Public Use Sample, Demographic File, 20 percent sample of the population.

itional language of East European Jews, is in third place. English, an important international language, is spoken as the only or primary language of only 1 percent of the sample, but if second and third languages spoken are included, the proportion increases to 9 percent. The top 12 languages after Hebrew are spoken as the only or primary language by 22 percent of the adult foreign-born men, or by 88 percent of those reporting an only or primary language other than Hebrew.

There is a strong relation between duration in Israel and proficiency in Hebrew (Table 10.2). The proportion of foreign-born men reporting Hebrew as the only or primary language increases with duration in Israel. The pattern of improvement is similar to the one observed 11 years later in the 1983 Census (Chiswick, 1998).[8] This suggests that the pattern is not due to inherently poorer Hebrew language ability among the more recent cohorts of immigrants that would put them at a linguistic disadvantage throughout their life in Israel, but rather reflects a longitudinal effect-Hebrew language skills improving with duration of residence.

The means and standard deviations of the variables used in the analyses are reported in Table 10.3 by the degree of Hebrew usage. Those who are more proficient in Hebrew, that is, they speak it as their only or primary

Table 10.2 Hebrew speaking skills by duration in Israel, 1972 (foreign-born Jewish men, age 25 to 64)

Duration (years)	*Sample size*	*Only lang. (%)*	*Primary lang (%)*	*Second + third lang. (%)*	*Does not speak (%)*	*Unknown (%)*	*Total (%)*
0–5	5,799	3.22	21.54	31.63	39.14	4.47	100.00
6–10	9,377	8.37	43.50	25.10	20.65	2.38	100.00
11–15	8,858	12.10	49.73	21.64	14.41	2.12	100.00
16–20	8,043	21.19	55.94	13.96	6.39	2.52	100.00
21–25(a)	42,582	27.46	55.38	10.56	4.64	1.97	100.00
26–30	4,574	33.95	59.93	5.18	2.32	1.62	100.00
31–35	3,831	31.66	59.07	5.69	2.01	1.57	100.00
Over 35	9,733	38.42	54.43	3.82	1.80	1.53	100.00
Total	92,797	23.65	51.70	13.52	8.97	2.15	100.00

Note

(a) These immigrants arrived in 1947–51, encompassing the year of independence and the subsequent large immigration from Europe, North Africa and the Middle East.

Source: 1972 Census of Population and Housing, Israel, Public Use Sample, Demographic File, 20 percent sample of the population.

language, compared to those with lesser proficiency, tend to be younger, better educated, resided longer in Israel, were married in Israel and have children, with at least some children born in Israel. These are, however, simple relationships.

The multiple regression analysis of spoken Hebrew language usage is reported in Table 10.4 using Ordinary Least Squares (OLS) and Logit analysis. Both procedures tell essentially the same story: using Hebrew as the only or primary language is greater the higher the level of education, the longer the duration of residence, and the younger the age at immigration (age when duration in Israel is held constant).[9] Those who married their current spouse prior to immigration are less likely to use Hebrew. The comparison of those who married after immigration with those not married shows an ambiguous pattern: lesser fluency for the former in the OLS analysis but no significant difference in the logit analysis. Children, especially if born in Israel, are associated with a greater use of Hebrew among their fathers.

Compared to the rest of Israel, those living in the more religiously observant and traditional Jerusalem are more likely to use Hebrew as their only or primary language. Compared to the rest of Israel other than Jerusalem, Hebrew is less likely to play this role in Tel Aviv. Even after controlling for place of residence in Israel through the Tel Aviv/Jerusalem city variables, the linguistic concentration variable (CONPR) is statistically significant. That is, immigrants living in a region in which a larger proportion of the foreign-born speak the same non-Hebrew language as the respondent are less likely to

Table 10.3 Means and standard deviation of variables used in language analysis, Israel, 1972 (foreign-born Jewish men, age 25 to 64)

Variable	*All*	*Hebrew only*	*Hebrew primary*	*Hebrew second or third*	*No Hebrew*
AGE	44.63	41.08	44.09	49.00	51.25
	(11.63)	(11.51)	(11.29)	(10.64)	(10.39)
EDUC	9.09	9.65	9.49	8.04	7.02
	(4.45)	(4.04)	(4.21)	(5.01)	(4.92)
YSM	21.14	25.19	22.29	15.87	12.35
	(9.81)	(8.82)	(8.96)	(9.02)	(9.31)
YSMSQ	543.00	712.53	576.89	333.34	239.10
	(466.97)	(513.69)	(453.38)	(331.32)	(317.08)
MARR	0.90	0.90	0.91	0.92	0.89
	(0.30)	(0.30)	(0.29)	(0.27)	(0.32)
MARROVER	0.37	0.16	0.33	0.71	0.79
	(0.48)	(0.36)	(0.47)	(0.45)	(0.41)
CHILDREN	0.68	0.74	0.72	0.61	0.48
	(0.47)	(0.44)	(0.45)	(0.49)	(0.50)
HCHILBIS	0.69	0.81	0.76	0.52	0.32
	(0.46)	(0.39)	(0.43)	(0.50)	(0.47)
TEL-AVIV	0.35	0.39	0.36	0.31	0.30
	(0.48)	(0.49)	(0.48)	(0.46)	(0.46)
JERUSALEM	0.07	0.07	0.08	0.07	0.04
	(0.26)	(0.26)	(0.26)	(0.25)	(0.21)
ASIA	0.21	0.30	0.21	0.15	0.11
	(0.41)	(0.46)	(0.41)	(0.35)	(0.31)
NORTHAFR	0.23	0.19	0.24	0.26	0.22
	(0.42)	(0.40)	(0.43)	(0.44)	(0.42)
USSR	0.06	0.06	0.05	0.05	0.12
	(0.23)	(0.25)	(0.21)	(0.22)	(0.32)
EASTEURO	0.41	0.37	0.41	0.44	0.48
	(0.49)	(0.48)	(0.50)	(0.50)	(0.50)
WESTEURO	0.06	0.06	0.06	0.05	0.03
	(0.23)	(0.24)	(0.24)	(0.21)	(0.18)
ENGLSPEA	0.01	0.005	0.01	0.04	0.02
	(0.12)	(0.07)	(0.10)	(0.20)	(0.15)
LATIAMER	0.01	0.005	0.02	0.02	0.02
	(0.12)	(0.08)	(0.12)	(0.13)	(0.14)
CONPR	10.19	0	13.62	14.07	13.84
	(10.18)	(0)	(9.42)	(9.76)	(9.83)
Sample size	92,797	21,947	47,977	12,550	8,328

Note: The variables are defined in the text.
Source: 1972 Census of Population and Housing, Israel, Public Use Sample, Demographic File, 20 percent sample of the population.

Table 10.4 Analysis of determinants of speaking Hebrew used as only or primary language, Israel, 1972 (foreign-born Jewish men, age 25 to 64)

Variables	*OLS*	*Logit*
CONSTANT	0.4544	0.0769
	(47.67)	[0.96]
AGE	–0.0073	–0.0595
	(–47.41)	[–42.59]
EDUC	0.0142	0.1137
	(45.53)	[42.25]
YSM	0.0339	0.2022
	(77.70)	[51.25]
YSMSQ	–0.0004	–0.0019
	(–44.94)	[–21.52]
MARR	–0.0190	0.0139
	(–4.10)	[0.35]
MARROVER	–0.0716	–0.4560
	(–20.20)	[–15.81]
CHILDREN	0.0262	0.3125
	(7.80)	[11.16]
HCHILBIS	0.0755	0.2811
	(21.04)	[9.88]
TEL-AVIV	–0.0080	–0.0776
	(–3.00)	[–3.23]
JERUSALEM	0.0143	0.1522
	(2.99)	[3.41]
ASIA	0.0556	0.4876
	(15.87)	[14.88]
NORTHAFR	0.0809	0.6203
	(23.01)	[20.17]
USSR	0.0012	0.0392
	(0.21)	[0.76]
WESTEURO	–0.0671	–0.6127
	(–12.23)	[–11.56]
ENGLSPEA	–0.1703	–1.2923
	(–16.25)	[–16.93]
LATIAMER	0.0552	0.0734
	(5.27)	[0.90]
CONPR	–0.0035	–0.0307
	(–28.02)	[–28.32]
R^2 adj.	0.3339	
Chi square		28,299.44
Sample size	81,594	81,594

Notes: Dependent variable: HEBSOP=1 if speak Hebrew as only or primary language, otherwise it is zero.
t ratios are in parentheses.
Asymptotic t ratios are in brackets.

Source: 1972 Census of Population and Housing, Israel, Public Use Sample, Demographic File, 20 percent sample of the population.

report they speak Hebrew as their only or primary language. This suggests that the greater the ease among immigrants to rely on their origin language, the less likely are they to use Hebrew.

Country of birth matters. Compared to Jewish immigrants born in Eastern Europe, those from North Africa and the Middle East (Asia) are more likely to use Hebrew as their only or primary language. These immigrants share with those of Eastern Europe a negligible propensity for return migration, but the language of their countries of origin (primarily Arabic) is linguistically closer to Hebrew than are the European languages. At the other extreme, those from Western Europe, and especially those from the English-speaking developed countries, have a much lower use of Hebrew. It is the immigrants from the English-speaking developed countries in particular that have a high degree of return migration (see Beenstock, 1996b; Blejer and Goldberg, 1980). Indeed, Hebrew usage by country of origin seems to be greater the lower the degree of return migration and the closer the language of origin is to Hebrew.

There is a strong relationship between Hebrew speaking frequency and ability to write in Hebrew, but the correlation is far from perfect.[10] Table 10.5 reports the OLS and logit analysis of the determinants of being able to write (a simple letter) in Hebrew. The patterns are generally the same as for speaking. Those with more schooling, who have been in Israel a longer period of time, who immigrated at a younger age, who did not marry prior to immigration, and who have children, especially children born in Israel, are more likely to be able to write a letter in Hebrew. Those who live in Jerusalem are more likely to be able to write in Hebrew in the logit analysis, but there is no difference between Tel Aviv and the rest of the country. Those who live in areas where more immigrants speak their origin language (CONPR) are less able to write in Hebrew.

Country of origin also matters for writing, with the patterns being similar to that for speaking. Those from North Africa and the Middle East have a greater propensity to be able to write in Hebrew, compared to Eastern European immigrants, even though the Arabic alphabet is very different from the Hebrew alphabet. Those from Western Europe and the English-speaking countries are less able to write in Hebrew than Eastern European immigrants.

Although the dependent variables are not strictly comparable, and the general patterns are very similar, there are some interesting differences in the partial effects of some of the explanatory variables on speaking and writing Hebrew. Education has a much larger impact on writing skills than on speaking, whereas duration in Israel has a larger impact on improving speaking than on improving writing. The negative impact on Hebrew proficiency of the linguistic concentration measure is stronger for speaking than for writing. These patterns for the differential effect on speaking and writing of these explanatory variables are remarkably similar to what was found in a study of English speaking and reading proficiency among illegal aliens who obtained legal status (primarily of Hispanic origin) in the United States (Chiswick and

Table 10.5 Analysis of determinants of Hebrew writing, Israel, 1972 (foreign-born Jewish men, age 25 to 64)

Variables	*OLS*	*Logit*
CONSTANT	0.4164	–1.2570
	(48.41)	[–13.39]
AGE	–0.0051	–0.0468
	(–36.22)	[–29.34]
EDUC	0.0339	0.3186
	(120.36)	[92.79]
YSM	0.0195	0.1488
	(49.46)	[32.82]
YSMSQ	–0.0002	–0.0012
	(–27.19)	[–10.86]
MARR	0.0104	0.2023
	(2.49)	[4.37]
MARROVER	–0.0408	–0.3965
	(–12.71)	[–11.69]
CHILDREN	0.0010	0.2172
	(3.28)	[6.65]
HCHILBIS	0.0453	0.2309
	(13.95)	[6.91]
TEL-AVIV	0.0033	0.0008
	(1.36)	[0.03]
JERUSALEM	–0.0041	0.1277
	(–0.95)	[2.45]
ASIA	0.0237	0.3574
	(7.48)	[9.65]
NORTHAFR	0.0347	0.4277
	(10.91)	[12.17]
USSR	–0.0165	0.0568
	(–3.30)	[0.96]
WESTEURO	–0.0533	–0.3764
	(–10.78)	[–5.31]
ENGLSPEA	–0.0170	–0.5139
	(–1.799)	[–4.65]
LATIAMER	0.0303	0.1167
	(3.22)	[1.06]
CONPR	–0.0005	–0.0038
	(–4.47)	[–3.11]
R^2 adj.	0.3077	
Chi square		26,293.33
Sample size	81,025	81,025

Notes: Dependent variable: HEBWRIT=1 if know how to write at least a simple letter in Hebrew, otherwise it is zero.
t ratios are in parentheses.
Asymptotic *t* ratios are in brackets.

Source: 1972 Census of Population and Housing, Israel, Public Use Sample, Demographic File, 20 percent sample of the population.

Miller, 1999). This provides additional support for the robustness of the findings within and across countries. The comparison with the US study also suggests very similar determinants of the two dimensions of literacy, reading and writing skills.

As a further refinement of the analysis, Table 10.6 reports the multinomial logit analysis of the four category Hebrew variable: only Hebrew, Hebrew primary, Hebrew secondary (or tertiary) and no Hebrew spoken daily.[11] Although the signs and levels of statistical significance are indicated in Table 10.6, the magnitudes can be difficult to interpret. To facilitate interpretation, Table 10.7 reports the probabilities that an individual with a particular set of characteristics will be in each of the four language groups, based on the multinomial logit analysis in Table 10.6. The probabilities in a row sum to unity.

The top row of Table 10.7 reports the probabilities for the person with "mean" characteristics, while the second row reports the probabilities for the base or reference person. The base or reference person in Table 10.7 has a mean age (age 44.6), level of education (9.1 years), and duration in Israel (21.1 years), is married, but married after immigration, and has children that were born in Israel. The reference person was born in Western Europe and does not live in Jerusalem or Tel Aviv.

The analysis indicates that the probability of speaking only Hebrew or speaking Hebrew as a primary language increases with education. The probability of speaking Hebrew as the only or primary language increases from 83 percent for those with base characteristics and 10 years of schooling to 90 per cent for those with 15 years of schooling. Although the probability of speaking Hebrew as a primary but not only language decreases from 61 percent to 55 percent from 16 to 38 years duration in Israel, the probability that it is the only language spoken daily increases from 11 percent to 41 percent. The effect of an older age at migration on speaking Hebrew (the age variable for the base duration in Israel) is dramatic; Hebrew usage is lower the older the age at migration.

Those who married overseas are less likely to speak only Hebrew or to speak it as a primary language. Not having children or having children who were not born in Israel is associated with lesser use of Hebrew.

The analysis also shows important differences by country of origin. Immigrants from English-speaking countries are much more likely to report Hebrew as their second (or third) language spoken or that they speak no Hebrew. English is presumably being predicted as the primary or only language spoken on a daily basis by 34 percent of those from English-speaking countries (28.5 percent speak Hebrew as the second or third language and 5.5 percent do not speak Hebrew on a daily basis).

Earnings

The mean annual earnings of wage and salary workers in 1972 in Israeli Liras are reported in Table 10.8 for adult Jewish immigrants by Hebrew speaking

Table 10.6 Multinomial logit analysis of Hebrew language usage, Israel, 1972 (foreign-born Jewish men, age 25 to 64)

Variable	*Primary lang. relative to only Hebrew*	*Second and third lang, relative to only Hebrew*	*No Hebrew relative to only Hebrew*
INTERCEPT	0.9195	0.6451	1.5274
	(11.44)	(5.84)	(11.95)
AGE	0.0442	0.0890	0.1126
	(36.24)	(48.41)	(50.88)
EDUC	–0.0031	–0.1025	0.1859
	(–1.30)	(–29.28)	(–43.18)
YSM	–0.0979	–0.2406	–0.3628
	(–22.23)	(–40.55)	(–56.59)
YSMSQ	0.0006	0.0018	0.0038
	(8.36)	(13.87)	(28.31)
MARR	–0.0814	0.0012	–0.1883
	(–2.35)	(0.02)	(–3.05)
MARROVER	0.1704	0.6542	0.5125
	(6.03)	(16.69)	(10.24)
CHILDREN	0.1033	–0.2151	–0.4807
	(0.39)	(–5.68)	(–10.92)
HCHILBIS	0.0081	–0.1810	–0.4631
	(0.28)	(–4.56)	(–9.96)
TEL-AVIV	–0.0065	0.0169	0.0076
	(–0.34)	(0.56)	(0.20)
JERUSALEM	0.0628	–0.0551	–0.3454
	(1.79)	(–0.99)	(–4.59)
ASIA	–0.3054	–0.6780	–1.0507
	(–12.36)	(–16.54)	(–19.40)
NORTHAFR	0.0403	–0.3462	–0.7519
	(1.49)	(–8.72)	(–15.24)
USSR	–0.2621	–0.4518	–0.1865
	(–6.35)	(–6.58)	(–2.56)
WESTEURO	0.2613	0.7274	0.5233
	(6.72)	(11.29)	(5.95)
ENGLSPEA	0.3041	1.6422	1.0617
	(2.60)	(12.85)	(7.04)
LATIAMER	0.5341	0.3305	0.2425
	(5.04)	(2.53)	(1.64)
Chi square	34,923.53	34,923.53	34,923.53
Sample size	81,602	81,602	81,602

Notes: Dependent variable: HEBPRIML=3 if speaking Hebrew only, HEBPRIML=2 if Hebrew is primary language, HEBPRIML=1 if Hebrew is used as second or third language, and HEBPRIML=0 if individual does not speak Hebrew. Base category HEBPRIML=3.
z statistics are in parentheses.

Source: 1972 Census of Population and Housing, Israel, Public Use Sample, Demographic File, 20 percent sample of the population.

Table 10.7 Predicted probability of being in each language category, Israel, 1972 (foreign-born Jewish men, age 25 to 64)

Variable	*Value*	*Only Hebrew*	*Hebrew primary*	*Hebrew second or third*	*No Hebrew*
MEANS		0.2068	0.6477	0.1096	0.0359
BASE		0.1594	0.6538	0.1455	0.0412
AGE	34	0.2502	0.6416	0.0887	0.0196
	45	0.1567	0.6532	0.1478	0.0423
	56	0.0891	0.6042	0.2238	0.0829
EDUC	5	0.1409	0.5854	0.1957	0.0780
	10	0.1629	0.6661	0.1354	0.0356
	15	0.1782	0.7177	0.0888	0.0154
YSM	16 (256)	0.1093	0.6243	0.2050	0.0613
(& YSMSQ)	27 (728)	0.2528	0.6528	0.0786	0.0158
	38 (1444)	0.4083	0.5519	0.0327	0.0071
MARR	1	Base	Base	Base	Base
	0	0.1498	0.6667	0.1366	0.0468
MARROVER	1	0.1242	0.6041	0.2181	0.0536
	0	Base	Base	Base	Base
CHILDREN	1	Base	Base	Base	Base
	0	0.1600	0.5919	0.1811	0.0669
HCHILBIS	1	Base	Base	Base	Base
	0	0.1521	0.6189	0.1664	0.0625
TEL-AVIV	1	0.1596	0.6505	0.1482	0.0416
JERUSALEM	1	0.1559	0.6809	0.1347	0.0285
Rest of Israel	0	Base	Base	Base	Base
EASTEURO	0	0.2104	0.6645	0.0928	0.0323
ASIA	1	0.2774	0.6456	0.0621	0.0149
NORTHAFR	1	0.2140	0.7038	0.0668	0.0155
USSR	1	0.2605	0.6332	0.0732	0.0331
WESTEURO	1	Base	Base	Base	Base
ENGLSPEA	1	0.1249	0.5349	0.2848	0.0554
LATIAMER	1	0.1389	0.7486	0.0853	0.0271

Notes: Sample Size = 81,602.
The base or reference person is a 44.6 year old Jewish male born in Western Europe with 9.1 years of schooling, who has lived in Israel 21.1 years, and is married but married after coming to Israel, has children born in Israel and does not live in Tel Aviv or Jerusalem. Row totals may not add to 1.0000 due to rounding.
Multinomial logit coefficients obtained from Table 10.6.

Source: 1972 Census of Population and Housing, Israel, Public Use Sample, Demographic File, 20 percent sample of the population.

Table 10.8 Mean earnings by Hebrew language skills, Israel, 1972 (foreign-born Jewish men, age 25 to 64, with positive gross annual earnings, wage and salaried workers)

	Speaking Hebrew				
Writing Hebrew	*Only*	*Primary*	*Second or third*	*No Hebrew*	*Total*
Yes	12,310.88 (14,708) [7,315.59]	12,518.10 (32,160) [7,736.96)	11,034.99 (5,352) [8,276.64]	9,996.84 (1,840) [7,217.98]	12,229.08 (54,060) [7,686.70]
No	7,594.10 (705) [4,983.46]	8,024.56 (2,865) [4,956.33]	7,782.43 (3,029) [4,894.87]	6,952.52 (3,286) [5,975.12]	7,563.29 (9,885) [5,319.15]
Total	12,095.13 (15,413) [7,292.14]	12,150.54 (35,025) [7,647.83]	9,859.47 (8,381) [7,405.50]	8,045.29 (5,126) [6,611.50]	11,507.81 (63,945) [7,561.08]

Notes: Number of cases (N) are in parentheses.
Standard deviations (STD) are in brackets.
Earnings in 1972 Israeli Lira.

Source: 1972 Census of Population and Housing, Israel, Public Use Sample, Demographic File, 20 percent sample of the population.

and writing proficiencies.[12] These simple relationships show that for each level of speaking Hebrew, earnings are higher for those who can write in Hebrew. For each writing level, earnings are higher for those who speak Hebrew as their only or primary language, compared to those for whom it is a second or third language or who do not speak Hebrew. The highest earnings are received by those for whom Hebrew is the primary language, but they also speak another language on a daily basis, and they can write a letter in Hebrew (IL 12,518). The lowest earnings are received by those who neither speak Hebrew on a daily basis, nor can they write in Hebrew (IL 8,045).

The statistical analysis of earnings is reported in Tables 10.9 and 10.10 to ascertain the effects of Hebrew speaking usage and Hebrew literacy, when other variables are held constant. Table 10.9 column (1) reports the basic earnings equation without the language variables, while speaking skills are added in the next three columns. In Table 10.10 the effects on earnings are examined for writing skills and for speaking and writing combined.

The effects of the basic variables on earnings are largely invariant with respect to the inclusion of speaking and writing variables. Earnings increase with additional human capital (Table 10.9, column 1). Earnings increase by about 4 percent per year of schooling (EDUC), which is lower than the 6 percent found in the analysis of immigrants in the 1983 Census, as well as lower than in the United States and Canada (Chiswick, 1998; Chiswick and Miller, 1992).

Earnings increase with pre-immigration labor market experience (EXP),

Table 10.9 Analysis of earnings with language variables, Israel, 1972 (foreign-born Jewish men, age 25 to 64, with positive gross annual earnings, wage and salaried workers)

Variables	*(1)*	*(2)*	*(3)*	*(4)*
CONSTANT	7.8380	7.7994	7.9506	7.9545
	(416.80)	(412.21)	(401.86)	(400.71)
EDUC	0.0441	0.0433	0.0429	0.0413
	(63.97)	(62.70)	(62.07)	(59.01)
EXP	0.0304	0.0305	0.0300	0.0305
	(36.14)	(36.25)	(35.51)	(36.04)
EXPSQ	–0.0006	–0.0005	–0.0005	–0.0005
	(–40.43)	(–38.81)	(–38.00)	(–38.25)
MARR	0.3044	0.3038	0.3029	0.3032
	(35.25)	(35.14)	(35.06)	(35.15)
YSM	0.0273	0.0224	0.0212	0.0216
	(33.76)	(26.27)	(24.66)	(25.17)
YSMSQ	–0.0003	–0.0002	–0.0002	–0.0002
	(–17.60)	(–14.18)	(–13.06)	(–14.01)
TEL-AVIV	–0.0163	–0.0159	–0.0159	–0.0168
	(–3.19)	(–3.10)	(–3.10)	(–3.28)
JERUSALEM	–0.0457	–0.0484	–0.0498	–0.0524
	(–4.99)	(–5.28)	(–5.45)	(–5.73)
ASIA	–0.2037	–0.2117	–0.2136	–0.2069
	(–30.45)	(–31.60)	(–31.86)	(–27.47)
NORTHAFR	–0.1376	–0.1482	–0.1523	–0.1419
	(–20.87)	(–22.42)	(–23.04)	(–19.24)
USSR	–0.1142	–0.1166	–0.1121	–0.1117
	(–10.45)	(–10.69)	(–10.28)	(–10.25)
WESTEURO	0.0073	0.0150	0.0141	0.0016
	(0.68)	(1.39)	(1.30)	(0.15)
ENGLSPEA	0.1989	0.2229	0.2078	0.0867
	(9.02)	(10.11)	(9.42)	(3.65)
LATIAMER	0.0799	0.0722	0.0680	0.0708
	(3.57)	(3.23)	(3.04)	(3.18)
HEBSOP	–	0.1282	–	–
	–	(18.79)		
HEBPRIM	–	–	0.0047	–0.0015
			(0.81)	(–0.24)
HEBSECTH	–	–	–0.0854	–0.0941
			(–9.53)	(–10.05)
HEBNONE	–	–	–0.2082	–0.2217
			(–18.80)	(–19.36)
ENGLOPS	–	–	–	0.1597
				(14.02)
ARABOPS	–	–	–	–0.0240
				(–3.21)
R^2 adj.	0.2427	0.2466	0.2483	0.2510
Sample size	61,622	61,416	61,416	61,416

Notes: Dependent variable: natural logarithm of gross annual earnings for wage and salaried workers.
t ratios are in parentheses. Earnings in 1972 Israeli Lira.

Source: 1972 Census of Population and Housing, Israel, Public Use Sample, Demographic File, 20 percent sample of the population.

Table 10.10 Analysis of earnings with language and writing variables, Israel, 1972 (foreign-born Jewish men, age 25 to 64, with positive gross annual earnings, wage and salaried workers)

Variables	*(1)*	*(2)*	*(3)*	*(4)*
CONSTANT	7.8042	7.7814	7.7780	7.9769
	(411.01)	(407.96)	(402.11)	(400.43)
EDUC	0.0405	0.0408	0.0408	0.0408
	(56.12)	(56.39)	(56.37)	(56.37)
T	0.0299	0.0301	0.0301	0.0301
	(35.37)	(35.60)	(35.53)	(35.53)
TSQ	–0.0005	–0.0005	–0.0005	–0.0005
	(–38.35)	(–37.46)	(–37.42)	(–37.42)
MARR	0.3033	0.3030	0.3030	0.3030
	(35.02)	(34.90)	(34.90)	(34.90)
YSM	0.0247	0.0213	0.0213	0.0213
	(29.73)	(24.67)	(24.67)	(24.67)
YSMSQ	–0.0003	–0.0002	–0.0002	–0.0002
	(–15.82)	(–13.45)	(–13.43)	(–13.43)
TEL-AVIV	–0.0176	–0.0171	–0.0171	–0.0171
	(–3.43)	(–3.32)	(–3.33)	(–3.33)
JERUSALEM	–0.0477	–0.0499	–0.0500	–0.0500
	(–5.20)	(–5.42)	(–5.43)	(–5.43)
ASIA	–0.2041	–0.2096	–0.2097	–0.2097
	(–30.37)	(–31.13)	(–31.14)	(–31.14)
NORTHAFR	–0.1408	–0.1484	–0.1485	–0.1485
	(–21.23)	(–22.31)	(–22.32)	(–22.32)
USSR	–0.1120	–0.1143	–0.1141	–0.1141
	(–10.25)	(–10.47)	(–10.45)	(–10.45)
WESTEURO	0.0140	0.0188	0.0187	0.01867
	(1.29)	(1.734)	(1.72)	(1.72)
ENGLSPEA	0.2014	0.2214	0.2200	0.2200
	(9.15)	(10.05)	(9.96)	(9.96)
LATIAMER	0.0808	0.0752	0.0748	0.0748
	(3.62)	(3.36)	(3.35)	(3.0.35)
HEBWRIT	0.1216	0.0888	0.0959	–
	(15.60)	(10.97)	(9.08)	
HEBSOP	–	0.1072	0.1183	–
		(15.00)	(9.23)	
HEBSPWR	–	–	–0.0153	–
				(–1.05)
YSPNWR	–	–	–	–0.0806
				(–7.16)
NSPYWR	–	–	–	–0.1030
				(–12.60)
NSPNWR	–	–	–	–0.1989
				(–20.52)
R^2 adj.	0.2430	0.2453	0.2453	0.2453
Sample size	60,893	60,689	60,689	60,689

Notes: Dependent variable: natural logarithm of gross annual earnings for wage and salaried workers.
t ratios are in parentheses. Earnings in 1972 Israeli Lira.

Source: 1972 Census of Population and Housing, Israel, Public Use Sample, Demographic File, 20 percent sample of the population.

at about 1.8 percent per year when evaluated at 10 years of experience. Experience in Israel (YSM) has a larger effect. Evaluated at 10 years, the effect of an extra year in Israel on earnings rather than an extra year in the country of birth is 2.1 percent when language skills are not held constant and about 1.7 percent when they are. Some of the effect of duration in Israel on earnings operates through language skills, that is, Hebrew language proficiency increases with duration and greater proficiency enhances earnings. Married men earn considerably more than observationally similar men who are not married. These patterns are similar to effects found in the 1983 Census and for other immigrant receiving countries.

Place of residence in Israel matters. Earnings are about 5 percent lower in Jerusalem and 1.6 percent lower in Tel Aviv than in the rest of the country (Table 10.9, column 1). This is slightly different from the pattern in the 1983 Census, an 8 percent lower earnings in Jerusalem and no significant difference in earnings in Tel Aviv. More striking is the difference between Israel, on the one hand, and the US and Canada on the other, where earnings tend to increase with population density or city size (Chiswick and Miller, 1992).

Country of birth also matters. Compared to immigrants from Eastern Europe, earnings are lower by about 20 percent among those from Asia (Middle East), 14 percent among those from North Africa, and 11 percent among those from the USSR (Table 10.9, column 1). There is no significant difference in earnings between Western European and Eastern European origin immigrants. Some immigrants, on the other hand, have earnings significantly higher than Eastern Europeans – about 20 percent higher for those from English-speaking developed countries and 7 percent higher among Latin American immigrants.

When the language variables are added to the earnings equation it is clear that Hebrew language skills matter. Speaking only Hebrew or speaking it as a primary language raises earnings by about 13 percent (Table 10.9, column 2). Compared to those who speak only Hebrew, there is no difference in earnings on the part of those who speak it as a primary language, but earnings are lower by about 9 percent for those who report it as a second or tertiary language, and by over 20 percent for those who do not speak it on a daily basis (Table 10.9, column 3).

Other variables being the same, speaking English on a daily basis is associated with about 15 percent higher earnings (Table 10.9, column 4).[13] On the other hand, speaking Arabic is associated with about 2 percent lower earnings. These patterns are very similar to the earnings differences for English and Arabic speakers in the 1983 Census (Chiswick, 1998). Greater earnings for English language skills, even when country of origin is held constant, may arise because English is an international language.[14] Those engaged in foreign trade or in tourism in Israel may have expanded opportunities if they have some degree of proficiency in English. Moreover, immigrants from the high income English-speaking developed countries have a high opportunity cost of remaining in Israel and a high propensity for return migration. Perhaps mainly those who "do well" in Israel remain.

Harder to explain are the lower earnings of those who speak Arabic on a daily basis, even after controlling for country of origin. Jewish immigrants from the Middle East and North Africa who speak Arabic on a daily basis may be less well skilled in Hebrew and less integrated into the mainstream Israeli economy than those who do not speak Arabic, other things being the same.

The analysis of literacy indicates that those who can write a letter in Hebrew earn about 12 percent more than those who cannot (Table 10.10, column 1). When Hebrew speaking and writing are both included in the analysis, each is statistically significant (Table 10.10, columns 2 and 3). Of particular note is Table 10.10, column (4). The benchmark is those for whom Hebrew is the only or primary language and who can write a simple letter in Hebrew. Those who can speak Hebrew but not write it have 8 percent lower earnings (YSPNWR). Those who do not speak Hebrew as an only or primary language but who can write Hebrew (NSPYWR) have 10 percent lower earnings. While those who neither speak nor write Hebrew (NSPNWR) have 20 percent lower earnings.

The analysis of earnings indicates that the skills of immigrants matter. Earnings increase with schooling and pre-immigration labor market experience. Post-immigration labor market experience and both speaking and writing Hebrew language skills are also important determinants of earnings. Earnings also vary by origin, with North African and Middle Eastern Jewish immigrants having the lowest earnings, perhaps because of the lower quality of schooling in their origins and their pre-immigration human capital was formed in economies at a much lower level of economic development than was Israel in 1972. Immigrants from the English-speaking developed countries have the highest earnings, in part because they speak the most important international language, come from highly developed economies with advanced school systems, and the high wages, political freedom and absence of persecution in their origin countries means that they have a high opportunity cost of staying in Israel. The high propensity for return migration and the high opportunity cost of staying is Israel may account for the high earnings of the immigrants from English-speaking countries that remain in Israel.

It is possible to estimate the rate of return on the investment in Hebrew language proficiency. It was found here that, at least in 1972, proficiency in Hebrew increased earnings by 20 percent, all other variables being the same. Suppose that this level of Hebrew proficiency can be obtained through a six-month full-time intensive Hebrew language ("ulpan") training program. The cost of this program is the forgone earnings plus the costs of the teachers, classroom, supplies, etc. For simplicity of exposition let us assume that these latter costs are also equal to six months' forgone earnings. If the total cost is the equivalent of a full year's potential earnings and if a long work life is assumed (and 30 years would be sufficiently long), the real social rate of return on this investment would be approximately 20 percent.[15] This would be a high rate of return on an investment in human capital. The rate of return would, of course, be lower if the immigrant is older (shorter remaining working life) or if the immigrant requires a longer or more expensive training period, and would be higher if the training costs were lower than what was used in this example.

Summary and conclusions

This study has used the 1972 Census of Israel to analyze the determinants of Hebrew speaking and writing proficiency among adult male Jewish immigrants. It also analyzes the effects of these skills on labor market earnings.

Hebrew speaking proficiency is measured by whether it is spoken on a daily basis as the only language or in conjunction with other languages, or not at all. Writing proficiency is measured by the ability to write a letter in Hebrew. The analysis demonstrates that the acquisition of these skills is consistent with the model of immigrant language acquisition developed for English-speaking destinations. In particular, Hebrew language proficiency among adult male immigrants is greater among those who: immigrated at a younger age, have been in Israel longer, and have more schooling. Hebrew skills are lower among those who married their current wife prior to immigration, and are greater among those with children, especially if they were born in Israel. Thus, skills and family structure matter.

Hebrew language skills are associated with where one lives in Israel. Those living in Tel Aviv are less likely to speak Hebrew, while those living in Jerusalem are more likely to speak and write Hebrew than immigrants living in the rest of the country. Those living in areas where many others use their origin language are less likely to speak Hebrew or to be able to write in Hebrew. That is, linguistic concentrations or enclaves retard Hebrew language acquisition.

Immigrants from North Africa and the Middle East show the greatest use of Hebrew, perhaps because it is linguistically close to Arabic and they have a low propensity for return migration. Those from Western Europe, and especially those from the developed English-speaking countries, have the least proficiency in Hebrew. The high propensity for return migration and the importance of English as an international language may discourage English language speakers from investing in the Hebrew language.

Hebrew language skills influence earnings. Those who speak Hebrew as an only or primary language earn about 13 percent more per year than those who use it less frequently or not at all, while those who can write in Hebrew have a similar earnings advantage. Although they are related, Hebrew speaking and writing proficiency have separate effects and both matter. Those who possess both skills earn about 20 percent more than those who lack both skills.

Hebrew is not the only language that effects earnings. Immigrants who speak English have an earnings advantage of about 16 percent, while those who speak Arabic have a 2 percent earnings disadvantage, other variables being the same.

The similarity of the findings with other immigrant receiving countries suggests that the underlying processes determining Hebrew language usage in Israel and its effect on earnings are the same as elsewhere. The large effects on earnings of Hebrew language skills indicates its importance in the labor market and for the successful absorption of immigrants.

Acknowledgements

This chapter is an outgrowth of Barry Chiswick's collaborative research with Paul W. Miller on the Economics of Language. This chapter was presented at the Conference on the Economics of Judaism and Jewish Human Capital, co-sponsored by the Departments of Economics at the University of Illinois at Chicago and Bar-Ilan University (Ramat Gan, Israel), held in Chicago, June 2000. Comments from Carmel U. Chiswick, Evelyn Lehrer, and Paul W. Miller on an earlier draft were very helpful. An earlier version was published in Sergio DellaPergola and Judith Evans (eds), *Papers in Jewish Demography, 1997*, Jerusalem: Hebrew University, Jewish Population Studies 29 (2000). We appreciate the assistance of Michael Beenstock, Department of Economics, Hebrew University, and of Michal Peleg, Director, and Natasha Volchkima, both of the Social Science Data Archive, Hebrew University, in making available the 1972 Census of Israel.

Notes

1 Few survey data sets on immigrants include information on literacy, that is, reading and/or writing skills. Studies that have examined immigrant literacy in the destination language include Beenstock (1993) for Israel, Rivera-Batiz (1996), Chiswick (1991), Chiswick and Miller (1998b and 1999) for the United States, Kee (1993) for the Netherlands and Dustmann (1994) for Germany. Compared to the census, these studies tend to analyze relatively small samples of selected foreign-born populations.

2 The Israeli Census is similar to censuses in other major immigrant receiving countries in not asking about specific pre-immigration experiences, other than country of birth.

3 For an interesting newspaper article on children serving as translators for their immigrant parents in the United States, see Hedges (2000).

4 Unlike the case of English among immigrants in the English-speaking developed countries, immigrants to Israel with a higher level of schooling are not more likely than their less educated counterparts to have been exposed to Hebrew in their pre-immigration secular schooling.

5 The automatic granting of Israeli citizenship to Jewish immigrants at entry means that this is not a relevant variable for Israel.

6 These regions are Jerusalem, Northern, Haifa, Central, Tel Aviv, Southern, and two very small groups, Jewish localities in the occupied territories (West Bank and Gaza) and area not specified.

7 These languages in order of frequency are Arabic, Yiddish, German, Romanian, French, Spanish-Ladino, Polish, English, Hungarian, Persian, Russian, and Kurdish. The substantive findings are unchanged if CONPR is expanded to include second or tertiary languages spoken.

8 For example, although in the 1972 Census among immigrants in Israel 6 to 10 years 8 percent spoke only Hebrew and 52 percent spoke Hebrew as an only or primary language, eleven years later in the 1983 Census, the cohort in Israel 16 to 20 years reported 19 percent and 79 percent, respectively. This is very similar to the 21 percent and 77 percent speaking only Hebrew or Hebrew as the only or primary language, respectively, among immigrants in the country 16 to 20 years as reported in the 1972 Census. See Table 10.2 and Chiswick (1998).

9 As the mean of the dependent variable is 0.75, multiplying the logit coefficients by 0.188 gives a partial effect that can be compared to the OLS coefficients. The majority of the effects in the logit model are slightly stronger than in the OLS model.

10 The cross-tabulation of writing and speaking among adult foreign-born men, expressed in percentage, is:

Writing	*Speaking Hebrew*				
	Only	*Primary*	*Second or third*	*No Hebrew*	*Total*
Yes	23.2	47.9	8.5	3.1	82.7
No	1.3	4.8	5.2	6.1	17.3
Total	24.5	52.7	13.7	9.2	100.0

11 Because the construction of the language concentration variable (CONPR) involves assigning values of zero to all individuals in the "Only Hebrew" category, the language concentration variable cannot be included in the multinomial logit model.

12 Other studies of the labor market adjustment of immigrants in Israel using various census and survey data include Beenstock (1993, 1996b), Beenstock and Ben-Menachem (1997), Chiswick (1998), Eckstein and Shachar (1995), Eckstein and Weiss (1998), Friedberg (2000), Neuman (1998), and Raijman and Semyonov (1998).

13 The dichotomous variables ENGLOPS and ARABOPS are unity for those who speak English and Arabic, respectively, on a daily basis as their only, primary or secondary language.

14 The effect of country of origin on earnings among immigrants in Israel from the high wage, high return migration English-speaking developed countries is presumably reflected in the coefficient of the English-speaking country of origin variable.

15 If going from lacking proficiency to having proficiency increases annual earnings by 100b percent, and if k is the cost of the investment expressed in full-year potential earnings, the rate of return on the investment is approximately r=b/k. In this example, b=0.20, k=1.0 and the rate of return on the investment is approximately 20 percent. If the cost were nine months' potential earnings, the rate of return would be approximately r=0.20/.75=26.6, or approximately 27 percent.

References

Beenstock, Michael (1993) "Learning Hebrew and Finding a Job: Econometric Analysis of Immigrant Absorption in Israel," Discussion Paper 93.05, Falk Institute, Hebrew University, Jerusalem.

Beenstock, Michael (1996a) "The Acquisition of Language Skills by Immigrants: The Case of Hebrew in Israel," *International Migration*, 34 (1): 3–30.

Beenstock, Michael (1996b) "Failure to Absorb: Remigration by Immigrants to Israel," *International Migration Review*, 30(4): 950–978.

Beenstock, Michael and Yitzhak Ben-Menahem (1997) "The Labor Market Experience of CIS Immigrants to Israel: 1989–1994," *International Migration*, 35 (2): 187–224.

Blejer, Mario I. and Itzhak Goldberg (1980) "Return Migration-Expectations versus Reality: A Case Study of Western Immigrants in Israel," *Research in Population Economics*, 2: 443–449.

Chiswick, Barry R. (1978) "The Effect of Americanization on the Earnings of Foreign-Born Men," *Journal of Political Economy*, 86 (5): 897–922.

Chiswick, Barry R. (1991) "Speaking, Reading and Earnings Among Low-Skilled Immigrants," *Journal of Labor Economics*, 9 (2): 149–170.

Chiswick, Barry R. (1998) "Hebrew Language Usage: Determinants and Effects on Earnings Among Immigrants in Israel," *Journal of Population Economics*, 11 (2):253–271.

Chiswick, Barry R. and Paul W. Miller (1992) "Language in the Immigrant Labor Market," in Barry R. Chiswick (ed), *Immigration, Language and Ethnicity: Canada and the United States*, Washington: American Enterprise Institute.

Chiswick, Barry R. and Paul W. Miller (1995) "The Endogeneity Between Language and Earnings: International Analyses," *Journal of Labor Economics*, 13(2):245–287.

Chiswick, Barry R. and Paul W. Miller (1998a) "English Language Fluency Among Immigrants in the United States," *Research in Labor Economics*, 17: 151–200.

Chiswick, Barry R. and Paul W. Miller (1998b) "Language Skill Definition: A Study of Legalized Aliens," *International Migration Review*, 32 (4): 877–900.

Chiswick, Barry R. and Paul W. Miller (1999) "Language Skills and Earnings Among Legalized Aliens," *Journal of Population Economics*, 12 (1): 63–91.

Dustmann, Christian (1994) "Speaking Fluency, Writing Fluency and Earnings of Migrants," *Journal of Population Economics*, 7 (2): 133–156.

Eckstein, Zvi and Ron Shachar (1995) "On the Transition to Work of New Immigrants: Israel 1990–92," mimeo, Department of Economics, Tel Aviv University.

Eckstein, Zvi and Yoram Weiss (1998) "The Absorption of Highly Skilled Immigrants: Israel, 1990–1995," mimeo, Department of Economics, Tel Aviv University.

Friedberg, Rachel (2000) "You Can't Take it With You? Immigrant Assimilation and the Portability of Human Capital," *Journal of Labor Economics*, 18 (2): 221–251.

Hedges, Chris (2000) "Translating America for Parents and Family," *New York Times*, Monday, June 19, New York Report, p. A19 (Midwest Edition).

Kee, Peter (1993) *The Economic Status of Male Immigrants in the Netherlands*, Amsterdam: University of Amsterdam.

Long, Michael H. (1990) "Maturational Constraints on Language Development," *Studies in Second Language Acquisition*, 12 (3): 251–285.

Neuman, Shoshana (1998) "Immigration: The Israeli Case," mimeo, Department of Economics, Bar Ilan University, Ramat Gan, Israel.

Raijman, Rebeca and Moshe Semyonov (1998) "Best of Times, Worst of Times, and Occupational Mobility: the Case of Russian Immigrants in Israel," Working Paper 98–04, Research on Immigration and Integration in the Metropolis, Vancouver.

Rivera-Batiz, Francisco L. (1996) "English Language Proficiency, Quantitative Skills, and the Economic Progress of Immigrants," in Harriet Orcutt Duleep and Phanindra V. Wunnava, (eds), *Immigrants and Immigration Policy: Individual Skills, Family Ties, and Group Identities*, Greenwich, Conn. JAI Press: 55–77.

Part III

The interaction of language and earnings among immigrants

11 The endogeneity between language and earnings

International analyses[1]

I. Introduction

It has become recognized in recent years that an important aspect of the overall adjustment of immigrants in the labor market is their linguistic adjustment.[2] Linguistic adjustment refers to the process by which immigrants who are not fluent in the destination's dominant language improve their fluency. Linguistic adjustment influences labor market outcomes, such as earnings. Previous research studies in this area have used ad hoc models and have not adequately addressed the issue of the extent to which language skills and earnings are mutually determined for immigrants.

This chapter addresses these deficiencies in the literature. First, it develops a systematic model based on economic principles for analyzing the determinants of immigrant fluency in the dominant language. Second, various techniques are used to analyze the effect on earnings of dominant language fluency. Third, the analysis explicitly examines the endogeneity between language fluency and earnings, in particular, whether those who anticipate greater earnings if they were to become fluent are in fact more likely to become fluent.

The empirical analysis is performed in detail for Australia, using the data from the 1981 and 1986 Australian Censuses. These results are then examined in a comparative framework with the findings obtained elsewhere for the United States, Canada, and Israel (Chiswick and Miller 1992; Chiswick 1993). There are several advantages to analyzing the Australia data. Australia has a large proportion foreign-born in its population (21% in 1986), compared to the United States (9% in 1990). The immigrants to Australia come from a diverse set of countries that differs from the major source regions for the United States. Among the foreign-born in Australia in 1981, 37% were born in Britain and Ireland, 43% are from various other parts of Europe (primarily southern Europe), 12% from Asia and Africa, 4% from New Zealand, and 3% from the Western Hemisphere. Unlike Spanish in the United States (which accounts for over half of all immigrants who speak a language other than or in addition to English at home), there is no single non-English language that plays a dominant role in Australia. Italian, Greek, and

Chinese are the three most common immigrant languages. Illegal aliens, who constitute a population that tends to avoid census enumeration, are very few in number in Australia, and again, unlike those in the United States, they play a trivial role in the low-skilled immigrant labor market. Thus, the problems of census undercount that may be important among low-skilled immigrants, particularly Hispanics, in the U.S. Census are of lesser importance in Australia. Moreover, the Australian question on language fluency is comparable to the question used in the U.S. 1980 Census. Finally, the study of the 1981 and 1986 Censuses of Australia permits a comparative analysis over time that is a test of robustness.

Section II develops the model of the determinants of English language fluency. The data used to test the model, the Australian Censuses of 1981 and 1986, are described in Section III. The empirical analysis of the determinants of English language fluency among immigrants is presented in Section IV. In Section V the effect of language skills on earnings is studied using alternative econometric techniques to ascertain the magnitude of the effect and whether there is endogeneity between language and earnings. These analyses are highly instructive even though it is not possible to demonstrate causality. Comparative analyses of the findings for Australia with parallel analyses performed on census data for the United States (1980), Canada (1981), and Israel (1983) are reported in Section VI. This international comparative analysis demonstrates the robustness of the model. Section VII closes this chapter with a summary and conclusion.

II. Determinants of language fluency: the theory

Language skills are an important form of human capital. They satisfy the three basic requirements for human capital: they are embodied in the person; they are productive in the labor market and/or in consumption; and they are created at a sacrifice of time and out-of-pocket resources.

Skills in one's "mother tongue" are acquired when young. At this stage the investments are made largely by parents or caregivers. This is a time in the life cycle when the human mind is especially efficient in creating language capital (Harley 1986; Long 1990). Among school-age children, language capital is acquired when other forms of human capital (e.g., physical maturation, schooling) are being acquired. Thus, their acquisition of spoken language skills in the mother tongue seems almost effortless.

Among immigrants, however, the acquisition of language capital relevant for the destination can be very costly and certainly does not appear to be an effortless process when this language differs sharply from the mother tongue.[3] Dominant language fluency among immigrants can then be expressed as a function of three conceptual variables: economic incentives, exposure, and efficiency.

As with other forms of human capital, *economic incentives* can be expected to be an important determinant of language capital acquisition. The economic incentives arise in part from the increment in the market wage rate, a higher rate of employment, and the decrease in the cost of consumption (including search costs) associated with a higher level of fluency. This suggests an endogeneity between earnings and language skills.

The economic incentives for language acquisition are also related to the expected future duration in the destination. Immigrants expecting to return to their country of origin would have a shorter expected future duration in the destination and, therefore, less of an incentive to make language investments specific to the destination and more of an incentive to make investments that retard the depreciation of language-of-origin skills. Thus, other variables being the same, destination-language fluency would be lower for immigrants from countries where there is a higher incidence of return migration.

Exposure refers to the learning by doing and the formal instruction aspects of acquiring fluency in the destination language, as well as formal language training. It includes the extent to which others, whether in person or through the media, use the destination language in one's presence and the extent to which the person himself or herself utilizes it. Exposure can be thought of as having three components: exposure prior to immigration, time units of exposure in the destination, and the intensity of exposure per unit of time in the destination.

Exposure to the destination language prior to immigration is greater the smaller the "linguistic distance" between this language and the immigrant's mother tongue. The linguistic distance between Spanish and Italian, for example, is smaller than that between Spanish and Korean. Therefore, including language (or country) of origin in an analysis of fluency in the destination language measures, in part, the effects of linguistic distance.

For immigrants from multilingual countries of origin (e.g., India), a measure of the linguistic distance between the immigrant's mother tongue and the destination language may be misleading. Exposure to the destination language while still in the country of origin, whether it is used as a lingua franca (e.g., English in India) or because of the presence of foreign nationals (e.g., overseas U.S. military bases), would enhance destination language skills. Preimmigration exposure may also include instruction in school in the destination language. This would result in a positive effect of schooling on destination-language skills.

Duration in the destination, as measured by the number of years since migration, is also an index of exposure. Other variables being the same, destination-language fluency would be expected to increase with duration, up to the asymptote of full fluency.

The intensity of exposure per unit of time in the destination is smaller for those living in an environment in which more people communicate in the immigrant's mother tongue. Thus, the greater the extent to which a minority language is spoken in the area in which the immigrant lives, whether by immigrants or natives, the poorer will be the fluency in the destination language.[4]

Perhaps the most important language environment is in the home. Language skills emerge in large part through the linguistic interaction of those living together. Marriage to a spouse from the same linguistic origin will detract from destination-language exposure and thereby reduce fluency, compared to marriage to a native speaker of the destination language.[5] Children in the family, particularly those born in the destination, are more likely to be fluent in the dominant language because of the effects of age on language acquisition and their enrollment in school. Thus, to the extent that immigrant parents acquire fluency from their children, adult immigrant-language fluency is, on the one hand, enhanced by the presence of children. On the other hand, rather than serving as teachers, children may serve as parental interpreters to the world outside the immigrant-language enclave. If so, children would inhibit destination-language fluency on the part of adult immigrants. The role, if any, of children as interpreters is likely to be less important in the labor market than in the market for consumer goods and services.

An important aspect of exposure per unit of time to the destination language is formal instruction in this language (e.g., English as a Second Language programs). The determinants of enrollment and its consequences for destination-language fluency are clearly important issues. Unfortunately, the data available for the study of language fluency generally lack information on participation in language-training programs.

Efficiency refers to the extent to which a given amount of destination-language exposure produces language fluency. The very young have an impressive ability (efficiency) to acquire language skills, even in more than one language simultaneously. With age, however, this facility appears to diminish.[6]

Efficiency in language acquisition may be enhanced by a higher level of education. This may arise because the more educated have a greater mastery of their mother tongue and are more efficient in learning new concepts and new terminology.[7] Furthermore, those with schooling in the destination would be expected to be more fluent in the destination language as fluency may be a prerequisite for school enrollment and the destination schooling itself would enhance fluency.

Refugees may appear to have a lesser efficiency in acquiring dominant language skills than economic migrants from the same linguistic origin for two reasons (Chiswick 1978). First, refugees are less intensively self-selected for the characteristics that enhance a successful adjustment in the destination. Factors other than successful adjustment play a larger role, if not a dominant role, in their decision to migrate. Second, refugees could be expected to have invested in less preparation for the move, particularly if sudden changes in political events are responsible for their refugee status. But to the extent that refugees have a lower probability of return migration, they would make greater investments specific to the destination. Ceteris paribus, this implies a lower initial level of fluency and a steeper rise in language fluency with duration among refugees than among economic migrants.

This analysis suggests the following conceptual equation:

LANG = *f* (economic incentives, exposure, efficiency),

where LANG is a measure of the immigrant's fluency in the dominant language. The empirical counterpart of this conceptual equation is:

LANG = *f* [expected wage increment for language fluency (+),
expected future duration (+), duration in destination (+),
married to native of destination (?),
married to native of origin (–), children (?),
minority language concentration (–), destination language instruction (+), linguistic distance (–),
age at migration (–), education (+), refugee status (–)],

where the expected partial effect (positive or negative) is indicated in parentheses.

Not all of these variables can be measured on an individual basis. There are no data on individual variations in expected increments in wages due to improved language fluency. It is, therefore, not possible to estimate directly the partial effect of expected wage increases due to fluency on the acquisition of fluency.

One approach would be to use observed earnings as a proxy for the expected wage increment variable. Then equations for language fluency and earnings could, in principle, be estimated simultaneously. This requires identifying instruments for both the language and earnings equations. The limitations inherent in determining appropriate identifying instruments, particularly for identifying the language equation, preclude this approach. Alternative procedures, to be discussed below, are used to address the relationship between language fluency and earnings.

In addition, in the data to be studied explicit information is not available for refugee status, linguistic distance, and expected future duration. Dichotomous variables for country of birth are used as proxy measures for these variables. Data are also not available for destination-language instruction. Data files that include one or more of these variables are deficient in other, far more serious, dimensions.

III. The data

The detailed empirical analyses reported in Sections IV and V below use data for Australia and test the robustness of the model by comparing findings for the 1981 and 1986 Censuses of Australia. Further tests of robustness are presented in Section VI through comparative analyses with findings for the United States, Canada, and Israel. This section presents a detailed discussion of the data used in the Australian analysis. The data for the United States and Canada are discussed in Chiswick and Miller (1992) and for Israel in Chiswick (1993).

The empirical analysis uses the microdata files (1/100 sample) released from the 1981 and 1986 Australian Censuses of Population and Housing. For most immigrant labor market analyses, the 1981 Census is far superior to the 1986 Census. The 1981 Census reports income in 14 categories, rather than the 8 available in the 1986 data, and period of immigration in single years (up to an upper open-end interval of 35 or more years), rather than the 5 broad categories used in presentation of the 1986 data. An advantage of the 1986 Census, however, is that it includes information on any second language spoken in the home by the respondent and thereby permits the creation of a better minority language concentration index. Thus, the empirical analysis of language fluency in Australia uses both Censuses so as to exploit as fully as possible the available data and test for robustness.

The statistical analysis is limited to foreign-born males aged 25–64 who were employed at the time of the Census. The 1981 Census asked if the respondent spoke a language other than English at home, and, if yes, also asked for the degree of fluency in spoken English—"very well," "well," "not well," and "not at all." In the 1986 Census, the respondent was also asked if a language other than English was spoken at home. This time, however, if the response was "yes," in addition to the question on fluency in English, the respondent was asked to identify the other language. Twelve minority languages, representing 74% of immigrants reporting a minority language, are coded on the data file.

It is useful to collapse the four-category English language proficiency variable into two categories for the statistical analysis: those who speak only English or speak it very well versus all other groups. The Australian Bureau of Statistics cautions that the "not well" and "not at all" categories will underestimate the extent of English-language deficiency (Australian Bureau of Statistics 1982, p. 1). Furthermore, preliminary tests indicated that among immigrants from non-English-speaking countries there is no difference in earnings between those who reported they spoke only English at home and those who spoke another language but spoke English very well, in contrast to the other three groups for which earnings were significantly lower by about 12%.[8]

Hence, in the analysis reported here, immigrants fluent in English are defined as those who speak only English at home and those who also speak another language but speak English very well. Using this definition, the fluency rate is 76.0% for adult male immigrants and 58.2% for those from non-English-speaking countries.[9] The fluency rate varies sharply across birthplace regions. It is relatively low for southern Europe (39.9%), other Asian countries (i.e., excluding Vietnam and South Asia, 46.9%), Vietnam (16.7%) and South and Central America (38.8%) but much higher for northern Europe (excluding the British Isles, 87.3%) and South Asia (87.8%).

The model in Section II above suggests the importance of a minority-language concentration variable in the analysis of language fluency. Specific minority languages are not identified in the 1981 Census, but there is substantial detail on country of birth. Linguistic country groups were created by combining countries with the same language (e.g., combining Portugal, Brazil, and Timor). Then the proportion of the population aged 15–64 in the region in which the immigrant lives that is of the same minority linguistic-country group as the immigrant is assigned to the respondent. The countries in which English is the mother tongue or a major lingua franca are assigned a value of zero.

In the 1986 Census, the 12 most important minority languages are specifically identified. The minority-language concentration variable is defined as the proportion of the population aged 15–64 in the region in which the respondent lives that reports the same minority language. Those who speak only English and the minority languages with so few speakers that they are not separately identified are assigned a value of zero.

It should be noted that, while the dependent variable is the respondent's level of English-language fluency, the concentration measure is not the mean level of English fluency in the person's region. In one instance (1981) it is the proportion of immigrants from the same country group, and in the other (1986) the proportion of immigrants and natives who speak the same non-English language at home. This mitigates, if not avoids, the endogeneity problem raised by Case and Katz (1991).

The relevant characteristics of the Australian Census, the variables discussed above, and the other variables used in the statistical analysis are defined in detail in the Appendix, where the means and standard deviations of the variables for Australia are also reported.

IV. Analysis of language fluency: Australia

The results of alternative specifications of the model for English language fluency are reported in Table 11.1. The model includes the exposure and efficiency variables discussed above. In the absence of data on the increment in wages, the effect of economic incentives is analyzed below through an analysis of residuals. The substantive findings in the ordinary least squares (OLS) and logit analyses are virtually identical, so only the former are discussed explicitly.

The analysis in Table 11.1 indicates that English-language fluency is related to education (years of schooling), duration of residence, age (which measures the effect of age at migration when duration is held constant), current marital status, whether married overseas, number and age of children, size of place, and dichotomous variables for country of birth. The minority birthplace concentration measure is included in columns 2–5. Columns 1–3 are OLS regressions, while column 4 is a logit equation and column 5 is OLS excluding immigrants from the major English-speaking source countries (Britain, Ireland, Canada, United States, British West Indies, and New Zealand).

Table 11.1 Regression estimates of English-language fluency among adult foreign-born men, Australia, 1981 (dependent variable: LANG)

	Total sample				
Variable	*OLS* (1)	*OLS* (2)	*OLS* (3)	*Logit* (4)	*Non-English-speaking OLS* (5)
Constant	.814	.819	.475	5.766	.292
	(30.24)	(30.53)	(9.94)	(10.44)	(5.71)
Education	.025	.024	.031	.230	.036
	(16.35)	(16.11)	(12.09)	(13.35)	(15.23)
Age	−.005	−.005	.001	−.068	−.010
	(10.95)	(10.91)	(1.59)	(11.23)	(11.33)
Years since migration (YSM)	.009	.009	.028	.116	.017
	(17.24)	(17.79)	(11.68)	(16.98)	(18.33)
Married	−.011	−.009	−.004	−.104	−.011
	(.91)	(.75)	(.37)	(.74)	(.51)
Married overseas	−.037	−.039	−.048	−.410	−.087
	(3.42)	(3.59)	(4.40)	(3.59)	(4.71)
Child <6 years only	−.022	−.022	−.026	−.224	−.038
	(1.57)	(1.60)	(1.92)	(1.42)	(1.63)
Child 6–17 years only	−.007	−.007	−.020	−.068	−.015
	(.73)	(.80)	(2.12)	(.66)	(.96)
Children <6 and 6–17 years	.004	.002	−.012	.059	−.002
	(.33)	(.15)	(.91)	(.41)	(.11)
Small urban location	.032	.011	.010	.045	.008
	(3.01)	(1.08)	(.93)	(.28)	(.36)
Rural location	.026	.006	.001	−.055	.007
	(1.97)	(.43)	(.11)	(.30)	(.26)
Minority birthplace concentration	. . .	−.047	−.048	−.275	−.053
		(5.35)	(5.48)	(5.10)	(5.75)
Birthplace:					
Southern Europe	−.578	−.472	−.475	−7.068	. . .
	(53.02)	(20.26)	(20.36)	(15.02)	

Northern Europe	−.186	−.137	−.146	−5.304	.277
	(15.79)	(8.93)	(9.46)	(11.32)	(14.36)
Eastern Europe	−.405	−.388	−.374	−6.999	.035
	(17.43)	(16.52)	(16.10)	(14.82)	(1.12)
Arab countries	−.499	−.462	−.452	−6.742	.037
	(14.68)	(13.35)	(13.23)	(13.91)	(.98)
South Asia	−.145	−.137	−.139	−4.759	.326
	(7.39)	(7.05)	(7.37)	(9.67)	(10.41)
Philippines	−.127	−.119	−.098	−4.928	.386
	(1.75)	(1.64)	(1.31)	(6.35)	(4.91)
Vietnam	−.787	−.764	−.741	−8.325	−.232
	(16.94)	(16.40)	(16.23)	(13.15)	(4.79)
Other Asian countries	−.541	−.526	−.518	−7.414	−.057
	(14.45)	(14.05)	(13.95)	(14.95)	(1.37)
South and Central America	−.558	−.535	−.532	−6.953	−.013
	(10.49)	(10.07)	(10.12)	(13.04)	(.25)
Africa	−.197	−.170	−.173	−5.262	.296
	(7.55)	(6.50)	(6.66)	(10.71)	(8.91)
Remainder group	−.154	−.152	−.146	−4.801	.271
	(3.32)	(3.30)	(3.22)	(6.52)	(4.75)
Age*YSM/100	. . .	. . .	−.034	. . .	. . .
			(9.45)		
Education*YSM/100	. . .	. . .	−.044	. . .	. . .
			(3.20)		
N	7,288	7,288	7,288	7,288	4,166
R^2	.4472	.4512	.4568		.3520
McFadden's R^2				.5108	

Note: t-statistics are in parentheses and have been derived using White's (1980) heteroscedasticity-consistent covariance matrix estimator. See the Appendix for variable definitions. The prediction success rate for the logit model is 86.72%.

Source: 1981 Australian Census of Population and Housing, 1/100 Sample of the Foreign Born.

The estimates for the baseline specification presented in Table 11.1, column 1, reveal that each additional year of education is associated with an increase in the fluency rate of 2.5 percentage points, or about 3.6 percentage points for those from non-English-speaking countries. Educational attainment therefore has a pronounced effect on language skills: there is, for example, a difference of 12.5 percentage points between the language fluency rates of individuals with the mean level of education (of around 11 years) and those who hold bachelor degrees, other variables being the same. This rises to 18.0 percentage points for non-English-origin immigrants.

Age at immigration is also an important consideration, with language proficiency declining the greater the age at migration. For example, immigrants who arrived in Australia at age 25 are predicted to have English fluency rates 10 percentage points greater than immigrants who arrived at 45 years of age, ceteris paribus, but this is 20 percentage points among the sample of non-English-origin immigrants.

Language skills improve rapidly with years in Australia. Each additional year of residence is associated with almost a 1 percentage point improvement in the language fluency rate among all immigrants and a 1.7 percentage point increase for those from non-English-speaking countries. The differences in the language fluency rates across duration of residence categories are impressive. For example, the language fluency rate of immigrants who have been in Australia for the mean period of residence (= 18.5 years) would be around 17 percentage points higher than for the most recent arrivals, while it would be 31 percentage points higher among those from non-English-speaking countries.

There is a distinctive pattern by marital status. The language proficiency of individuals who married in Australia does not differ significantly from the rate of individuals who are single. The rate for individuals who married prior to migration, however, is around 4 percentage points (8.7 percentage points for those who are not from the major English-speaking source countries) lower than for the remainder of the group. Marriage prior to migration—in general, marriage to a person of the same origin country—appears to reduce opportunities to acquire fluency in English.

The other family-related variables in the equation are the presence and age of children. These variables are generally statistically insignificant, although they are at the margin of statistical significance in the Table 11.1, column 3, specification that includes limited interaction terms.[10] Moreover, the coefficients are typically negative, although they are less negative or more positive when there are at least two children, one preschool age and the other school age. More children may result in English language interaction among the children, and hence greater parental fluency.

What role, therefore, do children play in parental language attainment? One possibility is associated with the desire to preserve the language of origin, perhaps because of an expectation of return migration, to maintain a cultural/national identity, or to maintain ties with relatives in the origin. This means teaching the minority language to children at home or in school.[11] In his review of the position of various immigrant groups in Australia for the 1975 Australian Government Commission of Inquiry into Poverty, Cox (1975, p. 85) comments on Polish immigrants: "The resulting emphasis upon teaching Polish and utilizing it in the home had obvious implications upon the second generation. . . . It also affected the parents' degree of fluency in English." The negative coefficients could also indicate that children act as interpreters for their parents, thereby reducing the benefits from the development of dominant language fluency and thus reducing the incidence of this skill. While children can serve this role in the household, they can hardly be claimed to fulfill the same function in the workplace.[12]

The emergence of multiculturalism in Australia in the 1960s would have lowered the non-labor-market benefits from learning English. Prior to the 1960s, Australia's "immigrants were expected to assimilate largely unaided, that is, to embrace wholeheartedly the Australian way of life and deny and forget their origins. By the 1960s, . . . there developed a greater acceptance of the role of language and cultural maintenance in facilitating settlement. This integration model envisaged that immigrants would adapt to a core of Australian institutions and values while maintaining their cultural traditions" (Committee to Advise on Australia's Immigration Policies, 1987, p. 14). Consequently, immigrant parents would be more likely to speak their mother tongue at home with their children, and the positive effect of children on parental English fluency would diminish as a result.

Country of birth is another important determinant of English language proficiency. Compared to the benchmark—immigrants from the English-speaking countries—each of the birthplace coefficients is negative and statistically significant. Moreover, the estimated coefficients are generally quite large. In Table 11.1 column 1, among the Europeans, there is a clear distinction between those from northern Europe (language fluency 18.6 percentage points lower than the reference group), southern Europe (57.8 percentage points lower fluency), and eastern Europe (40.5 percentage points lower fluency). Immigrants from Arabic-speaking countries have a language fluency rate 50 percentage points lower than the English-speaking country group, other variables being the same.

The four Asian variables indicate that exposure to English prior to migration has an important effect on language attainment in Australia. Thus, for immigrants from the Philippines (a region of considerable U.S. colonial influence and military bases), the rate of language fluency is only 13 percentage points lower than for the benchmark group; for those from South Asia (a region of British colonial influence), it is 15 percentage points lower than for the benchmark; while for immigrants from Vietnam (nearly all post-1975 refugees) and other Asian countries, the rate of English-language proficiency is substantially lower, with the deficit being 78.7 and 54.1 percentage points, respectively. Immigrants from South and Central America also seem to have a substantial language deficiency compared to immigrants from

English-speaking countries (55.8 percentage points lower fluency). Thus, the English-language proficiency of immigrants appears to decrease with the linguistic distance of the mother tongue, increase with exposure to English in the origin country, and be greater among economic migrants than among refugees.

The minority-linguistic/birthplace concentration variable is added to the equation in Table 11.1, column 2. The inclusion of this variable is associated with moderate reductions in the effect of being of European origin, but with only minor changes in the estimated coefficients of the other birthplace dichotomous variables (cf. Table 11.1, cols. 1 and 2). The estimated coefficient of the concentration measure is negative (–0.047) and highly statistically significant (*t*-ratio = 5.35). The estimated effect suggests that an increase in the linguistic/birthplace composition of the area favorable to an immigrant by 1 percentage point would be associated with a reduction in the language fluency rate of immigrants of around 5 percentage points.[13] This seems to be a quite powerful effect. It is not reflecting subtle effects of birthplace not captured by the dichotomous variables. As will be shown below, the effect persists when the equations are estimated within major birthplace regions.[14]

Two interaction terms are added to the estimating equation in Table 11.1, column 3. The negative coefficient on the age-duration interaction variable indicates that the effect of duration on English-language fluency is weaker the older the age at migration. The negative coefficient on the education-duration interaction indicates that the effect of duration on fluency is weaker for the better educated than for the less well educated. In other words, the English proficiency gap by level of education diminishes with a longer period of residence.

English-language fluency models were estimated separately for the sample of those not born in the English-speaking source countries. This is shown in Table 11.1, column 5, for the non-English-speaking source countries as a group and in Table 11.2 for the major regions of origin. The effects of schooling, duration, age at immigration, and whether married overseas are all larger when those from English-speaking countries are deleted from the data.

For each of the major birthplace groups, language proficiency is positively related to educational attainment (Table 11.2). The estimated effect varies from around 2 percentage points higher fluency per year of education for the most fluent groups (northern Europe, South Asia) to 6–7 percentage points higher fluency per year of education for immigrants from Arabic-speaking countries and South and Central America, who have relatively lower levels of language attainment. Overall, the simple correlation coefficient between mean level of fluency and the partial effect of education is –0.68, and this is statistically significant at the 10% level.

Age at migration has a significant negative effect on language fluency among those from non-English-speaking countries overall and in seven of the eight major source regions (Table 11.2, col. 2). It has an insignificant positive coefficient only for Central and South America, but this is based on a very small sample (67 observations).

Table 11.2 Selected regression coefficients for English-language fluency by place of birth, adult foreign-born men, Australia, 1981

Birthplace	*% Fluent*	*Education*	*Age*	*YSM*	*YSM*^{2*}	*Minority concentration*	*Married overseas*	*N*
All non-English-speaking	58.16	.036	−.010	.017	. . .	−.053	−.087	4,166
		(15.23)	(11.33)	(18.33)		(5.75)	(4.71)	
Southern Europe	39.88	.031	−.016	.024	. . .	−.048	−.053	1,921
		(7.80)	(11.55)	(14.34)		(4.27)	(1.89)	
Northern Europe	87.29	.022	−.005	.008	. . .	−.043	−.182	850
			(5.09)	(3.71)	(5.23)	(1.48)	(4.01)	
Eastern Europe	63.45	.024	−.010	.035	−.048	.012	−.216	394
		(2.81)	(3.10)	(3.39)	(1.88)	(.10)	(2.95)	
Arab countries	45.51	.061	−.001	.015	. . .	.048	−.211	178
		(6.51)	(1.50)	(3.17)		(.77)	(2.63)	
South Asia	87.85	.025	−.001	.027	−.042	−.201	.128	288
		(4.06)	(.20)	(3.94)	(2.81)	(2.66)	(2.40)	
Other Asian countries	46.90	.042	−.012	.048	−.065	−.358	.125	145
		(4.58)	(2.47)	(3.67)	(1.98)	(2.72)	(1.13)	
Central and South America	38.81	.066	.007	.013	. . .	−.444	−609	67
		(3.44)	(1.23)	(1.96)		(1.43)	(4.29)	
Africa	81.60	.036	−.008	.008	. . .	−.121	−.023	212
		(3.42)	(2.75)	(2.59)		(2.67)	(.31)	

Note: Additional control variables are age, whether married, child <6 only, child 6–17 only, children <6 and 6–17, small urban, rural. Equations for Vietnam, Philippines, and the remainder group are not presented owing to small sample size: Philippines (mean fluency rate of 86.36, sample size = 22), Vietnam (mean fluency rate of 16.67, sample size = 48), and the remainder group (mean fluency rate of 90.24, sample size = 41). See the Appendix for variable definitions. *t*-statistics are in parentheses and were derived using White's (1980) heteroscedasticity-consistent covariance matrix estimator.
* Variable was divided by 100.

Source: 1981 Australian Census of Population and Housing, 1/100 sample of the Foreign Born.

Table 11.3 Regression estimates of English-language fluency among adult foreign-born men, Australia, 1986 (dependent variable: LANG)

Variable	*Total sample*				
	OLS (1)	*OLS* (2)	*OLS* (3)	*Logit* (4)	*Non-English-speaking OLS* (5)
Constant	.781	.779	.592	4.110	.242
	(28.17)	(28.77)	(10.29)	(9.67)	(5.09)
Education	.023	.022	.024	.205	.033
	(15.69)	(14.66)	(7.98)	(13.02)	(14.97)
Age	−.005	−.005	−.001	−.054	−.009
	(11.55)	(10.40)	(.64)	(10.68)	(11.12)
Years since migration (YSM)	.009	.008	.018	.108	.017
	(20.48)	(21.20)	(7.24)	(20.23)	(22.51)
Married	−.011	−.009	−.009	−.130	−.018
	(.97)	(.88)	(.86)	(1.06)	(.95)
Child <6 years only	−.001	.003	.003	.117	.023
	(.03)	(.17)	(.18)	(.69)	(.90)
Child 6–17 years only	−.014	−.011	−.016	−.055	.023
	(1.56)	(1.25)	(1.81)	(.58)	(.90)
Children <6 and 6–17 years	−.001	.001	−.003	.017	−.016
	(.01)	(.02)	(.24)	(.12)	(1.03)
Small urban location	.011	−.012	−.013	−.119	−.016
	(1.08)	(1.22)	(1.26)	(.81)	(.74)
Rural location	.042	.015	.014	.298	.027
	(3.51)	(1.26)	(1.22)	(1.69)	(1.17)
Minority language concentration	. . .	−.075	−.074	−.362	−.070
		(13.38)	(13.26)	(11.73)	(12.60)
Birthplace:					
Southern Europe	−.527	−.389	−.390	−5.836	. . .
	(44.33)	(24.53)	(24.62)	(18.20)	
Northern Europe	−.151	−.129	−.128	−4.217	.241
	(11.15)	(9.76)	(9.66)	(12.51)	(12.42)

Other Europe	−.351	−.337	−.333	−5.474	.065
	(17.17)	(16.64)	(16.47)	(16.85)	(2.64)
Arab countries	−.457	−.393	−.391	−5.525	−.040
	(14.45)	(12.37)	(12.37)	(16.13)	(1.21)
Other Asian countries	−.285	−.250	−.250	−4.779	−.170
	(16.20)	(14.58)	(14.64)	(14.77)	(7.31)
Vietnam	−.831	−.753	−.736	−7.461	−.276
	(26.54)	(24.10)	(23.52)	(15.52)	(7.62)
South America	−.600	−.554	−.556	−6.117	−.095
	(12.42)	(11.56)	(11.59)	(15.37)	(2.00)
Other American countries	−.051	−.042	−.043	−2.155	.366
	(3.31)	(2.79)	(2.79)	(3.24)	(13.90)
Africa	−.116	−.084	−.087	−3.591	.343
	(5.05)	(3.67)	(3.80)	(9.62)	(11.59)
Age*YSM/100	. . .	. . .	−.016	. . .	. . .
			(1.19)		
Education*YSM/100	. . .	. . .	−.018	. . .	. . .
			(4.27)		
N	7,194	7,194	7,194	7,194	4,152
R^2	.3979	.4254	.4268		.3268

Notes: t-statistics are in parentheses and are derived using White's (1980) heteroscedasticity-consistent covariance matrix estimator. See the Appendix for variable definitions. OLS = ordinary least squares.

Source: 1986 Australian Census of Population and Housing, 1/100 Sample of the Foreign Born.

Language fluency improves with duration of residence for all birthplace groups, and the partial effect varies from around 1 percentage point per year of residence for the northern Europeans to 2.6 percentage points and 3.5 percentage points, respectively, for immigrants from eastern Europe and other Asian countries (evaluated at a duration of 10 years). Note the steep effect of duration on fluency for the eastern Europeans who are predominantly a refugee group. While those with lower fluency have a stronger effect of duration, the simple correlation coefficient (–0.43) is not significant at conventional levels. It is also noted that for three birthplace groups (eastern Europe, South Asia, and other Asian countries) there is evidence of a curvilinear relationship between language attainment and duration in Australia; fluency rises but at a diminishing rate with a longer residence.

Minority-language concentration is significant and negative overall and exists in four out of eight individual birthplace regions (Table 11.2). It is negative and insignificant in two cases and positive and insignificant for the remaining two birthplaces.

Finally, the foreign-marriage variable also performs satisfactorily within the disaggregated analysis. It is significant and negative both overall and in five out of the eight individual birthplaces.[15]

The 1986 Census provides the opportunity to construct a minority-language concentration variable that more closely matches the conceptual variable. Table 11.3 presents results from estimation of the language model using these data.[16] In spite of some minor changes in the definitions of a number of the other variables, comparison of Tables 11.1 and 11.3 reveals that the estimated effects are virtually identical.[17] For example, the partial effects of education in column 1 of Table 11.1 (1981 data) is 0.025, and that in column 1 of Table 11.3 (1986 data) is 0.023. Even though the age and duration of residence variables for the analysis of the 1986 Census data have been created from interval data, the partial effects are identical to those derived from the 1981 Census, where the data were provided in single years.[18] The similarity of the two sets of results attests to the robustness of the model.

Some caution is warranted when comparing the birthplace effects in the two analyses, owing to the somewhat different groupings of countries. However, for the southern Europe, Arabic, Vietnam, and Africa groups, the variables are the same. The estimated coefficients are again remarkably similar: southern Europe (–0.578 in 1981, –0.577 in 1986), Arabic (–0.499 in 1981, –0.457 in 1986), Vietnam (–0.787 in 1981, –0.831 in 1986), and Africa (–0.197 in 1981 and –0.116 in 1986). The model, therefore, appears to be quite robust.

The minority-language concentration measure is added in Table 11.3, column 2. This is negative (–0.075) and highly significant (t = 13.38). The estimated coefficient implies that in a region with 1% of the population speaking the same non-English language as the respondent, language proficiency will be a massive 7.5 percentage points lower than in a region in which none speak the language.[19]

Inclusion of the appropriately defined minority-language concentration variable (cf. Table 11.3, cols. 1 and 2) in the estimating equation results in slightly greater changes to the birthplace effects than recorded in Table 11.1. For example, the partial effect of a southern European origin declines by 13.8 percentage points, from –0.527 to –0.389, that for a Vietnamese origin from –0.831 to –0.753, and that for a South American origin from –0.600 to –0.554. It should be noted that Italian, Greek, Vietnamese, and Spanish are among the 12 languages used in the construction of the minority-language concentration variable. Yet the coefficients still indicate that language distance, exposure to English, and refugee status appear to explain the pattern of birthplace coefficients.

Table 11.3, column 5, reports the estimated equation for immigrants from non-English-speaking countries using the 1986 Census. The equations were also estimated separately for each of the nine non-English-speaking origins identified in Table 11.3. In summary, overall and for *each* of the nine non-English origin regions, education and duration have a positive effect on English-language fluency while age at immigration and the minority language concentration variable have a negative effect.

Table 11.4 examines the interrelationships between the minority-language concentration variable and education, age, and period of residence. The negative influence of the concentration variable (col. 4) is greater for the less well educated, for immigrants who arrive as adults, and for immigrants who have been in Australia for only a short time. Thus, ceteris paribus, for the very immigrants who have the lowest levels of language skills, the recently arrived, less educated, and older immigrants, living in a minority language enclave has a greater retarding effect on their acquisition of English language fluency.

V. Analysis of earnings: Australia

The analysis of earnings is based on the standard human capital earnings function modified for immigrant adjustment (Chiswick 1978). The structural model of the determinants of earnings among adult immigrant men assumes that, aside from current marital status, household characteristics (such as whether married prior to or after immigration and the number and age structure of the children) do not directly enter the earnings function and that, with competitive labor markets, wage rates for workers of a given level of skill are invariant with the size of an ethnic/linguistic enclave. These are implicitly standard assumptions in the immigrant adjustment literature.

As a result, the earnings equation specifies that the natural logarithm of earnings is a function of education, total potential labor market experience, duration in the destination, marital status, Australian citizenship, size of place, and country of birth. It is hypothesized that, with the exception of the birthplace variables, all of these partial effects are positive. These variables are defined, and their means and standard deviations are reported in the Appendix.

Table 11.4 Selected regression coefficients for English fluency model with minority language concentration interaction terms, adult foreign-born men, Australia, 1986 (N = 7,194)

Variable	(1)	(2)	(3)	(4)
Education	.022	.015	.016	.016
	(14.66)	(10.09)	(10.58)	(10.39)
Years since migration (YSM)	.008	.009	.009	.007
	(21.20)	(21.19)	(21.19)	(16.55)
Age	−.005	−.004	−.003	−.003
	(10.40)	(10.10)	(8.12)	(6.95)
Minority language concentration (CONC)	−.075	−.153	−.057	−.131
	(13.38)	(10.67)	(2.12)	(4.83)
CONC*Education	. . .	.008	.006	.007
		(5.85)	(4.39)	(4.54)
CONC*Age	. . .	. . .	−.002	−.003
			(4.18)	(6.59)
CONC*YSM	. . .	. . .	. . .	.005
				(9.18)
R^2	.3979	.4301	.4324	.4428

Note: These are selected regression coefficients gained from adding interaction variables to the Table 11.3 equations. In addition to the variables listed, all other control variables used in Table 11.3 are included in these equations. See the Appendix for variable definitions. Partial derivatives [from col. (4)] evaluated at sample means are

$$\partial \text{LANG}/\partial \text{EDUC} = 0.016 + 0.007\ \text{CONC} = 0.020,$$
$$\partial \text{LANG}/\partial \text{YSM} = 0.007 + 0.005\ \text{CONC} = 0.010,$$
$$\partial \text{LANG}/\partial \text{Age} = -0.003 - 0.003\ \text{CONC} = -0.005,$$

and

$$\partial \text{LANG}/\partial \text{CONC} = -0.131 + 0.007\ \text{EDUC} - 0.003\ \text{Age} + 0.005\ \text{YSM} = -0.075.$$

Source: 1986 Australian Census of Population and Housing, 1/100 Sample of the Foreign Born.

The 1981 Australian Census information on earnings was collected and released in the Public Use Sample File in 14 broad brackets, and in the case of the 1986 Census the data were aggregated to only eight intervals in the Household Sample File (Section of State). While there are statistical techniques that may be used to accommodate the grouping of income data (see Stewart 1983), a previous application of the technique using the 1986 Census [Miller 1989]) reveals that there are few gains from doing so. Accordingly, the analyses in this section are based on a dependent variable formed from the midpoints of the income intervals and by using a value of 1.5 times the lower threshold for the open-ended upper limit. This is the procedure employed in previous research based on the 1981 Australia Census by Chiswick and Miller (1985) and Stromback (1986).

Previous research (Chiswick and Miller 1985) suggests that the curvilinear relationship between earnings and duration of residence found in earnings functions for immigrants in other immigrant-receiving countries (e.g., the United States, Canada, and Israel) is not evident in the Australian data. Hence, only a linear duration of residence term is included in the estimating equations.

The basic regression equation is reported in Table 11.5, column 1, for the adult foreign-born men in the 1981 Australian Census. The partial effects are all statistically significant, have the hypothesized signs, and are consistent with other studies. Table 11.5, columns 2–5, address the issue of the effect of English-language fluency on the earnings of immigrants in Australia. The observed dichotomous English-fluency variable is added in column 2 and replaced in column 3 by a predicted measure using an instrumental variables (IV) approach.[20] In columns 4–7, the analyses are done separately for those fluent in English and those not fluent, where columns 4 and 6 are OLS equations, and columns 5 and 7 are equations corrected for the potential selectivity bias in such a dichotomy of the data.

Reading across the columns in Table 11.5, it is apparent that earnings rise among immigrants in Australia by about 6% per year of schooling for immigrants as a whole, but the effect is larger (8%) for those fluent in English and smaller (about 2%) for those not fluent in English. Earnings increase, but at a decreasing rate with a rise in total potential labor market experience, where the increase in Table 11.5, column 1, is about 1.3% per year for the first year and 0.8% when evaluated at 10 years.

The earnings of immigrants increase by about 0.4% per year in the country, and this effect is highly statistically significant.[21] Among those not fluent in English there is no significant effect of duration. Note, however, that part of the effect of schooling, experience, and years since immigration is to raise the level of fluency of immigrants. Marital status also matters. Earnings are about 12% higher for those currently married, regardless of fluency in English.

Earnings vary systematically by size of place. In the full sample, earnings are about 6% lower in small urban areas and about 20% lower in rural areas compared to the large urban areas. Among those not fluent in English, however, living in a small urban area results in earnings about 15 percentage points lower than in the large urban area. The partial effects of the size of place variables may be reflecting equalizing wage differentials, unmeasured differences in the cost of living or in immigrant skill.

Australian citizenship, in contrast, provides no additional earnings, when other variables including country of origin are the same. This finding is consistent with previous research (see, e.g., Chiswick and Miller 1985) and the observation by the Committee to Advise on Australia's Immigration Policies (1988, p. 11) that "citizenship is of little material value."

It is premature to link the absence of a pecuniary return to citizenship to the low propensity among immigrants to naturalize. The mean citizenship rate in Australia (56%) exceeds that in the United States (48%), even though analyses for the United States indicate a 5 percent earnings premium

Table 11.5 Regression estimates of earnings equations, adult foreign-born men, Australia, 1981 (dependent variable: Natural logarithm of annual income)

Variable	*Total sample*			*Fluent in English*		*Not fluent in English*	
	OLS (1)	*OLS* (2)	*IV* (3)	*OLS* (4)	*Selectivity corrected* (5)	*OLS* (6)	*Selectivity corrected* (7)
Constant	8.625	8.586	8.805	8.390	8.383	8.832	8.787
	(138.21)	(136.17)	(55.29)	(108.20)	(132.51)	(71.24)	(79.31)
Education	.061	.060	.066	.076	.077	.020	.011
	(19.66)	(19.14)	(14.03)	(19.40)	(23.23)	(4.02)	(1.94)
Experience	.013	.013	.013	.020	.020	.008	.012
	(4.44)	(4.48)	(4.44)	(5.37)	(6.08)	(1.32)	(1.84)
Experience^2/100	.027	−.027	−.029	−.039	−.039	−.016	−.017
	(4.92)	(4.82)	(5.37)	(5.52)	(6.53)	(1.77)	(1.69)
Years since migration	.004	.003	.006	.003	.004	.003	−.004
	(4.00)	(3.34)	(2.88)	(3.21)	(3.61)	(1.12)	(1.28)
Small urban	−.065	−.066	−.057	−.053	−.053	−.138	−.159
	(2.92)	(3.00)	(2.69)	(2.41)	(2.46)	(1.67)	(2.85)
Rural	−.196	−.197	−.190	−.222	−.221	−.009	−.031
	(5.98)	(6.01)	(7.68)	(6.46)	(8.64)	(.09)	(.48)
Married	.122	.123	.115	.117	.116	.147	.172
	(6.37)	(6.43)	(6.17)	(5.95)	(6.12)	(2.44)	(3.99)
Citizen	−.011	−.012	−.005	−.008	−.008	−.017	−.011
	(.72)	(.80)	(.31)	(.47)	(.48)	(.52)	(.38)
Ireland	−.049	−.049	−.048	−.053	−.053	. . .	. . .
	(1.63)	(1.64)	(1.15)	(1.77)	(1.29)		
Canada	−.076	−.074	−.086	−.104	−.105	. . .	. . .
	(.87)	(.85)	(.91)	(1.18)	(1.13)		
United States	.018	.019	.011	−.032	−.033	. . .	. . .
	(.18)	(.20)	(.18)	(.33)	(.53)		
New Zealand	.049	.048	.055	.039	.040	. . .	. . .
	(1.41)	(1.37)	(1.62)	(1.12)	(1.20)		

British West Indies	−.100	−.100	−.099	−.135	−.135	. . .	. . .
	(.63)	(.62)	(.45)	(.87)	(.62)		
Southern European	−.211	−.180	−.353	−.177	−.197	. . .	. . .
	(12.10)	(8.36)	(2.95)	(7.32)	(4.90)		
Northern Europe	−.101	−.091	−.148	−.103	−.110	.159	.029
	(4.20)	(3.67)	(3.27)	(3.87)	(4.20)	(2.61)	(.41)
Eastern Europe	−.131	−.109	−.231	−.130	−.145	.156	.124
	(4.06)	(3.13)	(2.59)	(2.83)	(3.30)	(4.20)	(2.47)
Arab countries	−.275	−.248	−.398	−.234	−.252	−.084	−.119
	(6.65)	(5.70)	(3.59)	(3.18)	(3.75)	(1.98)	(2.08)
South Asia	−.055	−.047	−.092	−.078	−.084	.153	−.039
	(1.94)	(1.65)	(1.99)	(2.53)	(2.23)	(2.06)	(.34)
Philippines	−.190	−.183	−.222	−.240	−.246	.302	.120
	(2.15)	(2.07)	(1.87)	(2.37)	(2.00)	(2.12)	(.39)
Vietnam	−.557	−.515	−.749	−.268	−.295	−.267	−.228
	(3.95)	(3.66)	(4.20)	(3.46)	(1.53)	(1.55)	(2.57)
Other Asian countries	−.121	−.092	−.255	−.163	−.182	.227	.217
	(2.30)	(1.66)	(2.12)	(1.72)	(2.52)	(4.16)	(3.42)
South and Central America	−.207	−.177	−.344	−.254	−.273	.123	.103
	(3.24)	(2.67)	(2.59)	(1.73)	(2.50)	(2.72)	(1.22)
Africa	−.084	−.073	−.134	−.075	−.083	.094	−.053
	(2.56)	(2.21)	(2.33)	(2.03)	(1.86)	(1.79)	(.53)
Remainder group	−.048	−.039	−.087	−.018	−.024	−.068	−.251
	(.68)	(.56)	(.95)	(.24)	(.27)	(.60)	(.94)
LANG	. . .	.053	−.243	. . .	. . .	. . .	. . .
		(2.54)	(1.20)				
λ	. . .	. . .	. . .	. . .	.025	. . .	−.180
					(.62)		(2.84)
R^2	.1479	.1487		.1460	.1459	.0484	.0523
N	7,288	7,288	7,288	5,540	5,540	1,748	1,748

Note: t-statistics are in parentheses and are calculated using White's (1980) heteroscedasticity-consistent covariance matrix estimator. OLS = ordinary least squares; IV = instrumental variables. See the Appendix for variabe definitions.

Source: 1981 Australian Census of Population and Housing.

associated with citizenship (Chiswick and Miller 1992). Perhaps the other factors to which the Committee to Advise on Australia's Immigration Policies (1988) drew attention may be responsible. These include the low symbolic value of citizenship in Australia, the absence of links between citizenship and welfare entitlements, the absence of special privileges granted citizens for sponsoring relatives (as is the case in the United States and Canada), and the special privileges granted British citizens.

The estimated birthplace effects are measured relative to the earnings of British immigrants. They fall into two groups. The first group comprises immigrants from the other English-speaking countries, for whom the estimated effects are not statistically significant. This result is not surprising for immigrants from Ireland, Canada, the United States, and New Zealand.[22]

The estimated partial effects of birthplace for the second group, the non-English-speaking countries, are all negative, and, with the exception of the small and heterogeneous remainder group, they are all statistically significant. Moreover, the partial effects are essentially invariant with respect to whether an English-fluency variable is added to the equation. The differential is "small" (13% or smaller) for South Asia, Africa, other Asian countries, northern Europe, and eastern Europe. There is, however, a "large" differential (20% or more), other variables being the same, for those from the Philippines, South America, southern Europe, Arab countries, and Vietnam.[23]

The basic estimating equation is augmented with the measure of dominant-language fluency in Table 11.5, column 2. English-language facility is associated with a statistically significant 5.3% higher earnings (t = 2.54). For the sample of immigrants from non-English-speaking countries, the effect of language fluency in a specification similar to that in column 2 is 6.4 percent (t = 2.83). Analysis of the limited income data released in the 1986 Unit Record Sample File reveals a partial effect of English-language fluency of 8.3% (t = 4.75) for the total sample and 9.3% (t = 5.10) for the portion of the sample from non-English-speaking countries. It appears there has been a rise in the premium for English language skills over the 5 years.[24] In other words, dominant-language fluency is a skill that is rewarded in the Australian labor market, and its importance appears to have increased in the early 1980s.

A test was performed to determine whether being in a minority/ethnic enclave has an effect on the earnings of immigrants independent of the person's own level of fluency. The birthplace concentration measure used in the language analysis and an ethnic concentration measure were each found to be statistically insignificant.[25]

The analysis with the instrumental variables measure of fluency is presented in Table 11.5, column 3. The dominant language fluency effect is negative and statistically insignificant, the t being only 1.20. This method of estimation is sensitive to the choice of instruments, and there is often little to guide the choice of "good" instruments that will yield the minimum asymptotic variance. Comparison of the OLS and instrumental variables estimates listed in columns 2 and 3, respectively, reveals that a number of birthplace coefficients (in particular, those for southern Europe, Arab countries, Vietnam, other Asian countries, South and Central America) change considerably, indicating a pronounced widening of the inter-birthplace wage differential under the instrumental variables method.

An analysis of the limited income data available in the 1986 Census adds further insights. First, it suggests that the widening in the inter-birthplace wage differentials when the instrumental variables method is used to obtain the English language fluency variable (Table 11.5, col. 3) is due to the use of a minority concentration variable based on birthplace (1981 Census) as an identifying variable, rather than one based on the preferred measure of minority languages (1986 Census).

Second, in the 1986 Census analysis using instrumental variables, the dominant language fluency variable becomes positive (0.043), although it is not statistically significant (*t*-ratio = 0.52). Thus, as might be expected on the basis of econometric theory (see, e.g., Koutsoyiannis 1977), the instrumental variable results are unstable. This emphasizes the importance of using alternative data sets to test the effects of language on earnings to determine the pattern of bias introduced into OLS estimation by selectivity. This is done in Section VI below.

Analyses for the sample fluent in English are presented in Table 11.5, columns 4 and 5. As the selectivity correction factor (λ) is statistically insignifcant, the two sets of results are quite similar. They reveal that, among groups fluent in English, the partial effect of education on earnings is higher than reported for the pooled analysis (7.7% compared to 6.1%). Similarly, the partial effect of preimmigration experience is marginally higher for the group proficient in English than for the aggregate-level results (1.2% compared with 0.7% when evaluated at 10 years of experience). The effect of duration of residence in Australia is, however, the same for the various language-fluency groups under investigation.

The findings for the 24% of the sample lacking dominant language fluency are presented in Table 11.5, columns 6 and 7. The effect of education, other variables being the same, is quite low and indicates that each additional year of education is associated with only 1.1% higher earnings. This low partial effect, and the contrast with the 7.7% effect for the group fluent in English, may be indicative of a complementarity between the skills represented by formal education and language. A similar interpretation may apply to the relatively low effects of preimmigration experience (0.86 percentage points, which is marginally lower than the 1.2 percentage point effect per year of experience estimated for the group possessing English-language fluency).

The duration of residence variable becomes statistically insignificant in Table 11.5, columns 6 and 7. In other words, labor market experience in Australia does not attract an earnings premium if the individual is not fluent in English. This is consistent with one of the explanations generally offered for the positive relationship between earnings and duration of residence—acquiring skills relevant for and learning about the institutions of the labor market. This learning is presumably impeded by inadequate facility in English.

A final feature of the results in Table 11.5, column 7, is that the coefficient on the selectivity correction term (λ) is negative and statistically significant (t = 2.84). The negative coefficient implies positive selection into the non-dominant language fluency market.

To complete the study of the relation between language and earnings among immigrants, tests of endogeneity based on the selectivity corrected estimates were conducted. The first test was the conventional asymptotic t-test on the selection terms included in the Table 11.5, columns 5 and 7, results. These tests did not reject exogeneity in the English-fluent group, but they did reject exogeneity in the case of the group lacking English fluency. In view of these mixed findings, a second test was performed. Following Robinson (1988, 1989), the tests based on the selectivity correction terms were expressed in terms of an F-test of the joint significance of the two sample selection terms in an equation estimated on data pooled across those fluent and those not fluent in English. The computed F-statistic was 5.33, which exceeded the critical value of 3.0. Thus there is evidence that dominant language fluency is endogenous to the process of earnings determination among immigrants.

VI. International comparisons

This section compares the findings for Australia with parallel econometric analyses of census microdata for adult male immigrants (aged 25–64 years) in the United States (1980 Census, 1/100 sample), Canada (1981 Census, 1/50 sample) and Israel (1983 Census, 1/20 sample). This comparison demonstrates the robustness of the model.

The language question in the U.S. 1980 Census is closest to that used in Australia. The U.S. data are from the self-reported responses to the question, "How well does this person speak English?" (very well, well, not well, or not at all). Those who speak only English or report speaking it well or very well are treated as fluent in English in this analysis.

The Canadian language question is, "Can you speak English or French well enough to conduct a conversation?" The responses are English only, French only, both English and French, and neither English nor French. Those who responded they could carry on a conversation in either of the two official languages are considered fluent. The U.S. analysis was also computed using a definition closer to the Canadian concept, where all but those who speak English "not at all" are considered fluent.

In the Israel 1983 Census, respondents were asked, "What languages do you speak daily? Do not report a language that you know if you do not speak it daily." The response categories delineated a primary and, if one existed, a secondary language. Those who reported Hebrew as their only or primary language were counted as fluent.

Table 11.6 Partial effects of selected variables on dominant language fluency among adult male immigrants, Australia, United States, Canada, and Israel

	Australia		*United States*		*Canada*	*Israel*
Variable	1981	1986	1980	1980, *Canadian Definition*	1981	1983
Education	.024	.022	.027	.009	.006	.010
Age	−.005	−.005	−.004	−.001	−.001	−.006
Years since migration[a]	.009	.008	.014	.006	.005	.026
Married	−.009*	−.009*	.012	.009	−.001*	.024
Married overseas	−.039	. . .	−.035	−.012	−.013	−.083
Minority-language concentration	−.047	−.075	−.014	−.005	−.018	−.014
Proportion fluent[b]	.76	.77	.80	.95	.97	.84

Notes:
a Evaluated at 10 years.
b Proportion fluent in the dependent variable, based on country-specific language variables and definitions.
* Estimated effect not statistically significant at 5% level.

Sources: For Australia, see Table 11.1 (col. 2) and Table 11.3 (col. 2); for the United States/ Canada, see Chiswick and Miller (1992), Table 7; for Israel, see Chiswick (1993), Table 4.

While the detailed econometric analyses can be found elsewhere (Chiswick 1993; Chiswick and Miller 1992), Table 11.6 compares the analysis of the determinants of destination-language fluency for the United States, Canada, and Israel with the findings from above for Australia.[26] The most striking feature is the similarity of the results—the coefficients are statistically significant, have the same hypothesized signs, and have very similar magnitudes.

Education raises fluency by 1.0–2.4 percentage points per year of schooling, except for Canada, where the smaller effect seems to be the result of a more lax definition of fluency. Age at migration (age when years since migration is held constant) decreases fluency by 0.4–0.6 percentage points per year of age, again except for the Canadian definition of fluency. Duration raises fluency (when evaluated at 10 years) by 0.5 percentage points per year in Canada, 0.8 percentage points per year in Australia, 1.4 percentage points per year in the United States, and in the predominantly refugee population in Israel by 2.6 percentage points per year. Those who married prior to migration have significantly lower levels of fluency in each of the four countries.[27]

For all four countries the minority-language concentration measure is based on nondominant languages spoken in the person's region of residence. It is negative and significant in all instances, ranging from –0.014 (United States) to –0.075 (Australia, 1986).

The Australian analysis indicates important interaction effects between the concentration measure and education, age and duration (Table 11.4). Similar patterns emerge for the other countries (Table 11.7). Living among others who speak the same language of origin has a greater depressing effect on fluency among those with less education, who immigrated at an older age, and who have arrived more recently.

Table 11.7 Sign and significance of the minority-language concentration interaction variables in analyses of language fluency: Australia, United States, Canada, and Israel

	Australia	*United States*	*Canada*	*Israel*
CONC	–	–	–	–
EDUC*CONC	+	+	+	+
AGE*CONC	–	–	–	–
YSM*CONC	+	+	+	+

Note: +: positive and significant;—: negative and significant. Education, age, duration, marital status, children, location, and country of origin are controlled for. See the Appendix for variable definitions.

Sources: For Australia, see Table 11.4, col. 4, and unpublished table for 1986 Census; for the United States and Canada, see Chiswick and Miller (1992); for Israel, see Chiswick (1993).

The measures of the presence of children in the family differ among the countries. The findings also differ. Children are found to have no significant effect on adult male fluency in Australia or Canada,[28] but a significant positive effect for the United States, and a larger and very highly significant effect for Israel.[29]

The equations were also estimated separately for the major source regions of immigrants for the United States, Canada, and Israel, as was done for Australia. Although the analyses are not reported here, the basic patterns within major source regions for the effects of education, age, duration, whether married overseas, and the language concentration measure are the same. They tend to have the expected signs and are statistically significant.

Moreover, there are systematic differences in the degree of fluency by country of origin, when other variables are the same. Destination-language fluency varies inversely with the probability of return migration. Fluency rates are low, ceteris paribus, among Mexican immigrants in the United States and North American immigrants in Israel, two groups with very high rates of return migration. Fluency rates are lower among refugees than among economic migrants. For example, fluency is lower among Cuban, Vietnamese, and Chinese immigrants (who are disproportionately refugees) relative to South American and other Asian immigrants in the United States and Canada, and lower among Soviet refugees than among other European immigrants in Israel. Fluency rates are also lower the greater the linguistic distance of the immigrant's language of origin from the destination language. Indeed, this may explain the higher level of Hebrew-language fluency of North African Jews in Israel, whose country-of-origin language, Arabic, is linguistically closer to Hebrew than any of the other immigrant languages.

The four country analyses of the effects of language fluency on earnings (OLS and IV) are summarized in Table 11.8. In the OLS analysis the partial effect of the language variable is positive and statistically significant. The magnitude is lower in Australia (5% and 8% in the 1981 and 1986 Census data, respectively) than in Israel (11%), Canada (12%), and the United States (17%). The coefficient in the instrumental variables analysis shows considerable instability. It is positive and statistically significant in the United States and Israel, but the magnitudes are surprisingly large.

Table 11.8 Partial effect on earnings of proficiency in the dominant language, OLS and IV estimation, Australia, United States, Canada, and Israel

Method	*Australia*				
	1981	1986	*United States*, 1980	*Canada*, 1981	*Israel*, 1983
OLS	.053	.083	.169	.122	.110
	(2.54)	(4.75)	(12.52)	(2.43)	(12.66)
IV	−.243	.043	.571	.414	.351
	(1.20)	(.52)	(5.43)	(1.34)	(4.25)

Note: Schooling, experience, years since immigration, marital status, country of origin, region variables, and for the United States and Canada weeks worked are controlled for. *t*-ratios are in parentheses. OLS = ordinary least squares; IV = instrumental variables.

Sources: For Australia, see Table 11.5 and unpublished table for 1986 Census (available from us on request); for the United States and Canada, see Chiswick and Miller (1992), Tables 8 and 9; for Israel, see Chiswick (1993), Table 8.

The responsiveness of dominant language fluency to these earnings gains was examined in the analyses for the United States and Canada using the sample selectivity tests reported above for Australia (Table 11.9). For each country analysis, as with the Australian study, the assumption that the dominant language fluency variable in the wage equation is exogenous is rejected (Chiswick and Miller 1992).

Table 11.9 Significance of selectivity correction terms for analyses of language/fluency: Australia, United States, and Canada

Sample	*Australia*	*United States*	*Canada*
Fluent in dominant language	no	yes	yes
Not fluent in dominant language	yes	no	yes
Pooled sample	yes	yes	yes

Note: yes = significant; no = not significant.

Sources: For the United States and Canada, see Chiswick and Miller (1992); for Israel, see Chiswick (1993); for Australia, analyses are from unpublished tables from the 1981 and 1986 Censuses.

Further support for the robustness of the model is provided in the analysis by Dustmann (1994) of fluency and earnings among southern European and

Turkish immigrants in Germany using the 1984 wave of the German Socio-Economic Panel. The analysis for about 1,000 male immigrants is not strictly comparable to the analyses reported above because of differences in the questions asked in the German survey, the coding of variables, and the specification of the estimating equations. Nonetheless, the observed patterns are the same. A younger age at entry, a longer duration of residence, and a higher level of schooling are all associated with greater fluency in German among immigrants in Germany. Those who are married to a German wife are the most fluent, while those married to a non-German wife are less fluent than the single immigrants. School-age children have a marginally significant positive effect on the father's fluency. A minority-language concentration index could not be constructed with the German data.

By country of origin, ceteris paribus, those from Yugoslavia are the most fluent, followed by immigrants from Greece and Italy, then Spanish immigrants, and finally Turkish immigrants, who are the least fluent. Note that, among these five origins, Yugoslavia has had the greatest Germanic influence over the past century, including being part of the Austro-Hungarian Empire. Moreover, Turkish is the most distant linguistically from German, and those of Turkish origin are the most numerous of the immigrants.

Dustmann (1994) also analyzes the effect of speaking fluency on earnings in an OLS equation. He finds that adult male immigrants who report they speak German very well or well earn 6.3% more than those who speak it badly or not at all, while those who speak at an intermediate level earn only a marginally significant 3.8% more than the less fluent group.

VII. Summary and conclusion

Our study was concerned with the determinants of dominant-language fluency among immigrants and the labor market effect of this fluency. The analysis explored the endogeneity between dominant language fluency and earnings. The empirical tests were conducted primarily for Australia, with international comparisons to the United States, Canada, and Israel.

We developed a theoretical model of language fluency, where fluency is a rising function of economic incentives for acquiring fluency, exposure to the destination language, and efficiency in achieving destination-language proficiency. *Economic incentives* include the increment in annual earnings with greater fluency and the expected duration in the destination. *Exposure* includes the preimmigration experience with the destination language, duration in the destination (learning by living), and the intensity of exposure per unit of time (i.e., the extent of postmigration training in the destination language), and destination language usage in the area and the household in which the immigrant lives. *Efficiency* refers to the extent to which exposure produces language fluency and is related to the level of other skills (e.g., schooling) and biological/maturational factors (i.e., age). Empirical counterparts are developed for the theoretical variables.

The model was tested, and the parameters were estimated for foreign-born men aged 25–64 years using large microdata files from the Censuses of Australia (1981 and 1986), Canada (1981), Israel (1983), and the United States (1980). These international comparisons demonstrate the robustness of the model. The analyses were reported in detail for Australia using the 1981 and 1986 1/100 samples. These two data sources were used because of the somewhat different attractive features of the files and to test robustness.

The empirical analysis is very robust. In each of the four countries, the following patterns hold overall and when they are done separately by region of birth. Fluency in the destination language (English in Australia and the United States, English or French in Canada, and Hebrew in Israel) increases with exposure. It is greater the greater the use of the destination language in the country of origin, the longer the duration of residence, the smaller the proportion of people in the immigrant's area of residence who speak his mother tongue, and if the spouse does not have the same mother tongue. Fluency increases with efficiency in language acquisition; it increases with the level of schooling and decreases with age at immigration. Other variables being the same, fluency rates are lower the greater the probability of return migration, the greater the linguistic distance between the origin and destination languages, and among refugees than among economic migrants.

Living among others who speak the same nondestination language as the immigrant retards the acquisition of destination-language fluency in a manner that varies with other characteristics. A minority-language enclave has a greater depressing effect on destination-language fluency among the immigrants with lowest levels of fluency, that is, those recently arrived, the less well educated, and those who immigrate at an older age.

The analysis of earnings uses the now standard human capital earnings function model augmented for immigrant adjustment. In the Australian 1981 Census, English-language fluency is shown to be associated with a statistically significant 5.3% higher earnings, which increases to 6.4% for those from non-English-speaking countries. For the Australian 1986 Census the effects are larger, 8.3% and 9.3%, respectively, suggesting an increase over time in the returns to English-language skills. The effects of fluency in the destination language are even larger in the United States (16.9%), Canada (12.2%), and Israel (11.0%).

These estimates of the labor market benefits of language fluency can be used to estimate the rate of return on the investment if there are also cost estimates. We can use as a guide the Israeli "ulpan" system for new immigrants that usually involves 6 months of intensive Hebrew language training to bring adults to a modest level of fluency and literacy (reading and writing). Suppose fluency for adults costs the equivalent of 6 months to 1 year of full-time potential earnings. Then, overall, for permanent immigrants (i.e., assuming a long working life) estimated rates of return based on labor market earnings would vary from about 9%–18% for Australia, 11%–22% for

Israel, 12%–24% for Canada, and 17%–34% for the United States. The rate of return would, of course, vary by age; it would be higher for young adult permanent immigrants, as they learn language skills more rapidly and have a longer remaining work life, than for older immigrants. These estimates ignore the consumption benefits from acquiring fluency in the destination language. Investments in language fluency, therefore, appear to be very profitable for immigrants who are not fluent in the dominant language of the four destinations we have studied.

Comparisons of the coefficients on the schooling and experience variables in samples of those fluent and not fluent indicate the complementarity among skills. That is, the effects on earnings of schooling and labor market experience are much larger for those fluent in the destination language.

We implemented various procedures to test for the endogeneity of language skills. These tests indicate that those who anticipate higher earnings for unmeasured reasons if they were to become fluent are more likely to acquire destination language fluency.

Our model has been shown to be applicable for a wide range of source countries and immigrant receiving countries. The findings indicate that the determinants of destination language skills among immigrants can be studied using econometric techniques, that fluency responds to incentives (economic, exposure, and efficiency), that language skills have an important effect in the labor market, and, finally, that earnings and language fluency are determined jointly.

Appendix: Definitions of the variables and descriptive statistics

I. Analysis of 1981 Australian Census of Population and Housing

The Australian Bureau of Statistics released two public use samples from the 1981 Census of Population and Housing: the 1/100 Persons Sample File and the 1/100 Households Sample File. These differ in terms of the target population (persons vs. households) and also in the extent of geographic information provided and the degree of detail contained in the classification scheme used for some characteristics. All geographic references other than for a code showing major urban/other urban/rural region of residence were removed from the Household Sample File to ensure that the confidentiality of individuals is protected. In the case of the Persons Sample File, some data at the state level are included, but at the cost of a reduction in the number of categories used when classifying birthplace (40 compared to 101) and period of residence (only five broad intervals for valid codes). Some of the birthplace classifications in the Persons File are less useful for analysis of linguistic effects. For example, Indonesia is grouped with the Philippines; South Africa is coded with Egypt. Furthermore, information on the duration of marriage of males and the number of children is unavailable in the Persons File. For these reasons the Households File is used in this study.

A. Definitions

Population.—These are foreign-born men aged 25–64 who were employed as wage, salary, or self-employed workers (excluding unpaid helpers) at the time of the census. The analysis is restricted to individuals living in private dwellings and who were members of the primary family in such dwellings (i.e., all single-family noninstitutional households and the primary family in multifamily households).

Earnings (LNEARN).—Respondents were asked to report the gross income (including pensions and/or allowances) that they usually receive each week from all sources. The answer was given simply by placing a checkmark in a box corresponding to 14 weekly income categories. This was converted to a usual yearly income by the Australian Bureau of Statistics by multiplying the weekly figures by 52. Hence the data are standardized for weeks worked. For the open-ended upper-income interval (over $26,000 per year) a mean value of $39,000 is imposed.

Years of Education (EDUC).—This variable records the total years of full-time education. It has been created from the census "Age Left School" and "Qualifications" variables. Years of education is calculated as "Age Left School Minus 5." Individuals who stated a school-leaving age of 19 or more years were assigned 13 years of education. Individuals who possess a diploma have been assumed to have the equivalent of 15 years of full-time education, individuals who possess a bachelors degree the equivalent of 16 years of full-time education, individuals who possess a graduate diploma have been assumed to have the equivalent of 17 years of full-time education, and individuals who have a higher degree (master's, Ph.D.) have been coded as having 19 years of education.

Years of Experience (EXP).—This is computed as age minus years of education minus 5 (i.e., EXP = AGE – EDUC – 5). A quadratic specification is used.

Years since Migration (YSM).—For individuals with fewer than 35 years of residence in Australia, information on years since migration is available in single years. The open-ended category of 35 years or longer is assigned a value of 40 years.

Birthplace.—The following birthplace regions were formed from the 99 valid country codes available in the original data: Britain, Ireland, northern Europe, southern Europe, eastern Europe, Arab countries, Philippines, Vietnam, South Asia (which primarily comprises the regions of British influence), other Asian countries, Canada, the United States, British West Indies, South and Central America, Africa, New Zealand, Other. For the study of language proficiency, immigrants from Britain, Ireland, Canada, the United States, British West Indies, and New Zealand are used as the control group, whereas for the study of earnings, the omitted category is restricted to immigrants from Britain. The remainder group includes other countries and country not specified.

English Language Proficiency (LANG).—LANG is set to one for individuals who speak only English at home, or if a language other than English is spoken in the home, speak English "very well." The LANG variable is set to zero where a language other than English is spoken in the home and the respondent speaks English either "well," "not well," or "not at all." No information was collected in the 1981 Australian Census of Population and Housing on second languages spoken. Census pretests indicate that the Census language information will provide only a broad indication of the level of proficiency in English. In particular, the Australian Bureau of Statistics notes: "Testing prior to the census compared responses to the question with assessments of language proficiency for a sample of respondents interviewed for the purpose. These results showed that those who responded 'not well' or 'not at all' were correctly identified as prospective 'clients' for English language tuition. However, some who responded 'well' or 'very well' did not rate highly in the interview assessment. Responses to the 1981 Census question therefore may underestimate the number of people who were not proficient in English" (Australian Bureau of Statistics 1982, p. 1).

Citizenship (CITIZEN).—This is a dichotomous variable, set to one for immigrants whose country of citizenship is recorded as Australia and set to zero for individuals stating a non-Australia country of citizenship and also for those who are classified as stateless.

Marital Status (MARRIED).—This is a binary variable, defined to equal one for individuals who are married (spouse present) and defined to equal zero for all other marital states. Information on whether the individual is married is derived from the census question on marital status. Information on whether the spouse is present is derived from the family structure of the public use sample file.

Married Overseas (FORMAR).—This binary variable is constructed using information on the duration of the current marriage and duration of residence in Australia. Information on duration of marriage is derived from the family structure of the file. Individuals having a duration of current marriage in excess of their duration of residence in Australia are assumed to have married in the country of origin and are coded with FORMAR equal to one. The variable is set to zero for all other individuals.

Children.—Three variables were constructed from the family structure of the public use sample file to parallel the variables included in the analysis of the U.S. Census presented in Chiswick and Miller (1992). The first of these records whether one or more children aged less than 6 years were living in the family and there were no older children. The second records whether one or more children aged between 6 and 17 years inclusive were living in the family and there were no younger children. The third variable records the presence of children aged less than 6 years and between 6 and 17 years.

Location.—The only geographic information contained in the Households public use sample distinguishes individuals living in major urban areas, other urban areas, and rural areas. On this basis, two dichotomous variables were formed, the first for individuals living in "other urban areas" and the second for individuals living in "rural" areas.

Minority Language Concentration (CONC).—This variable is defined as the percentage of the respondent's linguistic-country group (see following definition) that resides in the same sized locality (metropolitan area, other urban area, and rural area) as the respondent. It is set equal to zero for the English-speaking majority group.

B. Linguistic-country groups

The following groupings are used in the study: English (the United Kingdom and Ireland, Canada, British West Indies, the United States, Bangladesh, Burma, India, Pakistan, Sri Lanka, Kenya, Malawi, Tanzania, Uganda, Zambia, Mauritius, Republic of South Africa, Christmas Islands, Cocos Islands, Oceania other than New Caledonia), Spanish (Spain, Argentina, Boliva, Colombia, Ecuador, Mexico, Paraguay, Peru, Uruguay, Venezuela), Portugese (Portugal, Brazil, Timor), Arabic (Bahrain, Jordan, Saudi Arabia, Kuwait, Muscat and Oman, Qatar, Yemen, Iraq, Lebanon, Syria, Egypt), Scandinavian (Norway, Sweden, Denmark, Finland), Chinese (China, Hong Kong, Taiwan, Singapore), German (Austria, Germany, Switzerland), French (Belgium, France, Switzerland, New Caledonia), Russian (Estonia, Latvia, Lithuania, Ukraine, the Soviet Union), Italian (Italy), Greek (Greece, Cyprus), Slavic (Albania, Bulgaria, Czechoslovakia), Hungarian (Hungary), Dutch (the Netherlands), Polish (Poland), Romanian (Romania), Indonesian (Indonesia), Persian (Iran), Hebrew (Israel), Japanese (Japan), Cambodian (Kampuchea), Korean (Korea), Laotian (Laos), Malaysian (Malaysia), Tagalog (Philippines), Thai (Thailand), Turkish (Turkey), Vietnamese (Vietnam), Maltese (Malta), Serbian (Yugoslavia).

Note.—All variables for Australia are dichotomous except earnings, education, total experience, duration in the destination, and the minority concentration measure.

C. Language question: 1981 Australian Census of Population and Housing

Q. 15 Does the person speak a language other than English at home?

- No, only speaks English ☐
- Yes ☐

How well does this person speak English?

- Very Well ☐
- Well ☐
- Not Well ☐
- Not at All ☐

II. Analysis of 1986 Australian Census of Population and Housing

The analyses of the 1986 Australian Census presented in this chapter are based on the 1986 Household Sample File (Section of State). Two data files were released from the 1986 Census. The Section of State Household Sample File identifies "major urban areas," "other urban areas," and "rural areas." The State/Territory file identifies seven states/territories and "major urban" and "balance of State/Territory." The Section of State file is comparable to the 1981 Census Public Use Sample and hence has been used in these analyses. Because of differences in the way in which primary information has been coded in this and the 1981 Census, it is necessary to change the definitions of some variables. The new definitions are listed below.

A. Definitions

Years since Migration (YSM).—The 1986 Census data are released in categorical form, and only five broad categories are distinguished: 0–4 years, 5–9 years, 10–14 years, 15–19 years, and 20 years and over. A "continuous" duration of residence variable is created using the midpoints of the closed intervals, and a value of 30 years for the open-ended upper interval.

Birthplace.—Eleven broad birthplace regions are recognized in the study. They are the United Kingdom and Ireland, southern Europe, northern Europe (the Netherlands, Germany), other Europe (all other countries of Europe, including the Soviet Union), Arab countries, Vietnam, other Asian countries, South America, Other American countries (Canada, the Caribbean, El Salvador, Mexico, the United States, other American countries), New Zealand, other Oceania. The remainder group includes other countries and country not specified.

Foreign Marriage (FORMAR).—In the study of the 1981 Census, the foreign marriage variable was constructed using information on duration of marriage and duration of residence in Australia. For most respondents this information was available in single years. However, as noted above, the duration of residence data in the 1986 Census are in very broad intervals. The duration of marriage information is coded into 5-year intervals in the 1986 sample file. This method of presenting the primary data prevents the construction of a usable foreign marriage variable.

Age.—The age data are presented in 5-year intervals: 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64. The midpoints of these intervals have been used to create a "continuous" age measure.

Minority Language Concentration (CONC).—Twelve minority languages are coded in the Household Sample File. These are Arabic/Lebanese, Chinese, Dutch, French, German, Greek, Italian, Maltese, Polish, Serbian and Croatian, Spanish, and Vietnamese. The minority-language concentration variable (CONC) is constructed from these data as follows: each respondent is assigned a value equal to the percentage of the population aged 15–64 in the region (defined broadly using information on location) in which he lives that reports the same second (minority) language as the respondent. The percentage representation in each language group is displayed in Appendix Table 11.A2.

B. Language question: 1986 Australian Census of Population and Housing

Q. 17 Does the person speak a language other than English at home?

☐No, speaks only English

☐Yes (Specify language)________

[Answer question 18 for each person who speaks a language other than English at home]

Q. 18. How well does this person speak English?

Very Well	☐
Well	☐
Not Well	☐
Not at All	☐

III. Descriptive statistics

Table 11A.1 Means and standard deviations of variables by regions of origin for adult foreign-born men, Australia, 1981

Variable	*Total sample* (N = 7,288)		*English-speaking* (N = 3,122)		*Non-English-speaking* (N = 4,166)	
	Mean	*Standard deviation*	*Mean*	*Standard deviation*	*Mean*	*Standard deviation*
Education	10.961	2.80	11.266	2.455	10.732	3.012
Age	41.808	10.221	41.405	10.429	42.110	10.054
Experience	25.847	11.146	25.138	11.166	26.378	11.103
YSM	18.578	9.790	17.008	9.948	19.755	9.503
Married	.839	.368	.819	.385	.854	.353
Married overseas	.306	.461	.358	.480	.266	.442
Child <6	.124	.330	.118	.323	.129	.336
Child 6–17	.343	.475	.328	.470	.355	.479
Children <6 and 6–17	.131	.338	.115	.319	.144	.351
Small urban	.113	.317	.149	.356	.087	.282
Rural	.076	.265	.098	.297	.060	.237
Minority language concentration	.650	1.891	.000	.000	1.138	2.388
Citizenship	.558	.497	.322	.467	.734	.442
Birthplace:						
Britain	.346	.476	.808	.394	. . .	
Ireland	.025	.156	.058	.234	. . .	
Canada	.005	.067	.011	.102	. . .	
United States	.011	.102	.025	.155	. . .	
West Indies	.001	.029	.029	.044	. . .	
New Zealand	.041	.199	.002	.296	. . .	
Southern Europe	.264	.441	. . .		.461	.499
Northern Europe	.117	.321	. . .		.204	.403
Eastern Europe	.054	.226	. . .		.095	.293
Arab countries	.024	.154	. . .		.043	.202
South Asia	.040	.195	. . .		.069	.254
Philippines	.003	.055	. . .		.005	.072
Vietnam	.007	.081	. . .		.012	.107
Other Asian countries	.020	.140	. . .		.035	.183
South America	.009	.095	. . .		.016	.126
Africa	.029	.168	. . .		.051	.220
Remainder	.006	.075	. . .		.010	.099
Earnings	14,855	7,808	16,347	8,426	13,737	7,108
Log (earnings)	9.475	.574	9.577	.538	9.399	.588
LANG	.760	.427	.998	.045	.582	.493

Note: The English-speaking regions include Britain, Ireland, Canada, United States, New Zealand, and the British West Indies. See above for variable definitions.

Source: 1981 Australian Census of Population and Housing, 1/100 Sample of the Foreign Born.

Table 11.A2 Percentage representation of major minority-language groups by locality, Australia, 1986

Language	*Metropolitan*	*Small Urban*	*Rural*
Arabic/Lebanese	1.06	.04	.02
Chinese	1.55	.33	.09
Dutch	.61	.43	.55
French	.43	.25	.30
German	1.03	.63	.61
Greek	2.80	.37	.28
Italian	4.24	1.27	1.70
Maltese	.67	.09	.22
Polish	.74	.18	.19
Serbian, Croatian	.67	.06	.23
Spanish	.72	.07	.16
Vietnamese	.70	.07	.02
Other	4.94	1.62	2.50

Source: 1986 Australian Census of Population and Housing.

Notes

1 We appreciate the comments received on an earlier version of this chapter from Michael Kidd and at seminars presented at Flinders University, Hebrew University, Queen's University, Tel Aviv University, and the University of Illinois at Chicago, as well as at the European Science Foundation Conference on Migration and Development, Davos, Switzerland, 1992, the International Regional Science Association Annual Meeting, Chicago, 1992, and the Australian Conference of Economists, Melbourne, 1992.

2 See, e.g., Cox (1975); McManus, Gould, and Welch (1983); Evans (1986); Stromback (1986); Chiswick and Miller (1988, 1992); Kossoudji (1988); Robinson (1988); Tainer (1988); and Chiswick (1991), among others. Research on this issue has used U.S., Canadian and Australian survey and census data. Most of the literature has analyzed fluency rather than literacy in the destination language because data on speaking skills are more abundant than data on reading/writing skills.

3 This explains, in part, the preference on the part of international migrants for destinations with the same or similar mother tongue. See, e.g., Chiswick and Miller (1994).

4 This approach is similar to the neighborhood effects model used by Case and Katz (1991) in their study of family and neighborhood peer effects on the behavior of inner-city youths in the United States.

5 This is consistent with Evans's (1986, p. 234) hypothesis that "members of large immigrant groups with more inward-looking friendship networks and more in-marriage will be less skilled in the host country's language."

6 For detailed analyses of this issue, including a survey of the literature, see Harley (1986) and Long (1990).

7 As noted above, those with higher levels of schooling acquired in the origin may also have been exposed to the destination language in school if this was part of the curriculum.

8 Using the 1981 Census data for immigrants from non-English-speaking countries,

controlling for education, experience, duration, location, and citizenship, the earnings differentials from the benchmark (speaking only English at home) are

Speak	*Coefficient*	*t-Ratio*
Very well	−.020	−.84
Well	−.125	−4.91
Not well	−.128	−3.97
Not at all	−.111	−.88

9 The English-speaking countries for this purpose and in the regression analysis include Britain, Ireland, Canada, the United States, British West Indies, and New Zealand.

10 Alternative specifications of the children variables, including analyses within birthplace groups, also result in negative, but generally statistically insignificant, coefficients.

11 In addition to limited opportunities in public schools, there exist private day schools and after-school and weekend programs for transmitting the country -of-origin culture and language to native-born children.

12 If children serve as interpreters in household and community matters, their presence would tend to depress their parents' English language fluency and would be associated with a depressing effect on earnings, other variables being the same. Tests suggest, however, that children have an insignificant effect on the earnings of adult male immigrants in Australia. An analysis of gender differences in Australia indicates that children have a less positive or more negative effect on the English language fluency of their mothers than their fathers (Chiswick and Miller 1995).

13 This interpretation is consistent with the linguistic environmental effect discussed in Section II above. In another context, Manski (1993) argues that such an effect may be difficult to separate from effects deriving from the characteristics of the constituents of the region. The definition of the concentration variable used in Table 11.1 (birthplace) makes identification of the separate effects difficult. This is less problematic for the concentration measure used for the 1986 Census (see Table 11.3 below) as it is based on the minority languages spoken by immigrant and native-born Australians.

14 The analysis reported in Chiswick and Miller (1996) shows that the minority-language concentration measure reflects the effects of foreign language media (newspapers, radio, TV) and the presence of other relatives in Australia and, hence, with formal ethnic networks. This research also demonstrates that the minority-language concentration measure is a reasonable proxy for these formal ethnic networks when more direct measures are not available. This allays concern over the specification used here.

15 The foreign marriage variable is positive and significant only in the case of South Asia. This finding is not unique to Australia. In comparable analyses for the United States and Canada, the foreign marriage variable for South Asia also had positive although not statistically significant effects on language fluency compared to those who were not married, although the effect was negative for all other source regions (Chiswick and Miller 1992). This seemingly South Asian effect is undoubtedly reflecting an unmeasured variable. One explanation is the much higher rate of arranged marriages among those from South Asia. More so than for other countries, other variables being the same, postmigration marriage may involve a wife with the same non-English mother tongue who has only recently arrived in the destination.

16 The mean values of the dependent variable LANG are similar in the two Census samples. It is equal to 0.760 in the 1981 data and 0.769 in the 1986 data.

17 See the Appendix for these differences in definitions.
18 In the Table 11.3, col. 3, specification the interaction between age and duration of residence is insignificant, whereas it was highly significant in Table 11.1. This may be attributable to the fact that both variables are available only in interval form in the 1986 data.
19 The estimated partial effect, –0.075, is four times the size of that estimated for Canada and five times that estimated for the United States (Chiswick and Miller 1992). There are two factors that could be responsible for this stronger effect in Australia. First, Australian immigration is more intensely characterized by "waves" of immigrants from specific source countries than immigration in North America. This would provide a strong basis for language maintenance. Second, it is possible that the stronger measured effect of the language environmental factor reflects the greater emphasis on multiculturalism in Australia. That is, the "ethnics" may be more ethnic in Australia than elsewhere. This is consistent with one of the arguments advanced earlier concerning the (marginally) negative effect of children on their father's dominant language fluency.
20 The instruments are all of the explanatory variables in the language and earnings equations. The identifying instruments are whether married overseas, number and age of children, and the birthplace concentration variable. The predicted value has a much smaller standard deviation than the observed value, and this results in a much larger standard error of the estimate.
21 This effect of duration in Australia on earnings is smaller than what is found in the United States and Canada but consistent with other studies of Australia (see Chiswick and Miller 1985, 1988, 1992).
22 For the immigrants from the British West Indies, however, the finding contrasts with the evidence from Britain and North America (Chiswick 1980; Chiswick and Miller 1992), which indicates significantly lower earnings for (Black) West Indian immigrants. However, the sample of West Indian immigrants in Australia is very small (only six observations), and given the greater cost involved in migrating to Australia compared with the United States, Canada, and the United Kingdom, the West Indian immigrants in Australia are also likely to be a highly select group. Furthermore, the racial composition of these immigrants is not known.
23 Among the Vietnamese, the earnings differential is quite marked; the estimated coefficient of –0.557 implies an earnings differential of about 43 percentage points. Nearly all of the Vietnamese, however, were in Australia for less than 5 years, and this effect may not have been fully controlled by the linear duration of residence variable. Refugees would be expected to have particularly low earnings during the initial period of adjustment.
24 This difference in the returns to language fluency is not due to the different grouping of the income data. Repeating the analysis of the 1981 Census data after deriving the dependent variable from 8 rather than 14 intervals results in only a slight increase in the estimated effect on the dominant language fluency variable, from 0.053 (t = 2.54) to 0.057 (t = 3.40).
25 The ethnic concentration measure was based on father's country of birth. Birthplace regions were combined into 26 ethnic groups, e.g., Germanic (Austria and Germany), Arabic (Lebanon and other Middle Eastern countries), Spanish (Spain and Latin America), French, and Italian. The proportions of the population in each region with fathers of Germanic, Arabic, Spanish origins, etc., were computed, though in cases where the specific country was not identified (e.g., other European countries) the ethnic concentration index was assumed to be zero. The value of the ethnic concentration index was based on the respondent's birthplace and region in Australia.
26 Beenstock (1993) uses panel data from the 1970s on immigrants in Israel for up to 3 years to analyze Hebrew fluency and literacy, as well as employment,

unemployment, and occupational convergence. His data lack information on wages or earnings. Beenstock's findings are consistent with those reported here for Israel.

27 The language acquisition model offered no hypothesis for those who married after migration compared to the unmarried. The partial effect of marriage after immigration compared to those who are not married is found to vary in sign and significance across the four countries.

28 This finding for Canada is based on an analysis of the 1/100 Household and Family File of the 1981 Census of Canada discussed in Chiswick and Miller (1992, p. 246).

29 Perhaps this is confirmation of the Israeli myth that immigrant parents learn the nation's language (Hebrew) from their children.

References

Australian Bureau of Statistics. *Census 81-Language*. Catalogue No. 2152.0. Canberra: Australian Bureau of Statistics, 1982.

Beenstock, Michael. "Learning Hebrew and Finding a Job: Econometric Analysis of Immigrant Absorption in Israel." Paper presented at the Conference on the "Economics of International Migration: Econometric Evidence," Konstanz, Germany, February 1993.

Case, Anne C., and Katz, Lawrence F. "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths." Photocopied. Discussion Paper no. 1555. Cambridge, MA.: Harvard Institute of Economic Research, 1991.

Chiswick, Barry R. "The Effect of Americanization on the Earnings of Foreign-born Men." *Journal of Political Economy* 86 (October 1978): 897–921.

—— "The Earnings of White and Coloured Male Immigrants in Britain." *Economica* 47 (February 1980): 81–87.

—— "Speaking, Reading and Earnings among Low Skilled Immigrants." *Journal of Labor Economics* 9, no. 2 (April 1991): 149–70.

—— "Hebrew Language Usage: Determinants and Effects on Earnings among Immigrants in Israel." Paper presented at the Conference on "Immigrant Absorption: The Interface Between Research and Policy Making," Technion, Haifa, Israel, June 1993.

Chiswick, Barry R., and Miller, Paul W. "Immigrant Generation and Income in Australia." *Economic Record* 61 (1985): 540–53.

—— "Earnings in Canada: The Roles of Immigrant Generation, French Ethnicity, and Language." In *Research in Population Economics*, ed. T. Paul Schultz, pp. 183–228. Greenwich, CT: JAI, 1988.

—— "Language in the Labor Market: The Immigrant Experience in Canada and the United States." In *Immigration, Language and Ethnic Issues: Canada and the United States*, ed. Barry R. Chiswick, pp. 229–96. Washington DC: American Enterprise Institute, 1992.

—— "Language Choice among Immigrants in a Multilingual Destination." *Journal of Population Economics* 7, no. 2 (1994): 119–31.

—— "Language and Labor Supply: The Role of Gender among Immigrants in Australia." *Research in Income Inequality* 5 (1995): 153–89.

—— "Ethnic Networks and Language Proficiency among Immigrants." *Journal of Population Economics* 9, no. 1 (March 1996): 19–35.

Committee to Advise on Australia's Immigration Policies. *Understanding Immigration*. Canberra: Australian Government Printing Service, 1987.

—— *Immigration: A Commitment to Australia*. Canberra: Australian Government Printing Service, 1988.

Cox, David. "The Role of Ethnic Groups in Migrant Welfare." In *Commission of Inquiry into Poverty: Welfare of Migrants*, ed. Ronald F. Henderson, pp. 1–150. Canberra: Australian Government Publishing Service, 1975.

Dustmann, Christian. "Speaking Fluency, Writing Fluency and Earnings of Migrants." *Journal of Population Economics* 7, no. 2 (1994): 133–56.

Evans, M. D. R. "Sources of Immigrants' Language Proficiency: Australian Results with Comparisons to the Federal Republic of Germany and the United States of America." *European Sociological Review* 2, no. 3 (1986): 226–36.

Harley, Brigit. *Age in Second Language Acquisition*. Avon: Multilingual Matters, 1986.

Kossoudji, Sherrie A. "English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men." *Journal of Labor Economics* 6, no. 2 (April 1988): 205–28.

Koutsoyiannis, A. *Theory of Econometrics*. 2d ed. Hong Kong: Macmillan, 1977.

Long, Michael H. "Maturational Constraints on Language Development." *Studies in Second Language Acquisition* 12, no. 3 (September 1990): 251–85.

Manski, Charles F. "Identification Problems in the Social Sciences." *Sociological Methods* 23 (1993): 1–56.

McManus, Walter; Gould, William; and Welch, Finis. "Earnings of Hispanic Men: The Role of English Language Proficiency." *Journal of Labor Economics* 1, no. 2 (April 1983): 101–30.

Miller, Paul W. "An Economic Analysis of Citizenship in Australia." Photocopied. Nedlands: University of Western Australia, Department of Economics, 1989.

Robinson, Chris. "Language Choice: The Distribution of Language Skills and Earnings in a Dual Language Economy." *Research in Labor Economics* 9 (1988): 53–90.

—— "The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Models." *Journal of Political Economy* 97 (1989): 639–67.

Stewart, Mark B. "On Least Squares Estimation When the Dependent Variable Is Grouped." *Review of Economic Studies* 50 (1983): 737–53.

Stromback, T. "The Earnings of Migrants in Australia." Conference Paper no. 46. Canberra: Bureau of Labour Market Research, 1986.

Tainer, Evelina. "English Language Proficiency and Earnings among Foreign-born Men." *Journal of Human Resources* 23, no. 1 (Winter 1988): 108–22.

White, H. "A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (May 1980): 817–35.

12 Speaking, reading, and earnings among low-skilled immigrants[1]

The growing literature on the economic adjustment or economic assimilation of immigrants has focused on the human capital that is embodied in them, the relevance of this human capital to the destination labor market, and postmigration human capital investment. One important aspect of human capital is "language capital," that is, the speaking, reading, and writing skills in one or more languages.

Language capital, particularly spoken language, is partially developed during the course of a child's maturation, for example, the development of speaking fluency in one's "mother tongue." Important investments are made in school and elsewhere in developing further one's language capital in the mother tongue. For most immigrants, however, their mother tongue is not the majority or dominant language spoken in the destination. An immigrant who does not know the dominant language might find a language-minority enclave within which mother-tongue skills can be fruitfully used. A language-minority enclave may, however, limit training opportunities and job mobility, whether it is geographic, occupational, or employer mobility, and thereby limit earnings opportunities. Furthermore, greater dominant-language skills would enhance productivity in the enclave and the nonenclave labor market by increasing efficiency in job search and through greater productivity on the job. There is, therefore, a labor market incentive to acquire dominant-language skills. Whether, and under what circumstances, this incentive is worth the cost is of keen interest.

This chapter is concerned with both the determinants of fluency in dominant-language skills and how these skills are translated into labor market earnings. A unique data set, a sample of illegal aliens apprehended in the Los Angeles area, is used to study the issue.[2]

Section I briefly reviews the literature on the nexus between language and earnings in the labor market for immigrants. It indicates the strengths and limitations of this literature. The data used for this study are described in Section II. Section III is a multiple regression analysis of the determinants of fluency in speaking and reading English. This includes longitudinal changes in speaking skills. Section IV is a regression analysis of the determinants of earnings focusing on the roles of fluency in speaking and reading English.

The chapter closes (Section V) with a summary and conclusion, including suggestions for the collection of data on immigrant populations.

I. Language and earnings

Ever since its recent development, the literature on the economic status of immigrants has been concerned with the "Americanization" or adjustment of immigrants (Chiswick 1978). One of the important interpretations of the variable for duration in the destination has been the acquisition of destination-specific skills, including labor market information and language skills. The earliest research, using the 1970 Census of Population, was limited by the absence of data on language skills, except for what could be inferred from country of birth.[3]

Substantial progress on the role of language in immigrant adjustment could not be made until the 1976 Survey of Income and Education (SIE) became available. The SIE asked a battery of questions about languages spoken and the use of these languages.[4] The 1980 Census furthered research on language and earnings by including a self-reported question on fluency in spoken English at the time of the census, as well as a question on languages currently spoken in the home other than English, a pattern repeated in the 1990 Census questions.

Two data deficiencies in the SIE and the 1980 Census are corrected in the survey data studied in this chapter. First, the survey asked for self-reported fluency in English at the time of *first* arrival in the United States, as well as the SIE/census question on fluency at the time of interview.[5] Second, the survey included a question on self-reported fluency in reading English at the time of interview. Furthermore, the survey methodology included a bilingual interviewer and both English and Spanish versions of the survey instrument. This methodology should reduce reporting errors and nonresponse on the part of those least fluent in English.

Most of the American studies of English language fluency have focused on Hispanics. The earliest study was by McManus and his colleagues and concluded that once language skills are taken into account "the differentials in wages which are associated with Hispanic ethnicity, U.S. nativity, schooling abroad and time in the United States are no longer statistically significant" (see McManus, Gould, and Welch 1983, p. 121; see also Gould, McManus, and Welch 1982). They then indicate that the interpretation is not that these factors are unimportant but rather that "there effects are mediated through" measured English language skills. These findings, however, are the result of a specification error.[6]

Other studies have used the SIE and the 1980 Census for the United States and 1971 and 1981 Canadian Census data to analyze the effect of dominant and minority language proficiency at time of interview on the earnings or occupational status of immigrants.[7] In general, the studies find that dominant-language fluency, entered directly or using an instrumental variables approach, explains some (perhaps one-third) of the observed immigrant-native earnings differential, other variables the same, and accounts for some of the effect of duration in the destination on earnings.

Veltman (1988, pp. 545–46) notes that "no comprehensive account of the language shift process has as yet been produced for immigrants, although several relevant variables have been suggested." He cites only age at migration and length of time in the destination. Using the 1976 SIE data on Hispanics and univariate analysis, he confirms findings reported elsewhere that the propensity to speak English decreases with age at migration and increases with duration in the United States. He did not use the SIE data on schooling or other variables and, of course, did not have data on English fluency at migration. Chiswick and Miller (1992) used the 1980 U.S. and the 1981 Canadian Censuses to analyze dominant-language fluency as a function of demographic, human capital, household characteristic, and minority-language concentration variables. They also analyzed the effect of dominant-language fluency on earnings and the endogeneity of language skills. Their analyses were, of course, limited by the variables available in the censuses.

Research on the role of language in the labor market has been limited by the absence of data on English speaking ability at immigration. Furthermore, the research has not been able to resolve the issue as to whether speaking ability is sufficient or whether the speaking variable is reflecting some of the effects of an important unmeasured variable with which it is correlated, fluency in reading English. The analysis in this chapter addresses both issues.

II. The survey data

The data for this study are from a survey of illegal aliens apprehended by the Los Angeles District Office of the Immigration and Naturalization Service (INS) during the 12-month period starting October 1986.[8] The survey instrument was administered to all illegal aliens detained and processed during this period who satisfied the following criteria: age 15 and over, in the United States for at least 4 days during the current stay, non-violent, and not held for felony prosecution. The interviewer was fully bilingual in English and Spanish, and the survey instrument was available in both languages. The interviewer was clearly identified as not being an employee or agent of INS, and the interviews were conducted in private.

The questionnaire was designed to elicit information on the income, employment, and household structure of the illegal alien population of the United States who would not be eligible for legalization under the Immigration Reform and Control Act of 1986.[9] In addition to standard demographic, skill, and labor market questions the survey included the following language questions:

1) What languages did you usually speak at home as a child? (Circle all that apply.) Spanish, English, Other (specify).

2) How well did you speak English when you first came to the United States? Would you say: Very well, Well, Not well (a little bit), or Not at all?
3) Currently how well do you *speak* English? Would you say: Very well, Well, Not well (a little bit), or Not at all?
4) Currently how well do you *read* English? For example, an English language newspaper. Would you say: Very well, Well, Not well (a little bit), or Not at all?

Self-assessment of language skills is always problematical. A test of English language competency that may be more reliable would be very costly to implement for a large sample. The procedure adopted here also has the advantage of comparability to questions asked by the U.S. Bureau of the Census on English speaking fluency. Reliability should be enhanced by the survey procedure of having a bilingual interviewer and English and Spanish versions of the questionnaire. Furthermore, there is no reason to believe the procedure generates systematic biases in the interpretation of the findings.

The survey resulted in 836 completed interviews for males. There were only 14 refusals, for an interview refusal rate of only 1.6%. The item non-response rates were also very low. The average length of the interview was 36 minutes and did not differ between Mexican and non-Mexican men. Among the 836 men, 94% of the interviews were conducted in Spanish, 4% in English (primarily for men from Canada and the Eastern Hemisphere), and 2% in English and Spanish. In only two instances was it not possible to conduct the interview because a translator fluent in a third language was not available.

The sample demonstrates characteristics typical of illegal aliens in the Los Angeles labor market (Chiswick 1984, 1988*b*, and 1989; Kossoudji and Ranney 1984; and Massey 1987). In the sample, 84% of the men were from Mexico, 11% from Central America, 2% from South America, and 3% from Canada and the Eastern Hemisphere. Half of the Mexican men were born in the northern part of the Central Plateau, the home of 22% of the population of Mexico.

The mean age of the sample was young, only 23 years. The average for the Mexican men was 22 years and about 28 years for the others. They had a relatively short mean duration in the United States during their current stay, 1.5 years overall, 1.4 years for the Mexicans, and 2.2 years for the others. However, the Mexican men were more likely to have had previous stays or episodes; 28% for the Mexicans, only 15% for the others.

The schooling levels in this population are very low. The mean level of schooling outside the United States was 7.1 years overall and 7.0 years for both the Mexican and other Latin American aliens. It was 8.3 years for the Canadian/European men and 13.2 years for the other Eastern Hemisphere men. This generally low educational attainment was not substantially augmented by schooling in the United States. Among the Mexican men 77% had no schooling in the United States, and another 14% had less than 1 year. Among the non-Mexican men, 61% had no U.S. schooling, and another 20%

had less than 1 year. Among the small number currently enrolled in school, about half reported enrollment in an "English-as-a-second-language" program for both the Mexican and other aliens.

Reflecting the languages spoken in their countries of origin, nearly all of the Mexican and other Latin American aliens reported that only Spanish was spoken in the home when they were a child. Among the 18 Asian, African, and Middle Eastern aliens, all reported a language other than English, but nearly 40% also reported English was spoken in the home when they were a child.

III. Speaking and reading English

This section analyzes the English language speaking and reading skills of the sample of aliens. Although several studies have included analyses of current English language proficiency, this study is unique in being able to analyze speaking fluency at immigration and fluency in both speaking and reading English at the time of interview. This section first analyzes the speaking skills of the aliens. It closes with the analysis of English reading skills.

A. Speaking English

The aliens came to the United States with very poor English language skills. Among the Mexican men, nearly 80% reported that they could not speak English at all, another 20% reported that they spoke "not well," only 1% reported speaking "well," and none said "very well." For the men from other countries, English language skills at migration were only slightly higher: 70% spoke "not at all," nearly 20% reported "not well," 8% spoke "well," and only 5% (primarily from Canada and the United Kingdom) spoke "very well."

Language skills increased by the time of the interview, in spite of the short duration in the United States. Among the Mexican migrants, the proportion reporting that they spoke English "not at all" fell by half from four-fifths to two-fifths (see Table 12.1). Those reporting "not well" increased from one-fifth to over one-half. And 6% reported speaking "well" or "very well," in contrast with the 1% prior to coming to the United States.

The male aliens from other countries experienced greater improvements in their speaking skills (see Table 12.2). Less than 30% reported that they spoke English "not at all," a decline from nearly 70% at arrival. And nearly 30% reported speaking "well" or "very well," more than doubling the 13% at arrival.

The data on English language proficiency prior to first coming to the United States and at the time of interview permit a multivariate analysis of the determinants of increased fluency in English. It is hypothesized that, controlling for language skills at arrival, the longer aliens are in the United States, the greater their fluency in English. It is also hypothesized that, due to the complementarity of schooling and language fluency, in a low-fluency population those with higher levels of schooling would have a greater

Table 12.1 English speaking fluency of Mexican men

When first came to United States	*At time of interview*					
	Very well	*Well*	*Not well*	*Not at all*	*Total*	*%*
Very well	0	0	0	0	0	.0
Well	2	3	1	0	6	.8
Not well	1	15	119	1	136	19.3
Not at all	2	19	252	291	564	79.9
Total	5	37	372	292	706*	. . .
%	.7	5.2	52.7	41.4	. . .	100.0

* One nonrespondent to both questions.

Table 12.2 English speaking fluency of non-Mexican men

When first came to United States	*At time of interview*					
	Very well	*Well*	*Not well*	*Not at all*	*Total*	*%*
Very well	6	0	0	0	6	4.7
Well	2	8	0	0	10	7.8
Not well	1	10	13	0	24	18.6
Not at all	2	8	44	35	89	69.0
Total	11	26	57	35	129	. . .
%	8.5	20.2	44.2	27.1	. . .	100.0

increase in English language fluency. Furthermore, the effect of a higher level of schooling would be greater the longer the duration of residence. That is, controlling for initial speaking ability, education would have no separate effect at arrival but would have an increasing effect with duration of residence. Finally, it is hypothesized that the greater extent of temporary migration of Mexican aliens, because of the low cost of to-and-from migration, and the existing Spanish speaking Mexican-origin enclave in the Los Angeles area would retard their investments in developing English fluency.[10]

The variables used in the econometric analysis of speaking English (and the analyses below for reading English and for earnings) are defined in the Appendix.[11] The multiple regression analysis of speaking English is reported in Table 12.3. The dichotomous dependent variable SPEKWELL takes the value of one if the respondent reports speaking English "well" or "very well" at the time of interview; otherwise it is zero.[12] The equations are computed overall and separately for Mexican and other Latin American men, using ordinary least squares (OLS) and logit analysis.[13]

The first two columns in Table 12.3 report the simple linear regression for speaking well or very well (SPEKWELL) both with and without the statistical control variables for initial English speaking ability. The explanatory power of the equation is increased significantly (from 34% to 39%) when speaking skills at arrival are held constant. Perhaps most important, the partial effects of education and Canadian/Eastern Hemisphere origin are biased upward when speaking skills at arrival are not held constant. That is, part of the greater fluency of those with more schooling and from Canada/Eastern Hemisphere is due to their greater English fluency at arrival. There is little substantive difference between the results of the OLS specification and the logit specification (cf. Table 12.3, cols. 2 and 6).

Table 12.3 Analysis of fluency in speaking English (SPEKWELL) by country of origin, OLS, and logit

Variable	*OLS*					
	All	*All*	*All*	*Mexico*	*Other Latin American countries*	*Logit all*[a]
AGE	−.0003	−.0007	.0018	.0003	.0004	−.0374
	(−.216)	(−.592)	(1.376)	(.199)	(.128)	(−1.21)
EDUC	.0178	.0135	.0049	.0105	.0243	.2984
	(6.405)	(4.891)	(1.548)	(3.595)	(3.151)	(4.86)
DURNOW*	.0317	.0292	.0243	.0212	.0706	.3372
	(8.401)	(8.004)	(1.776)	(5.602)	(5.862)	(6.09)
SPOKE1	b	.4199	.3896	b	.6282	21.675
		(3.458)	(3.290)		(2.419)	(.0006)
SPOKE2	b	.4195	.4141	.6446	.3430	3.7311
		(6.239)	(6.316)	(7.278)	(1.328)	(2.97)
SPOKE4	b	−.0745	−.0694	−.0602	−.0700	−1.0672
		(−3.639)	(−3.473)	(−2.870)	(−.979)	(−3.01)
OTHLATIN	.0423	.0452	−.0552	b	b	.6307
	(1.651)	(1.835)	(1.649)			(1.42)
OTHER	.7576	.5905	.6671	b	b	3.3483
	(13.492)	(7.621)	(6.927)			(2.99)
(EDUC)(DURNOW*)	b	b	.0051	b	b	b
			(4.587)			
(AGE)(DURNOW*)	b	b	−.0013	b	b	b
			(−2.920)			
(OTHLAT)(DURNOW*)	b	b	.0366	b	b	b
			(3.328)			
(OTHER)(DURNOW*)	b	b	−.0878	b	b	b
			(−2.497)			
CONSTANT	−.1101	.0101	.0095	−.0115	−.1506	−4.5847
	(−3.028)	(−.248)	(−.212)			(−5.36)
R^2	.3438	.3947	.4339	.1732	.4940	. . .
Adjusted R^2	.3397	.3886	.4263	.1671	.4621	. . .
N	802	802	802	680	102	802

Note: Men who spoke only some English (SPOKE3) before coming to the United States are the benchmark in cols. (2)–(5). In the pooled equation Mexican men are also the benchmark. *t*-ratios are in parentheses.

a Logit analysis final value of log-likelihood ratio = −134.7. There are very few observations in the SPOKE1 category.

b Variable not included.

Controlling for speaking skills at arrival, there is a highly significant positive relationship between the ability to speak English well or very well and variables for duration in the United States, schooling, and a non-Mexican origin (Table 12.3). Overall, an extra year in the United States during the current stay is associated with a 3 percentage point higher probability of speaking well or very well, but the effect differs by country of origin. It is only 2 percentage points for Mexican aliens and 7 percentage points for other Latin American aliens, and the difference is statistically significant (Table 12.3, cols. 4 and 5).

An additional year of schooling is also associated with a higher probability of speaking well or very well. Overall the effect is 1.3 percentage points per year of schooling. However, it is 1.0 percentage point for Mexican aliens and 2.4 percentage points for other Latin American aliens. Again the difference is statistically significant.

Table 12.3, column 3, analyzes SPEKWELL by including interaction variables. As hypothesized, differences in schooling at immigration have no effect on language skills when initial speaking skills are held constant. However, the effect of a higher level of schooling increases with duration in the United States. At 3 years in the United States, an extra year of schooling raises the proportion speaking well or very well by 2.0 percentage points overall. Separate regressions by origin indicate the effect is 1.7 percentage points for Mexicans and 3.2 percentage points for other Latin American men.

The analysis indicates that older migrants have more difficulty adapting to English. As hypothesized, at arrival there is no effect of age on English skills, but the age-duration interaction variable indicates that the improvement in English-language skills with duration is significantly slower for older migrants, other variables the same. It is slower by 1.3 percentage points for each year difference in age.

The level and improvement in language skills also varies by country of origin. Although in Table 12.3, column 3, other Latin American aliens have a poorer fluency at arrival than Mexican aliens (coefficient = –0.055, $t = -1.65$), their skills increase more sharply with duration (coefficient = 0.037, $t = 3.3$), and they surpass the Mexican aliens after 18 months.[14]

The small sample of other aliens (Canadian and Eastern Hemisphere = OTHER) initially have much greater proficiency in English (Table 12.3, col. 3, coefficient = 0.667, $t = 6.927$). However, the interaction term indicates the difference narrows with duration (coefficient = –0.087, $t = -2.497$).

The primary purpose of the SPOKE variables in Table 12.3 is to control for initial conditions.[15] The coefficients indicate the not surprising result that those who had greater English speaking fluency at arrival were more likely to have greater fluency at the time of interview.

In summary, controlling for English speaking ability at immigration, spoken English fluency improves with duration in the United States after immigration. This improvement is steeper for those with higher levels of schooling, who are younger at immigration, and who came from Latin American countries other than Mexico. Those with greater speaking fluency at arrival also have greater fluency at the time of interview.

B. Reading English

It is unfortunate that questions on English literacy no longer appear in most surveys and censuses that have been used to study immigrant labor. Believing that this is still an important issue, especially for low-skilled immigrants, I included a question in the survey instrument on the self-reported ability to read English at the time of interview. The responses could fall into one of four categories: “very well,” “well,” “not well,” or “not at all.”

The Mexicans reported very low skills in reading English. Nearly two-thirds of the Mexican men reported “not at all,” and one-third reported “not well.” For other nationals, the situation was somewhat better. Nearly half reported “not at all,” over a third reported “not well,” and nearly one-quarter reported “well” or “very well.” Aliens who had been in the United States for 3 or more years during their current stay had a higher level of reading ability than more recent arrivals (see Table 12.4). Yet, only 11% of the Mexicans and 37% of other nationals in the United States for 3 or more years read English “well” or “very well.”

It is to be expected that English speaking fluency would be an important determinant of English reading skills. Those more fluent in speaking English would be more adept at learning how to read and at increasing their fluency. Therefore, the determinants of speaking skills discussed above are also determinants of reading skills. Yet the inquiry here is whether reading fluency is related to demographic and human capital variables after controlling for speaking fluency.

The acquisition of reading skills is a form of investment in human capital.

Table 12.4 Ability to read English at the time of interview by country of origin and duration in the United States

Reading ability	*Mexico**		*Other countries*		*Total*
	Less than 3 years	*3 or more years*	*Less than 3 years*	*3 or more years*	
Very well	2	1	5	6	14
Well	15	13	5	13	46
Not well	162	67	21	22	272
Not at all	392	51	46	11	500
Total	571	132	77	52	832

* Duration not reported for three Mexican males, and reading ability not reported for a fourth.

The accumulated stock of reading capital would increase with greater exposure to the United States, even when speaking skills are held constant. This implies that reading skills would increase with the duration of the current residence in the United States. It also implies that, for aliens from countries where multiple stays in the United States are not uncommon (such as Mexico), reading skills would increase with age when duration of the current stay is held constant.

Because of the complementarity among types of human capital, the costs involved in acquiring English reading skills would be smaller for those with more schooling, while the benefits from doing so would be larger. The effect of schooling, however, is expected to increase with the length of time in the United States.

The regression equations are reported in Table 12.5 for the dichotomous dependent variable, READWELL, which is unity for those who read "well" or "very well," using both OLS and logit analysis.[16] The simplest functional forms are presented in Table 12.5, columns 1–3, which examine the effects of adding speaking fluency to a reading skills equation. As indicated in column 1, English reading skills are significantly greater among those with more schooling, who have been in the United States a longer period of time, and who are of Canadian/Eastern Hemispheric origin. The addition of English speaking skills at immigration (STSPWELL) significantly increases the explanatory power of the equation (adjusted R^2 increases from 0.32 to 0.45). The inclusion of STSPWELL reduced by about one-quarter the partial effects on reading skills of schooling and duration and reduces by almost two-thirds the coefficient on Canada/Eastern Hemisphere, but these explanatory variables remain highly significant. Controlling for speaking skills at immigration, each extra year of schooling increases the probability of reading English well or very well by 1.3 percentage points, while each extra year in the United States raises it by 2.3 percentage points. Furthermore, as would be expected, those who spoke English well or very well at immigration had greater reading ability in English at the time of interview.

In column 3 of Table 12.5, the variable for English speaking skills at migration is replaced by the same variable at the time of the interview (SPEKWELL). Presumably because similar processes enhance speaking and reading skills, this substitution increases the explanatory power of the equation (adjusted R^2 increases from 0.45 to 0.57). Since current speaking skills have been shown above to increase with schooling level and duration in the United States, substituting current for initial speaking fluency lowers the partial effects of these variables. However, even after controlling for current English speaking skills, current reading ability is significantly greater for those with more schooling, in the United States a longer period of time, from Canada/Eastern Hemisphere, and for those who immigrated at an older age.

The logit equation in Table 12.5, column 5, demonstrates the statistical

Table 12.5 Regression analysis of fluency in reading English well or very well (READWELL), OLS, and logit

Variable	*OLS*				
	(1)	(2)	(3)	(4)	*Logit*[a] (5)
AGE	.0018	.0010	.0019	.0009	.0654
	(1.66)	(1.08)	(2.25)	(.79)	(1.94)
EDUC	.0168	.0131	.0072	.0060	.2433
	(6.78)	(5.82)	(3.56)	(2.27)	(3.01)
DURNOW*	.0266	.0229	.0095	−.0155	.1852
	(7.92)	(7.52)	(3.41)	(−1.33)	(2.49)
STSPWELL	b	.7051	b	.7319	b
		(13.43)		(11.01)	
SPEKWELL	b	b	.5390	b	4.4105
			(21.42)		(8.36)
OTHLATIN	.0096	.0105	−.0132	−.0495	−.6697
	(.42)	(.51)	(−.72)	(−1.77)	(−1.01)
OTHER	.6177	.2326	.2093	.3742	.3472
	(12.35)	(4.34)	(4.74)	(4.47)	(.44)
(AGE)(DURNOW*)	b	b	b	.0003	b
				(.76)	
(EDUC)(DURNOW*)	b	b	b	.0043	b
				(4.54)	
(OTHLAT)(DURNOW*)	b	b	b	.0228	b
				(2.46)	
(OTHER)(DURNOW*)	b	b	b	−.0743	b
				(−2.47)	
(STSPWELL)(DURNOW*)	b	b	b	−.0149	b
				(−1.18)	
CONSTANT	−.1551	−.1139	−.0958	−.0577	−8.4380
	(−4.79)	(−3.87)	(−3.69)	(−1.72)	(−6.88)
R^2	.3277	.4519	.5738	.4804	. . .
Adjusted R^2	.3234	.4478	.5705	.4732	. . .
N	802	802	802	802	802

Note: Mexican men are the benchmark. *t*-ratios are in parentheses.
a Logit analysis, final value of log-likelihood function = −72.1.
b Variable not included.

importance of the same variables as in the OLS analysis, schooling, duration, English speaking skills, and country of origin (cf. Table 12.5, cols. 3 and 5).

Interaction variables are added to the equation in Table 12.5, column 4, and regressions were also computed separately by country of origin. If speaking skills at migration are controlled for, an extra year of schooling increases English reading skills, with the partial effect increasing with the duration of residence. At 3 years of residence an extra year of schooling increases the probability of reading well or very well by 1.9 percentage points. The partial effect is 1.3 percentage points per year of schooling overall, but it is smaller for Mexican immigrants, 1.0 percentage point, compared to 2.1 percentage points for other Latin Americans. Age, however, shows no statistically significant effect on reading skills when initial speaking skills are held constant.

The partial effect of duration of residence is a highly statistically significant 2.3 percentage points per year in this sample controlling for initial speaking fluency. This effect varies with schooling level: it is larger for those with more schooling (Table 12.5, col. 4). It also varies by country of origin, being larger for the other Latin American migrants than for the Mexicans.

In summary, the analysis indicates that English reading ability among low-skilled immigrants is related to their overall skill level. Reading fluency is significantly greater for those with more schooling, in the United States a longer period of time, more fluent in speaking English at immigration, and from Canada/Eastern Hemisphere countries. Duration in the United States has a larger positive impact for those with more schooling and for Latin American men other than Mexicans.

IV. Earnings

This section reports the results of the multiple regression analysis of earnings for the sample of illegal aliens. Two dependent variables are considered: the usual weekly earnings during the current stay and the most recent hourly wage in the current stay. Because of missing values for one or more of the variables in the analysis, particularly the earnings variables, the analysis of usual weekly earnings is for about 380 observations, and the analysis for hourly wages is for 605 observations.[17]

Following standard practice, the natural logarithm of earnings is regressed on demographic and human capital variables.[18] It is hypothesized that earnings increase with the level of schooling attainment (EDUC), labor market experience in the current stay (DURNOW*), and total labor market experience (T) and that earnings are lower for those who are not currently married (SPOUSEAB). It is also hypothesized that earnings are greater for those more fluent in English (SPEKWELL and READWELL).

The regression analysis of usual weekly earnings is presented in Table 12.6 with a statistical control for the natural logarithm of usual hours of work per week (LNHOURS/WK) in columns 1 and 2 but not in columns 3 and 4. When hours per week are held constant, the coefficients of the other variables in the equation measure their effects on usual earnings per hour worked. Columns 2 and 4 include the speaking and reading variables (SPEKWELL and READWELL). The regression analysis for the most recent hourly wage is reported in Table 12.7 for the full sample and separately by country of origin, where the regressions in this table differ by the inclusion of the language variables.

As has been shown elsewhere, schooling has a highly significant effect on the earnings of the illegal alien (see, e.g., Chiswick 1984, 1988*b*; Kossoudji and Ranney 1984; and Massey 1987). In these data, weekly or hourly earnings rise by about 2% and 3%, respectively, for each additional year of schooling. The effect is somewhat larger, 3.5%–4%, for the aliens from other Latin American countries. These partial effects are comparable to coefficients

Table 12.6 Regression analysis of the natural logarithm of the usual weekly earnings during the current stay (Dependent Variable: LNWKEARN)

Variable	(1)	(2)	(3)	(4)
EDUC	.01906	.01447	.02370	.01967
	(2.500)	(1.858)	(2.599)	(2.102)
T	.01122	.01055	.01339	.01263
	(1.655)	(1.566)	(1.641)	(1.548)
TSQ	−.00024	−.00024	−.00038	−.00038
	(−1.429)	(−1.439)	(−1.920)	(−1.894)
DURNOW*	.01961	.01441	.03779	.03276
	(2.405)	(1.678)	(3.895)	(3.254)
SPOUSAB	−.08312	−.08742	−.13167	−.13818
	(−1.644)	(−1.736)	(−2.177)	(−2.283)
LNHOURS/WK	.72461	.73138	[a]	[a]
	(12.804)	(12.972)		
SPEKWELL	[a]	−.02862	[a]	.05836
		(−.320)		(.538)
READWELL	[a]	.26992	[a]	.16031
		(2.467)		(1.209)
OTHWHEM	−.03722	−.04660	−0.05654	−.06784
	(−.730)	(−.914)	(−0.923)	(−1.102)
OTHER	−.07369	−.25913	−0.23439	−.40360
	(−.671)	(−1.955)	(−1.779)	(−2.516)
CONSTANT	2.26849	2.29021	4.87120	4.91345
R^2	.3686	.3811	.0912	.0993
Adjusted R^2	.3550	.3644	.0743	.0777
N	380	380	385	385

Note: Mexican men are the benchmark. t-ratios are in parentheses.
[a] Variable not included.

found in other analyses for illegal aliens, although they are lower than what is found in studies of legal immigrants (Chiswick 1984 or 1988*b*).

Labor market experience in the United States during the most recent stay (DURNOW*) has a significant effect on usual weekly earnings (Table 12.6). When hours of work are not held constant and there are no controls for language fluency (Table 12.6, col. 3), the partial effect of duration in the United States during the current stay is 3.8% per year, with a t-ratio of 3.9. The statistical control for usual hours of work lowers the partial effect of current U.S. experience to 2.0% (t = 2.4) because usual hours worked per week increases with duration. The partial effect of duration on usual weekly earnings is reduced from 3.8% to 3.3% (or when hours are held constant, from 2.0% to 1.4%) when the language variables are held constant.

When the most recent hourly wage is the dependent variable, the coefficient of the duration variable is smaller and is less significant (Table 12.7). Indeed, when the language variables are included in the hourly wage equation, duration in the United States is not statistically significant. Controlling for

Table 12.7 Regression analysis of the natural logarithm of the most recent hourly wage during the current stay (Dependent Variable: LNWAGENW)

Variable	*All countries*				*Other Latin American countries* (5)
	(1)	(2)	(3)	*Mexico* (4)	
EDUC	.02962	.02814	.02587	.02266	.03434
	(4.065)	(3.778)	(3.456)	(2.630)	(2.304)
T	.02226	.02183	.02153	.01846	.03405
	(3.389)	(3.315)	(3.281)	(2.443)	(2.401)
TSQ	−.00036	−.00035	−.00035	−.00029	−.00060
	(−2.223)	(−2.169)	(−2.179)	(−1.590)	(−1.674)
DURNOW*	.01370	.01153	.00927	.01445	−.03593
	(1.726)	(1.398)	(1.120)	(1.560)	(−1.770)
SPOUSAB	−.04835	−.05039	−.04908	−.02105	−.10868
	(−.941)	(−.980)	(−1.958)	(−.338)	(−1.135)
SPEKWELL	[a]	.07669	−.04529	−.07627	.06082
		(.969)	(−.478)	(−.684)	(.318)
READWELL	[a]	[a]	.25881	.31667	.35649
			(2.326)	(2.320)	(1.623)
OTHLATIN	−.04827	−.05247	−.05277	[a]	[a]
	(−.951)	(−1.030)	(−1.039)		
OTHER	−.00520	−.06365	−.15758	[a]	[a]
	(−.046)	(−.493)	(−1.169)		
CONSTANT	1.09180	1.10725	1.12692	1.13753	1.03475
R^2	.0672	.0687	.0771	.0580	.2692
Adjusted R^2	.0563	.0562	.0631	.0447	.2044
Sample size	87	605	605	605	502

Note: Mexican men are the benchmark in the pooled equations. *t*-ratios are in parentheses.
[a] Variable not included.

language skills reduces the effect of duration in the United States on the hourly wage because, as was shown above, English language fluency itself increases with duration.

The coefficients of the variables for total labor market experience (T) and its square (TSQ) and marital status (SPOUSAB) are not sensitive to the inclusion of language variables. In Table 12.6, those who are not married have lower usual weekly earnings (by about 13%). About one-third of this differential arises because they work fewer hours and two-thirds because they earn less even when hours worked are held constant (about 8% lower earnings). In the analysis of hourly wages, however, there is generally no significant marital status effect.

Other variables the same, there is no difference in usual weekly earnings or hourly wages between Mexican and other Latin American aliens and no effect of adding statistical controls for language fluency. By way of contrast, although the coefficient is always negative, Canadian and Eastern Hemisphere aliens (OTHER) show no significant weekly earnings or hourly wage difference from the Mexican men when language variables are not included in the equation.[19] When English language fluency is held constant, however, the usual weekly earnings of the Canadian/Eastern Hemisphere illegal aliens (OTHER) are significantly lower than the earnings of Mexican aliens.[20]

Last, consider the coefficients of the English language fluency variables, SPEKWELL and READWELL. Alternative specifications, the most informative of which are presented in Tables 12.6 and 12.7, indicate that the variable for reading English consistently has a larger coefficient and a higher t-ratio than the variable for speaking. In the analysis for weekly earnings with a control for hours worked per week (Table 12.6, col. 2), reading well or very well increases earnings by a highly statistically significant 31% (converting the coefficient of 0.27 to a percent increase), while the speaking coefficient is very small and not significant (coefficient of –3%, $t = -0.3$). In the analysis of hourly wages (Table 12.7, col. 3), reading well or very well increases wages by a highly statistically significant 30% (converting the coefficient of 0.26 to a percent increase). Comparable findings appear when separate analyses are performed for Mexican and other Latin American aliens—speaking fluency has no separate effect, and reading well or very well increases wages by a statistically significant 37% and 42%, respectively, for the Mexican and other Latin American men.

Thus, reading skills dominate speaking skills in the analysis of the effect of English language fluency on earnings. Furthermore, the inclusion of language fluency variables reduces, but does not not eliminate, the measured effect on earnings of experience in the U.S. labor market. Finally, the inclusion of language variables alters the relative differences in earnings by country of origin. The relative earnings of Hispanic aliens is enhanced when there is an adjustment for their lower level of fluency in English.

V. Summary and conclusion

This article is concerned with the determinants of English language fluency and the effects of English language fluency on the earnings of a sample of low-skilled aliens. Using special survey data on over 800 illegal aliens, the analysis shows the importance of certain variables that are not available in the Census Bureau data that have been used previously to study immigrant labor market activities. These variables are English speaking fluency at migration and English reading fluency.

Using longitudinal data from self-reported responses to questions on English speaking fluency at arrival in the United States and at the time of interview, I show that English speaking fluency improves with duration in the United States. The improvement is greater for those with higher levels of schooling, presumably because of the complementarity of schooling and language acquisition and utilization. The improvement with duration is also greater for those who came to the United States at a younger age, reflecting the greater ease of language acquisition for younger people. The improvement with duration is slower for Mexican aliens. This may reflect the greater temporary nature of their stays and the adverse effects on English language acquisition of living in a language-minority enclave. Furthermore, tests indicate that analyses of English speaking fluency result in upward-biased estimated effects of schooling and non-Mexican origin if fluency at arrival is not held constant.

The men in the sample reported very poor English reading skills, particularly the Mexican nationals. The regression analysis of English reading ability demonstrates the large and highly significant effect of English speaking skills at migration and at the time of interview. Yet, even after speaking skills are controlled for, there are important effects on reading of demographic and human capital variables. Reading skills increase with schooling level and duration in the United States, and the increase with duration is greater for those with more schooling. This presumably reflects the complementarity of various types of human capital. Age at immigration apparently has no independent effect on reading fluency when speaking fluency at immigration is held constant, but it has a positive effect when speaking fluency at the time of the interview is held constant. Hispanic aliens reported poorer English reading skills than those from Canada/Eastern Hemisphere, even when other variables are the same. This may be reflecting adverse impacts on the acquisition of English reading skills of living in a language-minority enclave.

The analyses of the usual weekly earnings and most recent hourly wages of the illegal aliens show patterns consistent with other studies. Earnings increase with level of schooling, total labor market experience, and experience in the U.S. labor market. Adding variables for English language fluency (speaking and reading) reduces, but does not eliminate, the partial effect of duration in the United States on earnings. The coefficients of the schooling, marital status, and total experience variables are not affected.

In the analyses of weekly earnings and hourly wages, the variables measuring English reading proficiency dominates the variable measuring English speaking skills. That is, measures of reading skills are more important statistically for understanding labor market outcomes than merely measures of speaking English.

These findings indicate the importance of English language proficiency, especially reading and writing skills, for the labor market success of immigrants. They also suggest that future surveys of immigrants should include questions on English proficiency at arrival as well as at the time of interview and that questions on reading skills may be more useful than merely asking the respondent's fluency in spoken English. Furthermore, tests designed by the immigration authorities to evaluate the applicant's likely adjustment to the U.S. labor market for purposes of legalization, immigration, or naturalization would be more effective if they also measure English reading skills.

Appendix

Table 12A.1 List of variables used in the statistical analysis

Variable	*Code*	*Description*
Language skills	SPEKWELL, READWELL	Dichotomous variable, equal to unity if speak English or read English well or very well; zero otherwise.
	SPOKE1, SPOKE2, SPOKE3, SPOKE4	English speaking ability when came to the United States for the first time: 1 = very well, 2 = well, 3 = not well, 4 = not at all.
	STSPWELL	Dichotomous variable equal to unity if SPOKE1 or SPOKE2 are unity; otherwise zero.
Earnings	LNWKEARN, LNWAGENW	The natural logarithm of the usual weekly earnings, current stay or of the most recent hourly wage, current stay.
Schooling	EDUC	Total years of schooling.
Age and experience	AGE	Age in years.
	T	Years of labor market experience. (Age – schooling – 5, or years since age 15, for those with 10 or fewer years of schooling.)
Marital status	SPOUSAB	Dichotomous variable, equal to unity if divorced, widowed, or never married; zero otherwise.
Duration in United States	DURNOW*	Years in the United States, current stay. DURNOW*= (year and month of interview) minus (year and month last entered).
Hours of work	LNHOUR/WK	The natural logarithm of hours worked per week, current stay.
Country of birth[a]	MEXICO, OTHLATIN, OTHER	Dichotomous variable, equal to unity if born in Mexico, another Latin American country, or another country.

* OTHLATIN includes Belize, Colombia, Chile, Costa Rica, El Salvador, Equador, Guatemala, Honduras, Nicaragua, Peru, and Venezuela. OTHER includes Canada, India, Iraq, Israel, Italy, Korea, Lebanon, Morocco, Nigeria, Pakistan, Philippines, Syria, Taiwan, and United Kingdom.

Notes

1 The survey analyzed in this chapter was financed by a grant from the Immigration and Naturalization Service, U.S. Department of Justice, and was conducted by the Survey Research Laboratory, University of Illinois. The research for this article was financed by grants from the Sloan Foundation and the Institute of Government and Public Affairs, University of Illinois. I appreciate the research

assistance of Xiao-Bo Li and the comments received on an earlier draft from Evelyn Lehrer, Luis Locay, Evelina Tainer, and François Vaillancourt and from participants at a seminar at the University of Illinois at Chicago and a session of the American Economics Association annual meeting. I am, however, solely responsible for the contents of this chapter.

2 The importance of dominant-language skills, even for low-skilled workers, has been explicitly recognized in the amnesty program in the 1986 Immigration Reform and Control Act. To change their status from "temporary resident alien" to "permanent resident alien" within the 1-year grace period, those granted amnesty need to demonstrate a minimal command of English or enroll in at least 40 hours of English language instruction in an approved program (see Chiswick 1988*a*).

3 The person's "mother tongue," the language other than or in addition to English spoken in the home when the person was a child, was asked in the 1970 Census questionnaire administered to 15% of the population, but a key variable, duration in the United States, was asked only on the questionnaire administered to a non-overlapping 5% of the population.

4 The SIE also included a question on reading: "How often does (the respondent) read an English language newspaper?" with "most days," "occasionally," and "(almost) never" as the acceptable responses. While it is not clear what the reading question does measure, it is clearly not a satisfactory measure of English reading fluency.

5 The longitudinal data on a skill relevant in the labor market can be used to address the critique of Borjas (1985) that the improvement in earnings with duration in the destination observed in cross-sectional data is due to declining cohort quality, with no change in the skills relevant for the U.S. labor market as duration of residence increases. Although reestimations using the Borjas data and technique do find "assimilation" effects (see, e.g., Chiswick 1986; and LaLonde and Topel 1990), the longitudinal data in this study provide a more direct test.

6 McManus et al. (1983) used a two-step procedure (p. 121). First, standard earnings functions were computed "to identify important interactions and to identify important questions." Three language questions that had the highest explanatory power for earnings were retained. They then write: "Using interactive responses to these questions we identified seven groups that captured most of the information about wages in the SIE language questionnaire and that, at the same time, are arguably well ordered in terms of proficiency in English. By design, they are ordered in terms of wage predictions after the common variables [e.g., region, marital status, schooling, and experience] are taken into account." Thus, the seven English language proficiency groups used in the McManus et al. earnings analysis are proxies for earnings intervals or categories. Predictable results emerge. They find that their seven dichotomous English fluency variables are very highly statistically significant—far more so than in other studies. They also find that the effects of other determinants of earnings are reduced and that Hispanic ethnicity loses its statistical significance. The statistical methodology has insured that the partial effects of the variables other than language are biased downward. McManus et al. also analyze the determinants of their English language proficiency variable (pp. 119–20). They combine the seven categories into a single index to serve as a dependent variable. Weights are obtained from the earnings function with the dichotomous language variables on the right-hand side. They find that U.S. schooling and U.S. experience raise English language proficiency, but that foreign schooling and foreign experience lower it. What is less clear, however, is whether the analysis is reflecting the effects of the explanatory variables on the language categories or on the earnings weights.

7 For the United States, these studies include Reimers (1983), Grenier (1984), Chiswick (1987), Kossoudji (1988), Tainer (1988), Rivera-Batiz (1989), and

Chiswick and Miller (1992). For studies of the determinants of language fluency and the impact of language fluency on earnings in Canada, where promoting English-French bilingualism is official policy, see e.g., Carliner (1981), Grenier and Vaillancourt (1983), and Chiswick and Miller (1988, 1992). One of the few studies of language proficiency among women is in Boyd (1992).

8 A detailed discussion of the survey procedures, an analysis of the survey methodology, and discussion of the randomness of the sample and the characteristics of the population can be found in Chiswick (1989, app. A). Chiswick (1989) also provides an extensive analysis of these data.

9 For an analysis of the provisions of the 1986 Act and its implications for the characteristics of aliens not eligible for legalization, see Chiswick (1988*a*).

10 Chiswick and Miller (1992) show that in the United States and in Canada residence in an area in which many others speak the same minority language has a significant negative effect on the acquisition of the dominant language. It is not possible to explicitly test the minority-language concentration effect on language fluency in the survey under study that is limited to the Los Angeles area.

11 The means and standard deviations of the variables are reported in Chiswick (1989).

12 Tests indicate this is the most efficacious dichotomization of the four-category language variable for analyses of spoken language fluency.

13 Essentially the same results emerge from the OLS and logit analyses.

14 There is also a large and highly significant difference in the effect of duration on English speaking fluency between Mexican and other Latin American men when the equations are computed separately by origin, where the effect is larger for the latter group.

15 The statistical control for fluency at arrival may also control for individual differences in self-assessment of the same "objective" level of fluency.

16 Tests indicate this is the most efficacious dichotomization of the four-category variable. Similar results emerge when "not at all" is compared to all other reading categories.

17 The average usual weekly earnings during the current stay for the 398 adult men who responded to this question was \$174. The earnings were lower for the Mexican men (\$172) than for the men from other Latin America (\$182) or other countries (\$180).

18 For previous applications to illegal aliens, see Chiswick (1984 or 1988*b*), Kossoudji and Ranney (1984), and Massey (1987).

19 The only exception is the large and marginally significant effect (coefficient = –0.23, $t = -1.8$) when hours are not held constant. Mexican men have a longer workweek, 40.7 hours, in contrast to the 37.7 hours for the Canadian and Eastern Hemisphere men.

20 The coefficient for Canada/Eastern Hemisphere (OTHER) is –0.40 ($t = -2.5$) but declines to –0.26 ($t = -2.0$) when hours worked per week are held constant. These represent earnings that are lower by 33% and 23%, respectively. When the hourly wage is the dependent variable, the coefficient of OTHER becomes more negative but remains insignificant when the language variables are added to the equation. Many of the Eastern Hemisphere illegal aliens were students in the United States who had violated a condition of their visa, usually by working. Their low hourly wage may reflect the adverse effects on job opportunities of dovetailing work with schooling (Lazear 1977).

References

Borjas, George. "Assimilation, Changes in Cohort Quality and the Earnings of Immigrants." *Journal of Labor Economics* 3, no. 4 (October 1985): 463–89.

Boyd, Monica. "Adding Gender: Immigration Trends and Language Fluency Issues in Canada and the United States." In *Immigration, Language and Ethnicity: Canada and the United States*, edited by Barry R. Chiswick. Washington, D.C.: American Enterprise Institute, 1992: 305–72.

Carliner, Geoffrey. "Wage Differences by Language Group and the Market for Language Skills in Canada." *Journal of Human Resources* 16, no. 3 (Summer 1981): 384–99.

Chiswick, Barry R. "The Effect of Americanization on the Earnings of Foreign-born Men." *Journal of Political Economy* 85, no. 5 (October 1978): 897–921.

——. "Illegal Aliens in the United States Labor Market: Analysis of Occupational Attainment and Earnings." *International Migration Review* 18, no. 3 (Fall 1984): 714–32.

——. "Is the New Immigration Less Skilled than the Old?" *Journal of Labor Economics* 4, no. 2 (April 1986): 168–92.

——. "The Labor Market Status of Hispanic Men." *Journal of American Ethnic History* 7, no. 1 (Fall 1987): 30–58.

——. "Illegal Immigration and Immigration Control." *Journal of Economic Perspectives 2*, no. 3 (Summer 1988): 101–15. (*a*)

——. *Illegal Aliens: Their Employment and Employers*. Kalamazoo, Mich.: Upjohn, 1988. (*b*)

——. "Unlegalized Aliens: A Survey and Analysis of Aliens Not Eligible for Legalization." Report prepared for the Immigration and Naturalization Service, U.S. Department of Justice, Washington, D.C., April 1989.

Chiswick, Barry R., and Miller, Paul W. "Earnings in Canada: The Roles of Immigrant Generation, French Ethnicity, and Language." *Research in Population Economics* 6 (1988): 183–224.

——. "Language in the Labor Market: The Immigrant Experience in Canada and the United States." In *Immigration, Language and Ethnicity, Canada and the United States*, edited by Barry R. Chiswick. Washington, D.C.: American Enterprise Institute, 1992: 229–96.

Gould, William; McManus, Walter; and Welch, Finis. "Hispanic Earnings Differentials: The Role of English Language Proficiency." Unicon Report no. 82–02. Los Angeles: Unicon Research Corporation, 1982.

Grenier, Gilles. "The Effects of Language Characteristics on the Wages of Hispanic American Males." *Journal of Human Resources* no. 1 (Winter 1984): 35–52.

Grenier, Gilles, and Vaillancourt, François. "An Economic Perspective on Learning a Second Language." *Journal of Multicultural and Multilingual Development* 4, no. 6 (1983): 471–83.

Kossoudji, Sherrie A. (1988). "English Language Ability and the Labor Market Opportunities of Hispanic and East-Asian Immigrant Men." *Journal of Labor Economics* 6, no. 2 (April 1988): 205–28.

Kossoudji, Sherrie A., and Ranney, Susan I. "Wage Rates of Temporary Mexican Migrants to the U.S.: The Role of Legal Status." Paper presented at the Econometric Society annual meetings, Dallas, December 1984.

LaLonde, Robert J., and Topel, Robert H. "The Assimilation of Immigrants in the U.S. Labor Market." Mimeographed. Chicago: University of Chicago, January 1990.

Lazear, Edward. "Schooling as a Wage Depressant." *Journal of Human Resources* 12, no. 2 (Spring 1977): 164–76.

McManus, Walter; Gould, William; and Welch, Finis. "Earnings of Hispanic Men: The Role of English Language Proficiency." *Journal of Labor Economics* 1, no. 2 (April 1983): 101–3.

Massey, Douglas S. "Do Undocumented Migrants Earn Lower Wages than Legal Immigrants? New Evidence from Mexico." *International Migration Review* 21, no. 2 (Summer 1987): 236–74.

Reimers, Cordelia R. "Labor Market Discrimination against Hispanic and Black Men." *Review of Economics and Statistics* 65 (November 1983): 570–79.

Rivera-Batiz, Francisco. "English Language Proficiency and the Wages of Immigrants in the United States." Mimeographed. Rutgers, N.J.: Rutgers University, May 1989.

Tainer, Evelina. "English Language Proficiency and Earnings among Foreign-born Men." *Journal of Human Resources* 23, no. 1 (Winter 1988): 108–22.

Veltman, Calvin. "Modelling the Language Shift Process of Hispanic Immigrants." *International Migration Review* 22, no. 4 (Winter 1988): 545–62.

13 The linguistic and economic adjustment of Soviet Jewish immigrants in the United States, 1980–2000

With Michael Wenz

Introduction

This chapter constitutes an extension of earlier work by one of the authors on the economic status of turn-of-the-20th century Russian Jewish immigrants, as well as work on Soviet Jewish immigrants to the United States in the late 20th century (Chiswick, 1991, 1992, 1993, 1997, 1999). The specific purpose of this chapter is to continue this line of research on the linguistic and labor market adaptation of adult male Soviet Jewish immigrants in the United States in the post-1965 period.[1] Linguistic adaptation, that is, the acquisition of English language proficiency, is important for many reasons, including increasing access to US schooling and job training and success in the labor market, whether measured by employment or earnings. Moreover, it is important for acquiring US citizenship and thereby expanding job opportunities and increasing political influence. Labor market success is an important element in a family's economic well being and determines current consumption, as well as having an influence on marital formation and stability, fertility, and parental investments in the human capital of their children.

The data under study are from the 2000 Census of Population of the United States, Public Use Microdata Sample (Census, 2003), 5 percent random sample of the population, as well as comparable data from the 1980 and 1990 Censuses.[2]

Migration from the former soviet union

The extent of migration

With the impending and actual collapse of the Soviet Union in 1989 a massive exodus began of the Jewish population. Between 1989 and 2003, 1.6 million Jews and their non-Jewish relatives left the former Soviet Union (FSU), 200,000 each in 1990 and 1991 alone, with the numbers declining thereafter to only 35,000 in 2003 (Tolts, 2004a,b). The primary destination was, of course, Israel, which received over 950,000, or 61 percent of the emigrants. The emigration data suggest that about 315,000 Jews and their non-Jewish relatives left the FSU for the United States, or about 20 percent of the emigrants. Another 160,000 (10 percent) went to Germany and about 20,000 went to Canada, with the remainder settling in a wide range of destinations.[3]

From the start of official record keeping in the United States in 1820, to the present, approximately, 4.0 million people are recorded as having immigrated (permanent resident aliens) to the United States from the Russian Empire or the FSU (Table 13.1). The peak decade was 1901–1910 when 1.6 million immigrants were recorded, followed by 1911–1920 with 0.9 million immigrants (Table 13.2). Immigration from the Soviet Union declined sharply thereafter, with less than 600 recorded in the 1940s, rising to nearly 700 in the 1950s, 2,500 in the 1960s, 39,000 in the 1970s, 58,000 in the 1980s, and nearly 463,000 in the 1990s (1991–2000), for a total of 560,000 over the

Table 13.1 Immigration to the United States from Russia and the Soviet Union, 1820–2002[a].

Time period	*Number of immigrants*
1820–1830	89
1831–1840	277
1841–1850	551
1851–1860	457
1861–1870	2,512
1871–1880	39,284
1881–1890	213,282
1891–1900	505,290
1901–1910	1,597,306
1911–1920	921,201
1921–1930	61,742
1931–1940	1,370
1941–1950	571
1951–1960	671
1961–1970	2,465
1971–1980	38,961
1981–1990	57,677
1991–2000	462,874
2001	55,099
2002	55,464
Total	4,017,143

Note

[a] Individuals granted permanent resident alien status. Includes all constituent units of the Russian Empire and of the FSU.

Source: US Department of Justice, 1993 Statistical Yearbook of the Immigration and Naturalization Service, Washington, DC., September 1994; US Department of Justice, 2001 Statistical Yearbook of the Immigration and Naturalization Service, Washington, DC., February 2003; and US Department of Homeland Security, 2002 Yearbook of Immigration Statistics, Washington, DC., October 2003.

Table 13.2 Soviet refugee and asylee arrivals and admissions, FY 1961–2002.

Year	*Dept of Justice*[a]	*Dept of State*[b]
1961–69	456	
1970	209	
1971	88	
1972	228	
1973	591	
1974	2,221	
1975	3,209	6,211
1976	5,882	7,450
TQ 1976	1,208	
1977	5,296	8,191
1978	9,931	10,688
1979	27,135	24,449
1980	28,692	28,444
1981	11,244	13,444
1982	2,838	2,756
1983	1,449	1,409
1984	791	715
1985	674	640
1986	833	787
1987	3,728	3,694
1988	18,880	20,421
1989	39,831	39,553
1990	53,130	50,716
1991	57,587	38,661
1992	66,026	61,298
1993	51,983	48,627
1994	NA	43,470
1995	NA	35,716
1996	NA	29,536
1997	NA	27,072
1998	NA	23,349
1999	NA	17,220
2000	NA	15,103
2001	NA	15,749
2002	NA	23,150
Total	394,140	598,519

Notes

[a] Soviet refugee and asylee approvals, fiscal year 1961–1993. TQ1976 means transition quarter when fiscal year was adjusted to start October 1 rather than July 1.

[b] Refugee admissions from the Soviet Union, 1976–2002, including all republics from the FSU.

Source: US Department of Justice, 2001 Statistical Yearbook of the Immigration and Naturalization Service, Washington, DC, February 2003, Table 24. US Department of Homeland Security, Yearbook of Immigration Statistics, 2002, Washington, DC, October, 2003.

period 1965–2000. Because of these trends, the analysis is limited to those who first came to the United States to stay in 1965 or later.

The 2000 Census suggests that there were about 700,000 people living in the United States who were born in the former Soviet Union. They may have entered with permanent resident alien visas or under other visas and provisions of immigration law, and some of these subsequently became permanent resident aliens. A large proportion entered as refugees or asylees (Table 13.2).

The refugee experience

Many who sought to leave the Soviet Union would not have had an incentive to leave if not for the anti-semitism and generalized repression. Many were motivated, at least in part, by these factors and not simply conventional economic incentives. There had been a pent up demand for emigration from the Soviet Union, but there had been little expectation that it could be realized. Most emigrants had a limited ability to prepare for the move because of the seemingly arbitrary nature of the Soviet bureaucracy and the apparent randomness as to whose application for an exit visa would be approved, or when it would be approved. Many who sought to leave before the collapse of the Soviet Union experienced various degrees of reprisals and persecution, including loss of their jobs and imprisonment or internal exile. The unexpected and sudden opening for emigration with the collapse of the Soviet Union was accompanied by fears that the door could close at any time accompanied by a resurgence of anti-semitism and repression. Thus, the Soviet Jewish migrants to the United States are more appropriately characterized as refugees than as economic migrants.

Refugees have a different adjustment in the destination than do economic migrants (Chiswick, 1978, 1979). They have more skills specific to the origin and fewer skills that are destination specific or internationally transferable. As a result, at arrival they would be expected to have lower levels of human capital specific to the destination, including language skills, and hence lower earnings than economic migrants with similar measured characteristics. As they make implicit and explicit investments in the destination to increase the transferability of previously acquired skills and to create new skills, it would be expected that they would exhibit a more rapid improvement in language skills and earnings than economic migrants. Yet, because refugees are likely to be less favorably selected for economic success in the destination than otherwise similar economic migrants, it would be expected that the gap between them and economic migrants would narrow, but never close (Chiswick, 2000).

Moreover, because of the lesser degree of the transferability of the skills acquired in the origin in school and on the job (labor market experience) among refugees, the effects of these variables on their earnings in the US would differ from that of economic migrants. In particular, refugees would be expected to have a smaller effect of schooling and pre-migration experience on earnings than would be the case for economic migrants.

While Soviet Jewish immigrants would reflect these refugee characteristics, these might be offset by the different labor market characteristics that have been exhibited by Jews in the US, whether immigrants or native born. American Jews have had high rates of occupation and earnings mobility, have a larger effect of schooling on earnings, and have obtained higher earnings, compared with observationally similar non-Jews (Chiswick, 1999).

As a result the linguistic and labor market progress of Soviet Jewish immigrants in the United States, in comparison to other (non-Jewish) economic migrants would be expected to reflect both their refugee and Jewish experiences and backgrounds.

Who is a Soviet Jew?

The first step in an analysis of "Soviet Jews" in the United States is to define each of the two terms. For the purpose of this study, persons born in any of the constituent republics of the FSU are referred to as "Soviet immigrants". Thus, the analysis is not to be limited to those born in "Russia" loosely defined or in the Russian Federation.

Defining Jews is more problematic. The Census of the United States, unlike censuses in some other countries, such as Australia, Canada, and Israel, has never asked religion. In the 2000 Census microdata file anyone who responds to the question on ethnic ancestry by revealing a religion is assigned the same ancestry code (998) as all other religious responses. Any response indicating Jewishness, even if the response is "secular Jew", is combined with and thereby masked with other religious responses.

Yet, clearly, not all respondents from the FSU are Jews. Those who report an Armenian ancestry or who report that they speak Armenian or Ukrainian at home are not likely to be Jewish. Thus, for a first approximation for the purposes of this paper, persons born in the FSU who do not report an Armenian ancestry, or Armenian or Ukrainian as a language spoken at home are the subject of this analysis and for simplicity of exposition are considered "Soviet Jews"[4] (Chiswick, 1993, 1997).

This study is limited to the analysis of adult (aged 25–64) males. For younger and older persons school enrollment and retirement decisions have a major impact on labor supply and choice of jobs, and hence earnings. Similarly, the labor market attachment of women is strongly influenced by marital status and child care responsibilities. Analyses of these labor supply decisions are beyond the scope of this study.

Descriptive statistics

Table 13.3 reports the means and standard deviations of selected variables relevant for the analysis. The Soviet Jewish immigrants, as defined here, are less proficient in English than either European or Asian immigrants. Among the Soviet Jews, 73 percent reported that they speak only English at home or

Table 13.3 Selected characteristics of adult males who immigrated since 1965 by region of birth, 2000.

Variable[a]	*FSU*[b]	*Europe (Excluding FSU)*	*Asia*	*Latin America*	*Total*[c]
(A) Means and S.D.[a]					
Age	42.4	42.4	41.1	38.3	39.6
	(10.8)	(10.4)	(10.3)	(9.6)	(10.0)
Education (years)	14.8	13.6	14.1	9.3	11.5
	(3.2)	(3.8)	(4.0)	(4.7)	(4.9)
Earnings ($)	37,555	50,889	42,370	22,966	32,704
	(48,691)	(61,390)	(53,979)	(29,191)	(45,018)
Log of earnings	9.35	9.95	9.65	9.19	9.43
	(2.30)	(1.98)	(2.08)	(1.90)	(1.99)
Weeks worked	46.5	47.6	46.8	45.3	46.1
	(11.1)	(9.9)	(10.7)	(11.6)	(11.2)
(B) Percents					
Period of Immigration					
1995–2000	31.2	22.6	19.1	17.8	19.4
1990–1994	38.3	14.5	18.3	17.0	17.5
1985–1989	12.7	13.2	17.0	21.2	18.8
1980–1984	5.9	11.9	18.4	17.4	16.8
1975–1979	9.6	11.0	15.1	11.9	12.6
1970–1974	1.7	12.8	8.0	9.4	9.1
1965–1969	0.7	14.1	4.2	5.3	5.8
Total	100.0	100.0	100.0	100.0	100.0
Married	73.8	68.9	67.4	56.3	61.1
Speaks English[d]	72.7	88.8	81.9	57.3	69.6
With children at home	50.6	44.3	54.3	67.6	60.2
Rural (non-metropolitan area) residence	0.5	0.8	0.6	1.6	1.2
Southern states	10.1	20.5	19.9	32.6	27.6
Unemployed[e]	4.3	2.8	3.1	4.9	4.1
Sample size	9,384	42,911	1,25,487	2,50,828	4,51,844

Notes

[a] Mean values. Standard Deviations within parentheses. Percents with specific characteristics.
[b] FSU excludes persons of Armenian ancestry or who speak Armenian or Ukrainian at home.
[c] Total includes groups not shown separately (23,234 observations), primarily from Canada and Oceania.
[d] Speaks only English at home or speaks another language but speaks English very well or well.
[e] Unemployed as a percent of the labor force.

Source: 2000 Census of Population, Public Use Microdata Sample, 5 percent sample.

speak another language, but speak English "very well" or "well" (Tables 13.3 and 13.4). Twenty-seven percent reported that they spoke English "not well" or "not at all". In contrast, 89 percent of the European immigrants and 82 percent of the Asian immigrants satisfy this definition of English proficiency. Among those with earnings, the Soviet Jews earned nearly $37,600 in 1999, considerably less than the earnings of other European ($50,900) and Asian ($42,400) immigrants, but substantially more than Latin American immigrants ($23,000).

The Soviet Jewish immigrants have some characteristics that would enhance their language proficiency and earnings potential, but other characteristics that would have a negative impact. Their educational level is very high, an average of 14.8 years of schooling, far greater than even the 14.1 years among Asian immigrants, the 13.6 years among other European immigrants, or the 11.5 years among all (including Soviet) immigrants. On the other hand, the Soviet immigrants had a very short period of residence in the US. In 2000, among those who immigrated in 1965 or later, 70 percent of the Soviet Jews had been in the US 10 or fewer years, in contrast to 37 percent overall.

The two measures of employment tell a similar story. Among those who worked, the weeks worked in 1999 were lower for Soviet Jews (46.5 weeks) than for European (47.6 weeks) or Asian (46.8 weeks) immigrants, although greater than among Latin American immigrants (45.3 weeks). Among those in the labor force in the reference week, the last week in March 2000, 4.3 percent of the Soviet Jewish immigrants were unemployed, in contrast to 2.8 percent and 3.1 percent for European and Asian immigrants, respectively.

Table 13.4 provides greater detail on the English language proficiency of immigrants. The Soviet Jews are least likely to speak only English at home (4.5 percent compared to 13.4 percent for all immigrants) and are more likely (26.8 percent) than European and Asian immigrants to report that they speak English "not well" or "not at all" (11.2 and 17.9 percent, respectively). Only the Latin American immigrants have a greater proportion (42.7 percent) in these two least proficient categories.

Appendix Tables 13.A1–13.A3 report the ethnic ancestry, language spoken at home if it is not exclusively English and the republic of birth for the sample

Table 13.4 Fluency in English among adult male immigrants who immigrated since 1965 by region of origin (percent)[a].

English Fluency	*FSU*[b]	*Europe (excluding FSU)*	*Asia*	*Latin America*	*All*
Speaks only English at home	4.5	32.3	7.4	10.7	13.4
Speaks another language at home and speaks English:					
Very well	30.1	36.6	45.3	22.0	31.1
Well	37.8	19.9	29.4	24.6	25.1
Not well	22.5	9.6	15.4	28.2	21.4
Not at all	4.3	1.6	2.5	14.5	9.0
Total	100.0	100.0	100.0	100.0	100.0
Sample size	8,373	42,590	1,24,735	2,50,826	4,51,844

Notes: Detail may not add to total due to rounding.

[a] All immigrants include groups not shown separately.

[b] FSU excludes persons of Armenian ancestry and persons who speak Armenian or Ukrainian at home.

Source: 2000 Census of Population, Public Use Microdata Sample, 5 percent sample.

of Soviet Jews under study by sub-period of immigration to the US since 1965. There appears to be relatively little variation in these characteristics across the sub-periods.[5]

Methodology for the statistical analysis

A multivariate statistical analysis (ordinary least-squares regression analysis, OLS) is used to compare Soviet Jewish immigrants to other immigrants, when other measured variables are held constant. That is, controlling for factors such as age, schooling, marital status, and duration in the United States, do Soviet Jews differ in English language proficiency and earnings from other immigrants?[6]

The statistical analysis uses the adult (aged 25–64) male respondents in the 2000 Census Public Use Microdata Sample, 5 percent sample of the population, as the unit of observation. The means and standard deviations for the dependent and explanatory variables are reported in Table 13.3.

Language skills are measured by a dichotomous variable defined to equal unity for those who speak only English at home or if they speak another language they speak English "very well" or "well". It is zero for those who speak English "not well" or "not at all" (see Table 13.4).

The earnings variable is the natural logarithm of annual earnings in 1999, where earnings are the sum of wage, salary, and self-employment income. Those who reported zero earnings or did not work in 1999 are deleted from the analysis. Those who reported earnings of less than \$100, including the negligible number reporting negative earnings, were assigned a value of \$100 since the natural logarithm is not defined for zero or negative values.[7]

The econometric model for the analysis of language proficiency is based on earlier research that specifies three fundamental concepts (Chiswick & Miller, 1998). These are exposure to the destination language, efficiency in destination language acquisition, and economic incentives for learning the destination language. In the empirical application the measurable variables reflecting these concepts include two continuous variables, years of schooling and years of age, and a set of dichotomous variables. The dichotomous variables include marital status (whether married, with spouse present), whether there are children under age 18 currently living in the household, and whether the respondent lives in a rural area or a southern state (the swath of 17 states from Texas to the Atlantic Ocean, from Maryland to Florida, including Washington, DC).[8]

The Census asks, when did this person come to the United States to stay? The Census does not ask the type of visa used to enter the United States or whether permanent resident status was obtained. Given that many Soviet Jews entered the United States as asylees only to become permanent resident aliens (immigrants) at a later date, the Census question is more appropriate for this analysis than would be the year the respondent obtained permanent resident alien or immigrant status. Since few Soviet Jews subsequently left the United States to return to the FSU or go to a third country, such as Israel, the emigration from the United States of Soviet immigrants does not pose a selectivity problem (Ahmed & Robinson, 1994; Mulder, 2003).[9]

Variables for duration in the United States are central to the analysis and they are entered as period of arrival dichotomous variables.[10] This specification was chosen to permit a finer determination of non-linearities than would a quadratic specification of a continuous duration variable. Moreover, it increases comparability with earlier research on Soviet Jews in the United States. When duration is held constant, the age variable reflects the effect of age at migration on English language proficiency.

Another key variable is country of birth. A person born in any of the republics that constituted the FSU (other than those who reported Armenian ancestry or language or the Ukrainian language) is considered to be a Soviet Jewish immigrant (FSU).[11] Data are not available on when the person left the FSU or on country of last permanent residence, so it is not possible to identify whether there was a destination prior to coming to the US. In this analysis, the country categories Europe and Asia constitute all of Europe and Asia, other than the designated parts of the FSU. Other countries of origin groups are Canada, Latin America (including the Caribbean), and other countries (Africa, Oceania, etc.). Europe other than the FSU serves as the benchmark.

The econometric analysis of earnings is based on the human capital earnings function, modified for immigrant adjustment (Chiswick, 1978). The natural logarithm of annual earnings in 1999 is regressed on years of schooling completed, years of potential labor market experience (age minus schooling minus 5 years), and its square, the natural logarithm of weeks worked, and dichotomous variables as defined above for being proficient in English, married spouse present, living in a rural area and living in a southern state. The same dichotomous variables are used, as defined above, for period of arrival and country of origin. Controlling for period of arrival, the labor market experience variable measures the effect on earnings in 1999 of experience in the country of origin.

Econometric analysis

Language—Soviet and other immigrants

The results of the multiple regression analysis for adult males for the dependent variable, proficient in English, are reported in Tables 13.5 and 13.6. The variable is unity for those who speak only English at home or who speak another language but speak English very well or well, otherwise the English fluency variable is zero.

Table 13.5 reports the equation for all immigrants by sub-period and for

Table 13.5 Regression analysis of fluency in English among adult males who immigrated since 1965: 2000.

Dependent variable = ENGSPK	*2000 Census*							
Immigration period	*1965–2000*		*1965–1979*		*1980–1989*		*1990–2000*	
Variable	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
CONSTANT	0.7238 (169.12)	0.7241 (169.07)	0.7980 (114.50)	0.7980 (114.50)	0.7618 (94.72)	.07615 (94.69)	0.6048 (89.67)	0.6063 (89.89)
EDUCYRS	0.0349 (255.53)	0.0349 (255.37)	0.0335 (150.77)	0.0335 (150.74)	0.0363 (147.89)	0.0363 (147.87)	0.0342 (139.85)	0.0342 (139.83)
AGE	−0.0067 (−99.89)	−0.0067 (−100.05)	−0.0068 (−61.96)	−0.0068 (−61.94)	−0.0079 (−62.49)	−0.0079 (−62.50)	−0.0056 (−49.66)	−0.0057 (−49.82)
IM95_00	−0.2405 (−120.21)	−0.2380 (−118.47)					−0.1041 (−50.05)	−0.1006 (−47.84)
IM90_94	−0.1396 (−68.98)	−0.1415 (−69.46)					[a]	[a]
IM85_89	−0.0631 (−31.99)	−0.0637 (−32.24)			−0.0614 (−28.54)	−0.0617 (−28.60)		
IM75_79	0.0573 (26.28)	0.0571 (26.08)	−0.0382 (−17.28)	−0.0381 (−17.20)				
IM70_74	0.0958 (39.38)	0.0959 (39.38)	[a]	[a]				
IM65_69	0.1230 (42.54)	0.1232 (42.57)	0.0381 (14.19)	0.0382 (14.21)				
IM95_00*FSUJEW		−0.0908 (−7.64)						−0.1118 (−10.15)

IM90_94*FSUJEW	0.0502							
		(4.46)						
IM85_89*FSUJEW	0.0574				0.0494			
		(3.70)				(1.89)		
IM75_79*FSUJEW		0.0182		−0.0202				
		(1.10)		(−0.89)				
IM70_74*FSUJEW	−0.0021		[a]					
		(−0.06)						
IM65_69*FSUJEW		0.0066		−0.0436				
		(0.12)		(−0.87)				
MARRSP	0.0409	0.0410	0.0220	0.0220	0.0264	0.0264	0.0537	0.0539
	(29.96)	(30.09)	(9.53)	(9.53)	(10.53)	(10.53)	(23.08)	(23.15)
RURAL	0.0082	0.0082	0.0139	0.0139	0.0142	0.0142	.0075	0.0075
	(1.55)	(1.56)	(1.52)	(1.52)	(1.47)	(1.48)	(0.86)	(0.85)
SOUTH	0.0081	0.0080	0.0094	0.0094	0.0166	0.0166	0.0041	0.0041
	(6.08)	(6.02)	(4.25)	(4.25)	(6.83)	(6.82)	(1.83)	(1.80)
CHILD	−0.0257	−0.0255	−0.0086	−0.0086	−0.0201	−0.0201	−0.400	0.0398
	(−19.19)	(−19.05)	(−3.94)	(−3.94)	(−7.83)	(−7.84)	(−17.64)	(−17.52)
FSU	−0.1043	−0.1050	−0.0276	−0.0118	−0.0390	−0.0555	−0.1733	0.1354
	(−23.85)	(−12.44)	(−2.69)	(−0.61)	(−3.07)	(−3.61)	(−29.65)	(−19.53)
ASIA	−0.0766	−0.0764	−0.0658	−0.0660	−0.0907	−0.0902	−0.0804	0.0839
	(−35.26)	(−34.58)	(−20.19)	(−20.19)	(−20.10)	(−19.94)	(−21.72)	(−22.58)
LATAMER	−0.1725	−0.1723	−0.0911	−0.0913	−0.1622	−0.1618	−0.2562	0.2588
	(−79.66)	(−78.95)	(−29.53)	(−29.54)	(−35.95)	(−35.81)	(−66.94)	(−67.49)
CANADA	0.0705	0.0704	−0.0015	−0.0017	0.0397	0.0401	0.1263	0.1234
	(14.45)	(14.42)	(−0.21)	(−0.23)	(3.57)	(3.61)	(15.99)	(15.62)

(Continued Overleaf)

Table 13.5 Continued

Dependent variable = ENGSPK	*2000 Census*							
Immigration period	*1965–2000*		*1965–1979*		*1980–1989*		*1990–2000*	
Variable	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
OTHER	0.0584	0.0584	−0.0052	−0.0054	0.0195	0.0199	0.0882	0.0854
	(16.59)	(16.53)	(−0.86)	(−0.89)	(2.79)	(2.85)	(15.62)	(15.11)
SAMPLE SIZE	4,51,843	4,51,843	1,24,512	1,24,512	1,40,887	1,40,887	1,66,684	1,66,684
STANDARD ERROR	0.3925	0.3924	0.3344	0.3344	0.3983	0.3983	0.4178	0.4176
R^2	0.2723	0.2726	0.2586	0.2586	0.2299	0.2299	0.2739	0.2744
ADJUSTED R^2	0.2722	0.2725	0.2585	0.2585	0.2298	0.2298	0.2739	0.2743

Notes: t-ratios in parentheses.
[a] Omitted as benchmark; 1980–1984 and Europe are benchmarks unless otherwise noted.

Source: 2000 US Census of Population, Public Use Microdata Sample, 5 percent sample.

Table 13.6 Regression analysis of English fluency among adult males who immigrated since 1965: 2000, 1990, 1980.

Dependent variable = ENGSPK					
Immigration period	*2000 Census 1965–2000*		*1990 Census 1965–1989*		*1980 Census 1965–1979*
Variable	(1)	(2)	(1)	(2)	(1)
CONSTANT	0.7177 (74.75)	0.7188 (74.87)	0.6031 (114.99)	0.6027 (114.94)	0.5243 (41.60)
EDUCYRS	0.0347 (113.15)	0.0347 (113.12)	0.0356 (192.66)	0.0355 (192.59)	0.0388 (89.92)
AGE	–0.0066 (–44.38)	–0.0067 (–44.49)	–0.0067 (–73.07)	–0.0067 (–73.30)	–0.0050 (–20.26)
IM96_00	–0.2492 (–52.88)	–0.2443 (–51.50)			
IM91_95	–0.1626 (–35.26)	–0.1639 (–35.27)			
IM87_90	–0.0876 (–18.99)	–0.0891 (–19.23)	–0.1387 (–51.87)	–0.1341 (–49.77)	
IM85_86	–0.0347 (–6.07)	–0.0345 (–6.04)	–0.0698 (–23.07)	–0.0693 (–22.85)	
IM75_79	0.0544 (11.07)	0.0546 (11.06)	0.0796 (33.12)	0.0789 (32.64)	–0.0956 (–19.20)
IM70_74	0.1069 (19.64)	0.1073 (19.68)	0.1312 (50.16)	0.1316 (50.21)	*
IM65_69	0.1222 (19.04)	0.1222 (19.02)	0.1690 (58.01)	0.1695 (58.13)	0.0574 (11.18)
IM96_00*FSUJEW		–0.1580 (–6.36)			
IM91_95*FSUJEW		–0.0147 (–0.53)			
IM87_90*FSUJEW		0.03281 (1.00)		–0.2527 (–10.31)	
IM85_86*FSUJEW		0.0036 (0.04)		–0.0456 (–0.85)	
IM75_79*FSUJEW		–0.0386 (–0.99)		–0.0249 (–1.00)	
IM70_74*FSUJEW		–0.0421 (–0.58)		–0.0346 (–0.87)	
IM65_69*FSUJEW		0.0514 (0.57)		0.0263 (0.45)	
MARRSP	0.0397 (12.98)	0.0399 (13.05)	0.0404 (20.53)	0.0409 (20.78)	0.0134 (2.17)
RURAL	0.0188 (1.58)	0.0189 (1.59)	0.0177 (4.55)	0.0177 (4.55)	0.0102 (1.16)
SOUTH	0.0083 (2.79)	0.0083 (2.79)	0.0174 (8.99)	0.0175 (9.03)	0.0030 (0.60)

(Continued Overleaf)

Table 13.6 Continued

Dependent variable = ENGSPK					
Immigration period	*2000 Census 1965–2000*		*1990 Census 1965–1989*		*1980 Census 1965–1979*
Variable	(1)	(2)	(1)	(2)	(1)
CHILD	–0.0149 (–4.98)	–0.0147 (–4.88)	–0.0267 (–14.16)	–0.260 (–13.80)	0.0093 (1.98)
FORMAR[a]					–0.0431 (8.32)
FSU	–0.0979 (–10.00)	–0.0492 (–2.18)	–0.1374 (–16.05)	–0.0273 (–1.31)	–0.1384 (–7.47)
ASIA	–0.0739 (–15.05)	–0.0754 (–15.20)	–0.0632 (–22.43)	–0.0631 (–22.43)	–0.0431 (–6.97)
LATAMER	–0.1768 (–36.33)	–0.1777 (–36.38)	–0.1514 (–54.87)	–0.1514 (–54.88)	–0.1445 (–25.13)
CANADA	0.0825 (7.73)	0.0812 (7.60)	0.0739 (11.05)	0.0739 (11.06)	0.1265 (9.26)
OTHER	0.0621 (7.91)	0.0601 (7.73)	–0.0228 (–6.10)	–0.0227 (–6.08)	0.0202 (2.41)
Sample size	90,383	90,383	2,27,554	2,27,554	35,915
Standard error	0.39307	0.3929	0.3879	0.3877	0.3790
R^2	0.2711	0.2718	0.2649	0.2656	0.3047
Adjusted R^2	0.2710	0.2716	0.2649	0.2655	0.3044

Notes: *t*-ratios in parentheses.
[a] Variable cannot be reconstructed for 1990, 2000 Censuses.
* Omitted as benchmark; benchmark is 1980–1984 and Europe unless otherwise noted.

Source: 2000 Census of Population, Public Use Microdata Sample, 5 percent sample; 1990 Census of Population, Public Use Microdata Sample, 5 percent sample; 1980 Census of Population, Public Use Sample, B and C Sample Files Combined, 2 percent sample.

the whole period 1965–2000. As shown in column (1), consistent with what has been found elsewhere for immigrants, English language proficiency increases with years of schooling (3.5 percentage points more are proficient for each extra year of schooling). Proficiency is lower for those who immigrated at an older age. Five years older at immigration is equivalent to about one fewer year of schooling. Men who are married are more proficient (by 4 percentage points), but children at home detracts from their proficiency (by 2.6 percentage points per child).

Duration in the United States has a major impact on English language proficiency. The coefficients are highly statistically significant and show a consistent gradient of increased proficiency with duration in the US, with the effect of an extra year in the United States becoming smaller the longer the duration of residence. With those who immigrated in 1980–1984 as the benchmark, other variables the same, the most recent immigrants (1996–2000) were 24 percentage points less proficient in 2000, or the equivalent of the effect of seven years of schooling. The earliest cohort, 1965–1969 immigrants, was 12 percentage points more proficient than the 1996–2000 cohort or the equivalent of 3.5 years of schooling.

Other variables the same, Soviet immigrants are about 10.4 percentage points less likely to be proficient in English than other European immigrants. They are even less proficient than Asian immigrants (Asians are at a 7.7 percentage points disadvantage compared to European immigrants), but less disadvantaged than those from Latin America (17.3 percentage point differential compared to European immigrants).

It is possible to test whether the effect of duration in the US on proficiency in English differs between Soviet and other immigrants. The statistical analysis (Table 13.5, column 2) shows that during the first four years the negative effect on proficiency of being an immigrant is much greater for Soviet Jews than it is for other immigrants. Compared to other recent European immigrants, Soviet Jews who arrived in 1995–2000 are 20 percentage points less proficient (–0.1050 – 0.0908 = –0.20). Soviet Jews experience a steeper improvement in proficiency with duration in the US so that the disadvantage is only 5.5 percentage points –0.1050 + 0.0502 = –0.055) for those who immigrated in 1990–1994 (6–10 years in the US), and 4.8 percentage points (–0.1050 + 0.0574 = –0.048) for those who immigrated 1985–1989 (11–15 years in the US). Indeed, the very large proportion of Soviet immigrants in the US a short period of time and the very low English proficiency of this group are very important determinants of the overall low proficiency among Soviet immigrants.

The analysis was also performed for sub-periods within the 1965–2000 period (Table 13.5). For each of these sub-periods the effects of schooling, age at immigration, marital status, and children are quite similar.[12] That is, their partial effects on proficiency in 2000 do not appear to vary by period of immigration. The effects of duration do vary by period of immigration. One fewer year in the US has a larger negative effect on proficiency the more recently the immigrant cohort arrived in the US, which is consistent with the non-linear effect of duration on proficiency.

The results reported here for the 2000 Census can be compared with analyses reported previously for Soviet Jews and other immigrants who came to the US in 1965 or later using the microdata files from the 1980 and 1990 Censuses (Chiswick, 1993, 1997) (see Table 13.6). The effects on English language proficiency of schooling, age, marital status, and rural residence are virtually identical across the three censuses, although the positive effect of being married was much smaller in the 1980 Census and the positive effects of living in the South is smaller in 2000 than in 1990.[13] The negative effects of children in the household are also smaller in absolute value in 2000 than in 1990, but it was not significant in 1980. The strong positive effect of duration in the US on proficiency is also observed in these earlier censuses.

The 10 percentage points disadvantage of being from the FSU compared to another part of Europe in the 2000 data is somewhat smaller than the 14 percentage points in the 1980 and 1990 Censuses. When the interaction terms of Soviet origin with duration are added, the Soviet intercept is a highly significant –5 percentage points in 2000, compared to a non-significant +3 percentage points in 1990. The negative effect of being in a particular immigrant cohort compared to an earlier arrival cohort diminishes from the 1980 to the 2000 Census as the cohorts are in the US a longer period of time. Most striking is that in 1990, the only Soviet-duration of residence interaction term whose coefficient was large or statistically different from the benchmark (1980–1984) was the most recent cohort, 1987–1990 (coefficient of –25 percentage points). Ten years later, compared to the same benchmark, the 1987–1990 interaction term has a coefficient of only 3 percentage points and it is not statistically significant.

These results suggest that the sharp gradient of English language proficiency with duration in the US is not a consequence of declining proficiency among more recent cohorts. Rather it appears to be reflecting a longitudinal or adjustment effect, that is, the acquisition of English language proficiency as a cohort has a longer duration in the US. Moreover, this initial deficiency and speed of adjustment (improvement) appear to be more intense for Soviet Jews than for other immigrants. This may reflect their refugee motivated migration, the limited ability to prepare for the emigration because of the arbitrary nature of the Soviet bureaucracy, and the unexpected and sudden opening for emigration from the Soviet Union, with uncertainty as to how long emigration would be possible.

Earnings—Soviet and Other Immigrants

The analysis of earnings (Table 13.7, column 1) indicates that an extra year of schooling raises the earnings of immigrants by about 4.6 percent, that earnings increase at a decreasing rate with an increase in total labor market experience, that earnings rise by about 0.85 percent for each one percent increase in weeks worked (about one half of a week), and that earnings increase with duration of residence in the US. Indeed, compared to those who immigrated in 1980–1984, those who recently arrived (immigrated 1996–2000) have about 16 percent lower weekly earnings, while those who immigrated in 1965–1969 had about 11 percent higher weekly earnings.

The effects of country of origin are quite large. Compared to European immigrants, those from the Soviet Union had weekly earnings that were nearly 20 percent lower, other measured variables being the same. Only Latin American immigrants had a larger earnings disadvantage (about 32 percent) compared to those from Europe, while Canadian immigrants showed a large earnings advantage over Europeans (about 13 percent). Other factors that resulted in higher earnings are being proficient in English (about 17 percent),

Table 13.7 Regression analysis of earnings among adult males who immigrated since 1965, 2000.

Dependent variable = LNEARN

	2000 Census							
Immigration period	*1965–2000*		*1965–1979*		*1980–1989*		*1990–2000*	
Variable	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
CONSTANT	6.248 (451.41)	6.248 (451.29)	6.030 (214.00)	6.030 (213.99)	6.581 (280.50)	6.578 (280.29)	6.143 (291.94)	6.142 (291.77)
EDUCYRS	0.0461 (130.43)	0.0458 (129.27)	0.0547 (75.35)	0.0546 (75.03)	0.0438 (77.98)	0.0437 (77.63)	0.0415 (70.28)	0.0412 (69.59)
EXP	0.0108 (22.35)	0.0108 (22.40)	0.0113 (11.34)	0.0113 (11.37)	0.0127 (14.85)	0.0127 (14.84)	0.0084 (10.62)	0.0084 (10.50)
EXPSQ	−0.00017 (−19.61)	−0.00017 (−19.78)	−0.00014 (−8.20)	−0.00014 (−8.26)	−0.00022 (−14.06)	−0.00022 (−14.09)	−0.00016 (−10.21)	−0.00016 (−10.19)
LNWW	0.8461 (314.94)	−0.8459 (314.93)	0.8656 (151.51)	0.8656 (151.51)	0.7757 (170.96)	0.7757 (170.96)	0.8879 (218.39)	0.8876 (218.35)
IM95_00	−0.1554 (−35.35)	−0.1503 (−34.06)					−0.0530 (−12.56)	−0.0483 (−11.29)
IM90_94	−0.1041 (−24.20)	−0.1027 (−23.71)						
IM85_89	−0.0434 (−10.46)	−0.0434 (−10.44)			−0.0459 (−11.26)	−0.0457 (−11.16)		
IM75_79	0.0562 (12.25)	0.0561 (12.18)	−0.0388 (−6.82)	−0.0392 (−6.88)				
IM70_74	0.0965 (18.68)	0.0986 (19.04)	[a]	[a]				
IM65_69	0.1184 (19.18)	−0.1214 (19.64)	0.0270 (3.87)	0.0277 (3.96)				

(Continued Overleaf)

Table 13.7 Continued

Dependent variable = LNEARN	*2000 Census*							
Immigration period	*1965–2000*		*1965–1979*		*1980–1989*		*1990–2000*	
Variable	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
IM95_00*FSUJEW		–0.3447						–0.1395
			(–9.29)					(–5.85)
IM90_94*FSUJEW		–0.2299						
		(–6.36)						
IM85_89*FSUJEW		–0.1048				–0.0306		
		(–2.49)				(–0.68)		
IM75_79*FSUJEW		–0.1148		0.0101				
		(–2.61)		(0.13)				
IM70_74*FSUJEW		–0.2233		[a]				
		(–2.80)						
IM65_69*FSUJEW		–0.2416		–0.1297				
		(–2.00)		(–0.94)				
ENGSPK	0.1742	0.1738	0.1371	0.1373	0.1643	0.1645	0.2018	0.2007
	(54.05)	(53.92)	(18.18)	(18.21)	(32.43)	(32.48)	(40.27)	(40.04)
MARRSP	0.2115	0.2121	0.2661	0.2659	0.2168	0.2168	0.1699	0.1707
	(78.13)	(78.35)	(47.16)	(47.14)	(49.57)	(49.58)	(38.95)	(39.13)
RURAL	–0.0804	–0.0807	–0.1349	–0.1348	–0.0536	–0.0536	–0.0740	–0.0741
	(–7.10)	(–7.12)	(–5.64)	(–5.64)	(–2.91)	(–2.91)	(–4.16)	(–4.17)
SOUTH	–0.0288	–0.0292	–0.0551	–0.0552	–0.0230	–0.0231	–0.0158	–0.0162
	(–10.25)	(–10.37)	(–9.67)	(–9.68)	(–5.02)	(–5.03)	(–3.49)	(–3.58)
FSU	–0.1850	–0.3106	–0.0309	–0.1983	–0.1178	–0.3136	–0.2545	–0.3448
	(–19.46)	(–15.30)	(–1.17)	(–3.17)	(–5.76)	(–7.36)	(–21.26)	(–14.39)

ASIA	−0.1673	−0.1608	−0.0521	−0.0492	−0.2302	−0.2232	−0.2280	−0.2198
	(−36.48)	(−34.46)	(−6.23)	(−5.84)	(−27.39)	(−26.24)	(−30.89)	(−28.93)
LATAMER	−0.3247	−0.3211	−0.2137	−0.2122	−0.3838	−0.3789	−0.3927	−0.3881
	(−70.91)	(−69.62)	(−26.89)	(−26.65)	(−45.74)	(−44.86)	(−51.24)	(−50.10)
CANADA	0.1327	−0.1371	0.0316	0.0333	0.1309	0.1366	0.2070	0.2131
	(−13.09)	(13.50)	(1.76)	(1.85)	(6.31)	(6.57)	(13.47)	(13.82)
OTHER	−0.2071	−0.2025	−0.0870	−0.0849	−0.2181	0.2125	−0.3113	−0.3054
	(−27.88)	(−27.16)	(−5.62)	(−5.47)	(−16.73)	(16.25)	(−27.26)	(−26.93)
FSU*EDUCYRS		0.0244		0.0129		0.0174		0.0112
		(11.02)		(2.29)		(5.01)		(6.31)
Sample size	3,98,520	3,98,520	1,10,840	1,10,840	1,45,315	1,45,315	1,42,363	1,42,363
Standard error	0.7833	0.7831	0.8121	0.8120	0.7621	0.7769	0.7774	0.7772
R^2	0.3580	0.3583	0.3225	0.3226	0.3176	0.3217	0.3886	0.3888
Adjusted R^2	0.3580	0.3583	0.3224	0.3225	0.3175	0.3213	0.3885	0.3887

Notes: t-ratios in parentheses.
[a] Omitted as benchmark; benchmark is 1980–1984 and Europe unless otherwise noted.

Source: 2000 Census of Population, Public Use Microdata Sample, 5 percent sample.

being married (21 percent), living in an urban area (8 percent), and living outside the south (3 percent).

Other variables the same, as shown in Table 13.7, column 2, an extra year of schooling is associated with 7.0 percent higher earnings for the Soviet Jewish immigrants, in contrast to the 4.6 percent for other immigrants, and the difference is highly statistically significant (t = 11.0). Also, other things the same, the earnings of Soviet Jewish immigrants are much lower (and the difference is highly significant) than those of other immigrants who came in the same time period during the first few years in the US (immigrated 1996–2000 or 1991–1995). The magnitude diminishes but does not disappear for those who have been in the United States for 10 or more years in 2000.

Thus, the earnings gap between Soviet and other immigrants varies with duration in the US and level of schooling. At the mean level of schooling of Soviet immigrants (14.8 years), those who immigrated in 1980–1984 (16–20 years in the US) had about 5 percent higher weekly earnings than other European immigrants (the partial effect is: –0.3106 + (14.8)(0.0244) = 0.051).

The comparison of these results with the 1990 and 1980 Census analyses is striking (Chiswick, 1997) (Table 13.8). In 1990, the effect of schooling on earnings was larger for Soviet Jewish immigrants by 1.9 percentage points, in 1980 by 2.8 percentage points, and in 2000 by 2.6 percentage points, all of which were significantly different from zero, but not from each other. In 1990, the Soviet immigrant duration of residence interaction term for the most recent arrivals was large and highly significant compared to the benchmark (1980–1984 cohort), as was the case in 1980 (1970–1974 benchmark), but the differential shrank with duration. Although only in the US 6–10 years at the time of the 1990 Census, at the mean level of schooling for Soviet immigrants (14.9 years), the earnings of the 1980–1984 cohort of Soviet Jews was only 1 percent lower than that of other European immigrants. As in the 2000 Census, the larger return from schooling narrowed the earnings gap between Soviet Jews and other immigrants in spite of a larger initial earnings disadvantage.

Among the Soviet immigrants (Table 13.8), the 31 percent greater earnings disadvantage of the 1987–1990 cohort compared to the 1984–1985 cohort in 1990, shrank to a marginally significant (t = 1.6) 14 percent disadvantage 10 years later in 2000. This too suggests that what is being observed is an immigrant assimilation process rather than a change (deterioration) in the earnings potential of more recent cohorts.

For most of the other explanatory variables their partial effects on earnings did not change by much across the three censuses. Perhaps the most dramatic change is the increase in the negative effects of living in a rural area. This may be due to the change in the definition of rural from the old census definition of rural (farm and non-farm) to only those living on a farm.

Moreover, the lower initial earnings and the steeper rise in earnings with duration of residence in the US of the Soviet Jewish immigrants,

Table 13.8 Regression analysis of earnings among adult males who immigrated since 1965: 2000, 1990, 1980.

Dependent variable = LNEARN	*2000 Census*		*1990 Census*		*1980 Census*
Immigration period	*1965–2000*		*1965–1979*		*1980–1989*
Variable	(1)	(2)	(1)	(2)	(1)
CONSTANT	6.198	6.201	5.204	5.208	4.360
	(200.12)	(200.29)	(303.00)	(303.01)	(102.18)
EDUCYRS	0.04507	0.0447	0.0480	0.0479	0.0462
	(57.14)	(56.49)	(103.10)	(102.62)	(40.89)
EXP	0.0102	0.0103	0.0268	0.0267	0.0300
	(9.40)	(9.43)	(42.81)	(42.71)	(19.77)
EXPSQ	−0.0002	−0.0002	−0.0004	−0.0004	−0.0005
	(−8.72)	(−8.83)	(−34.33)	(−34.26)	(−16.88)
LNWW	0.8679	0.8675	0.9534	0.9526	1.048
	(145.45)	(145.44)	(270.03)	(269.66)	(114.66)
IM96_00	−0.1598	−0.1501			
	(−15.34)	(−14.37)			
IM91_95	−0.1229	−0.1205			
	(−12.53)	(−12.19)			
IM87_90	−0.0692	−0.0705	−0.0949	−0.0910	
	(−7.13)	(−7.23)	(−16.32)	(−15.57)	
IM85_86	−0.0448	−0.0429	−0.0708	−0.0698	
	(−3.75)	(−3.59)	(−11.46)	(−11.28)	
IM75_79	0.0561	0.0557	0.1062	0.1069	−0.1345
	(5.42)	(5.360)	(21.85)	(21.90)	
IM70_74	0.0913	0.0931	0.1787	0.1797	[a]
	(7.92)	(8.06)	(33.67)	(33.78)	(−13.12)
IM65_69	0.1124	0.1152	0.1996	0.2006	0.0804
	(8.17)	(8.36)	(33.64)	(33.76)	(7.60)
IM96_00*FSUJEW	−0.5036				
	(−6.39)				
IM91_95*FSUJEW	−0.3033				
	(−3.97)				
IM87_90*FSUJEW	−0.1352		−0.3090		
	(−1.60)		(−5.65)		
IM85_86*FSUJEW	−0.3156		−0.1979		
	(−1.73)		(−1.79)		
IM75_79*FSUJEW	−0.0911		−0.1458		
	(−0.95)		(−2.85)		
IM70_74*FSUJEW	−0.0100		−0.1137		
	(−0.06)		(−1.40)		

(Continued Overleaf)

Table 13.8 Continued

Dependent variable = LNEARN	*2000 Census*		*1990 Census*		*1980 Census*
Immigration period	*1965–2000*		*1965–1979*		*1980–1989*
Variable	(1)	(2)	(1)	(2)	(1)
IM65_69*FSUJEW	–0.1193		0.0029		
	(–0.56)		(0.02)		
ENGSPK	0.1755	0.1743	0.1723	0.1717	0.1632
	(24.40)	(24.24)	(39.38)	(39.25)	(14.84)
MARRSP	0.2022	0.2030	0.2093	0.2099	0.1718
	(33.38)	(33.52)	(57.12)	(57.26)	(16.11)
RURAL	–0.1159	–0.1163	–0.0183	–0.0186	–0.0190
	(–4.51)	(–4.53)	(–2.36)	(–2.40)	(–1.03)
SOUTH	–0.0386	–0.0389	–0.0925	–0.0925	–0.0312
	(–6.13)	(–6.19)	(–23.52)	(–23.51)	(–2.97)
FSU	–0.1937	–0.2899	–0.1759	–0.3021	0.0895
	(–9.11)	(–5.11)	(–9.36)	(–3.42)	(–0.66)
ASIA	–0.1592	–0.1559	–0.1955	–0.1953	–0.1862
	(–15.35)	(–14.89)	(–34.34)	(–34.32)	(–14.44)
LATAMER	–0.3224	–0.3216	–0.3227	–0.3231	–0.2612
	(–31.39)	(–31.20)	(–57.83)	(–57.90)	(–21.63)
CANADA	0.1760	0.1777	0.0936	0.0937	0.1375
	(7.92)	(8.00)	(6.99)	(7.00)	(4.83)
OTHER	–0.2180	–0.2159	–0.2511	–0.2511	–0.2276
	(–13.14)	(–12.99)	(–32.60)	(–32.60)	(–13.03)
FSU*EDUCYRS		0.0256		0.0194	–0.0280
		(5.41)		(3.79)	(–3.08)
Sample size	79,582	79,582	2,02,113	2,02,113	35,915
Standard error	0.7830	0.7827	0.7456	0.7455	0.7898
R^2	0.3646	0.3652	0.4267	0.4268	0.3895
Adjusted R^2	0.3645	0.3650	0.4266	0.4268	0.3892

Notes: t-ratios in parentheses.

[a] Omitted as benchmark; 1980–1984 and Europe are benchmarks unless otherwise noted.

Source: 2000 Census of Population, Public Use Microdata Sample, 5 percent sample; 1990 Census of Population, Public Use Microdata Sample, 5 percent sample; 1980 Census of Population, Public Use Sample, B and C Sample Files Combined, 2 percent sample.

compared with other immigrant groups, is a phenomenon to be expected among refugee populations. Since their motives for migrating are not strictly economic, refugees tend to be less prepared for the move, especially Soviet migrants, and to have skills that are less readily transferable to the destination.

Language and earnings—Soviet Jewish immigrants

Parallel analyses to those reported above were performed separately for just the Soviet Jewish immigrants (Appendix Tables 13.A5 and 13.A6). The statistical significance of many of the variables is reduced because of the much smaller sample size. Of particular interest is whether there are differences among Soviet immigrants depending on their reported ethnic ancestries. Excluding those of Armenian ancestry or language and Ukrainian language, four groups are defined, Russian (53 percent of the sample), Ukrainian (18 percent), a response that revealed a person's religion (12 percent), and all other responses (17 percent). Those of Russian ancestry serve as the benchmark. The coefficients and significance levels of the other variables do not change when the ethnic ancestry variables are entered into the equation.

In the language analysis, other variables being the same, few differences are found in English language proficiency by ancestry (Appendix Table 13.A5). Those of Ukrainian origin are 2 percentage points less proficient in English than those of Russian ancestry, but the difference is at the margin of being significant ($t = 1.7$). Those of "other ancestries" are one percentage point less proficient than the Russians, but this is not statistically significant ($t = 0.7$). There is no difference from those of Russian ancestry among those who gave a response indicating their religion (the coefficient indicates a 2.5 percentage point higher proficiency with $t = 1.8$).

The analysis of earnings, other variables being the same, presents a similar picture (Appendix Table 13.A6). For the post-1965 immigrants, there is no difference in earnings between the Russian, Ukrainian, and religious revealing ancestries. Compared to the Russians, the Ukrainians had 2.1 percent lower earnings, but a $t = -0.7$, while those who gave a religious response had 4.5 percent higher earnings, but a $t = 1.3$. Only the heterogeneous group of "other ancestries" showed an earnings differential, a marginally significant ($t = 1.9$) 5.5 percent higher earnings.

The coefficient on the education variable in the earnings analysis limited to Soviet Jewish immigrants is about 7.3 percent, whether or not the Soviet ancestry variables are held constant. This is a very large coefficient for an immigrant population in the United States and is significantly greater than for other immigrants. That it does not change when ancestry is held constant suggests that it holds across the ancestry groups that in this study are used to identify Soviet Jews.

Summary and conclusion

This chapter is concerned with the English language proficiency and labor market earnings of adult (aged 25–64 years) male Soviet Jews who immigrated to the United States since 1965. The data for the empirical analysis are from the 2000 Census of Population, Public Use Microdata Sample, and is for a five percent sample of the population. Comparisons are made to earlier parallel analyses using the 1980 and 1990 Censuses. Because of the absence of direct information on who is Jewish or of Jewish ancestry, the empirical analysis is based on persons born in the FSU who are not of Armenian ancestry and do not speak Armenian or Ukrainian at home. This definition should capture most Soviet Jews but include few non-Jewish immigrants from the FSU.

The Soviet Jews were less proficient in English than other European and Asian immigrants. Under the definition of proficiency used in this study, 72 percent of the Soviet Jews were proficient, compared to 89 percent for European immigrants, 82 percent for Asian immigrants, and 57 percent for those from Latin America. Their earnings (at $37,600 in 1999) were considerably less than the earnings of other European ($50,900) and Asian immigrants ($42,400), but were greater than the earnings of Latin American immigrants ($23,000).

The much higher level of schooling of the Soviet immigrants would tend to enhance their English proficiency and earnings; 14.8 years for the Soviet Jews, compared to 14.1 for Asian immigrants, 13.6 years for European immigrants, and 9.3 years for Latin American immigrants. On the other hand, the refugee motivations for their move and their recency of arrival would tend to lower their English language skills and earnings. Among those who immigrated since 1965, 70 percent of the Soviet Jewish migrants were in the United States 10 or fewer years, compared to only 37 percent of those from Europe, 37 percent of the Asians, and 35 percent of the Latin Americans.

Multiple regression analysis is used to examine the effects of being a Soviet Jewish immigrant compared to coming from another region, when all other measured variables are held constant. It is found that recently arrived Soviet immigrants have a lower level of English proficiency than other European immigrants, but they have a faster rate of improvement with duration in the US. As a result, the difference virtually disappears for those in the United States from 16 to 20 years. The 1980 and 1990 Census data analyses show a similar pattern for recent immigrants. This appears to be a longitudinal phenomenon reflecting their refugee experience, rather than inherently poorer English proficiency that will persist among the most recent cohorts.

Thus, the low level of English proficiency among Soviet immigrants is due to the low proficiency among recent arrivals and the large proportion that recently arrived. It is a temporary and not a permanent phenomenon.

The analysis of earnings, other measured variables the same, also shows much lower earnings among recent Soviet Jewish immigrants, but a steeper improvement with duration in the United States. The Soviet immigrants have a much larger positive effect of schooling on earnings compared to other immigrants. An extra year of schooling raises the earnings of Soviet Jewish immigrants by about 7.3 percent, compared to only 4.6 percent for other immigrants. As a result there is an earnings catch-up coming sooner the higher the level of schooling. Similar patterns were found in the analyses for the 1980 and 1990 Censuses.

Again, this suggests that the earnings disadvantage of Soviet Jewish immigrants as a group is short-lived and is due to the low earnings of recent arrivals and the disproportionate number of recent arrivals in the 2000 Census.

Analyses of English language proficiency and earnings were also performed for those classified here as Soviet Jewish immigrants by the ancestry they reported in the 2000 Census: Russian, Ukrainian, an ancestry response that reveals one's religion, and all other ancestry responses. In the language analysis, there was essentially no difference in English proficiency, other variables the same, between those of Russian and "other ancestries", although those who indicated Ukrainian had slightly lower proficiency while those who indicated a religion were marginally more proficient. In the earnings analysis, other variables the same, there were no significant differences among these three groups, although the heterogeneous group of other ancestries showed a marginally significant 5 percent earnings advantage. The addition of ancestry variables to the language and earnings equations does not alter the effect of schooling.

Overall, it appears that Soviet Jewish immigrants adjust very well in the United States compared to other European immigrants. Their initial disadvantages in English language skills and earnings may be due to the refugee motivations for migration.[14] With the passage of time this disadvantage disappears. For earnings it disappears most rapidly for those with higher levels of schooling. This very high level of schooling and the greater effect of schooling on earnings among Soviet Jewish immigrants compared to other immigrants parallels patterns found among Jews and non-Jews born in the US (Chiswick, 1999). Thus, the Soviet Jews appear to be reflecting patterns that are specific to both refugees and Jews in the United States.

Appendix: Data regarding the 2000 Census analysis on immigrants to the US

Table 13.A1 Ancestry or ethnic origin of adult male Soviet immigrants who immigrated since 1965, 2000 (percent).

Ethnic ancestry	*Period of immigration*							
	1965–2000		*1965–1979*		*1980–1989*		*1990–2000*	
	All	*Excl. Armenian/ Ukrainian*[a]	*All*	*Excl. Armenian/ Ukrainian*[a]	*All*	*Excl. Armenian/ Ukrainian*[a]	*All*	*Excl. Armenian/ Ukrainian*[a]
Russian	41.1	52.6	36.7	49.9	39.1	51.9	41.4	53.2
Religion[b]	9.6	11.7	10.9	10.3	9.6	13.7	9.7	11.5
Armenian	10.8	–	12.7	–	17.1	–	9.8	–
Ukrainian	19.9	18.1	18.9	19.8	18.3	14.9	20.9	18.5
Not reported	5.8	5.6	6.7	8.4	7.3	6.9	5.4	4.8
Soviet Union, n.e.c.[c]	4.7	2.6	1.3	0.4	1.1	2.3	5.4	3.1
Lithuanian	1.5	1.5	0.8	1.1	1.3	1.8	1.5	1.5
Latvian	0.8	0.8	1.6	1.1	1.5	1.0	0.5	0.7
Polish	0.3	0.5	1.3	1.3	0.0	0.7	0.1	0.2
All other	5.5	6.7	9.1	7.5	4.7	6.7	5.3	6.5
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: Detail may not add to total due to rounding.
[a] Excludes persons of Armenian ancestry or who speak Armenian or Ukrainian at home.
[b] Response to ancestry question indicating the person's religion or religious origin, ancestry code 998.
[c] Includes Azerbaijani, Belorussian, Estonian, Ossetian, Moldavian, Tatar, Turkestani, Uzbek, Georgian, Tajik, and those who reported Soviet Union.

Source: 2000 Census of Population, Public Use Microdata Sample, 5 percent sample.

Table 13.A2 Language spoken in the home by adult males who immigrated from the FSU since 1965, 2000 (percent).[a]

	Period of Immigration			
Language	*1965–2000*	*1965–1979*	*1980–1989*	*1990–2000*
English only	4.0	9.2	5.0	3.1
Russian	71.6	63.3	69.4	73.4
Armenian	9.4	12.7	17.3	8.0
Ukrainian	7.2	2.5	2.5	8.4
Yiddish	0.2	0.4	0.0	0.1
Other	7.6	11.9	5.8	7.0
Total	100.0	100.0	100.0	100.0

Notes: Detail may not add to total due to rounding.
[a] Language currently spoken in the home other than or in addition to English.

Source: 2000 Census of Population, Public Use Microdata Sample, 5 percent sample.

Table 13.A3 Republic of birth of immigrants from the FSU, adult males, by period of immigration who immigrated since 1965, 2000.

Republic of Birth	*1965–2000*		*1965–1979*		*1980–1989*		*1990–2000*	
	Non-Armenian/ Ukrainian[a]	*All*	*Non-Armenian/ Ukrainian*[a]	*All*	*Non-Armenian/ Ukrainian*[a]	*All*	*Non-Armenian/ Ukrainian*[a]	*All*
Estonia	4	4	2	2	1	1	1	1
Latvia	120	120	24	24	40	40	56	56
Lithuania	155	155	18	18	35	35	102	102
Armenia	17	919	3	133	8	298	6	488
Azerbaijan	114	168	7	7	15	16	92	145
Belarus	417	421	15	15	87	87	315	319
Georgia	106	121	7	8	14	15	85	98
Moldova	228	236	25	25	24	24	179	187
Russia	3,540	3,610	428	434	636	647	2,476	2,529
Ukraine	2,180	2,848	319	344	361	425	1,500	2,079
USSR[b]	454	488	100	102	115	121	239	265
Kazakhstan	22	23	0	0	2	3	20	20
Kyrgyzstan	5	5	0	0	0	0	5	5
Tajikstan	8	8	0	0	3	3	5	5
Turkmenistan	1	2	0	0	0	0	1	2
Uzbekistan	250	256	15	15	26	26	209	215
Total	7,621	9384	963	1127	1,367	1741	5,291	6,516

Notes:
[a] Excludes persons reporting Armenian ancestry, or who speak Armenian or Ukrainian at home.
[b] Persons reporting USSR rather than a specific republic.

Source: 2000 US Census of Population, Public Use Microdata Sample, 5 percent sample.

Table 13.A4 Period of immigration for all adult male immigrants born in the FSU, including Armenians, 2000 (percent).

Period of Immigration	*All Years*	*Since 1965*
1995–2000	37.4	38.7
1990–1994	26.3	27.4
1985–1989	12.7	13.2
1980–1984	8.2	8.6
1975–1979	6.1	6.4
1970–1974	2.7	2.8
1965–1969	2.1	2.2
1960–1964	1.9	—
1950–1959	1.5	—
Before 1950	0.9	—
Total	100.0	100.0

Note: Detail may not add to total due to rounding.

Source: 2000 Census of Population, Public Use Microdata Sample, 5 percent sample.

Table 13.A5 Regression analysis of fluency in English among adult Soviet Jewish males who immigrated since 1965.

Dependent variable = ENGSPK	*2000 Census*							
Immigration Period	*1965–2000*		*1965–1979*		*1980–1989*		*1990–2000*	
Variable	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
CONSTANT	0.7773 (22.19)	0.7793 (21.83)	0.7989 (13.55)	0.7897 (13.12)	0.8356 (14.06)	0.8364 (13.76)	0.6814 (17.25)	0.6921 (17.07)
EDUCYRS	0.0391 (26.09)	0.0388 (25.57)	0.0172 (6.05)	0.0173 (6.06)	(0.0285 (9.50)	0.0282 9.25	0.0452 (23.39)	0.0446 (22.81)
AGE	–0.0104 (–22.69)	–0.0104 (–22.74)	–0.0027 (–3.52)	–0.0027 (–3.57)	–0.0081 (–9.27)	–0.0081 (–9.32)	–0.0129 (–21.03)	–0.0129 (–21.08)
IM95_00	–0.3547 (–18.16)	–0.3515 (–17.97)					–0.2552 (–21.93)	–0.2544 (–21.83)
IM90_94	–0.1057 (–5.62)	–0.1031 (–5.47)						
IM85_89	–0.0408 (–1.93)	–0.0388 (–1.83)			–0.0378 (–2.15)	–0.0373 (–2.11)		
IM75_79	0.0386 (1.77)	0.0428 (1.96)	0.0051 (0.22)	–0.0081 (0.35)	[a]	[a]		
IM70_74	0.0662 (1.81)	0.0678 (1.85)						
IM65_69	0.1089 (2.01)	0.1063 (1.95)	0.0033 (0.08)	–0.0021 (–0.05)				
MARRSP	0.0123 (1.08)	0.0128 (1.12)	–0.0078 (–0.39)	–0.0063 (–0.31)	0.0141 (0.65)	0.0142 (0.66)	0.0161 (1.06)	0.0160 (1.06)

RURAL	0.0247	0.0255	0.1399	0.1328	–0.0421	–0.0395	0.0126	0.0136
	(0.38)	(0.39)	(0.87)	(0.83)	(–0.33)	(–0.31)	(0.15)	(0.17)
SOUTH	0.0186	0.0187	–0.0021	–0.0025	0.0210	0.0218	0.0198	0.0197
	(1.40)	(1.40)	(–0.08)	(–0.10)	(0.73)	(0.75)	(1.17)	(1.16)
CHILD	–0.0074	–0.0070	0.0252	0.0252	–0.0108	–0.0110	–0.0085	–0.0075
	(–0.77)	(–0.72)	(1.44)	(1.44)	(–0.59)	(–0.60)	(–0.66)	(–0.58)
UKRAINE		–0.0198		0.0047		0.0053		–0.0315
		(–1.68)		(0.23)		(0.22)		(–2.04)
RELIG		0.0245		0.0121		0.0312		0.0258
		(1.75)		(0.45)		(1.29)		(1.37)
OTHANCS		0.0081		0.0197		0.0061		0.0041
		(0.67)		(0.90)		(0.28)		(0.25)
Sample size	6,492	6,492	856	856	1,240	1,240	4,394	4,394
Standard error	0.3440	0.3438	0.2235	0.2238	0.2833	0.2835	0.3733	0.3731
R^2	0.2347	0.2356	0.0651	0.0661	0.1196	0.1208	0.2422	0.2436
Adjusted R^2	0.2333	0.2338	0.0563	0.0539	0.1146	0.1136	0.2410	0.2419

Notes: t-ratios in parentheses.

[a] Omitted as benchmark; 1980–1984 is benchmark unless otherwise noted.

Source: 2000 Census of Population, Public Use Microdata Sample, 5 percent Sample.

Table 13.A6 Regression analysis of earnings among adult Soviet Jewish males who immigrated since 1965.

Dependent variable = LNEARN	*2000 Census*							
Immigration Period	*1965–2000*		*1965–1979*		*1980–1989*		*1990–2000*	
Variable	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
CONSTANT	5.024 (40.48)	5.010 (39.92)	4.229 (10.11)	4.176 (9.91)	4.798 (14.20)	4.833 (14.10)	5.029 (38.91)	5.016 (38.32)
EDUCYRS	0.0732 (19.17)	0.0736 (19.00)	0.0791 (6.36)	0.0789 (6.34)	0.0885 (9.74)	0.0880 (9.52)	0.0683 (15.32)	0.0689 (15.26)
EXP	0.0082 (2.02)	0.0081 (2.00)	0.0246 (1.85)	0.0236 (1.78)	0.0107 (1.08)	0.0111 (1.11)	0.0042 (0.89)	0.0043 (0.90)
EXPSQ	–0.00022 (–2.74)	–0.00023 (–2.74)	–0.00062 (–2.37)	–0.00061 (–2.33)	–0.00024 (–1.17)	–0.00024 (–1.19)	–0.00014 (–1.44)	–0.00014 (–1.45)
LNWW	1.045 (42.21)	1.044 (42.17)	1.175 (13.51)	1.179 (13.55)	1.021 (12.96)	1.017 (12.85)	1.028 (38.19)	1.027 (38.14)
IM95_00	–0.3272 (–6.76)	–0.3215 (–6.64)					–0.1233 (–4.63)	–0.1257 (–4.71)
IM90_94	–0.2093 (–4.60)	–0.2024 (–4.44)						
IM85_89	–0.0228 (–0.45)	–0.0175 (–0.34)			–0.0147 (–0.28)	–0.0102 (–0.19)		
IM75_79	0.0736 (1.40)	0.0836 (1.58)	0.0549 (0.58)	0.0650 (0.67)				
IM70_74	0.0270 (0.31)	0.0259 (0.29)						
IM65_69	0.0592 (0.45)	0.0399 (0.30)	0.0711 (0.43)	0.0430 (0.26)				

ENGSPK	0.3133	0.3112	0.3109	0.3028	0.2654	0.2660	0.3258	0.3241
	(10.43)	(10.35)	(2.16)	(2.11)	(3.11)	(3.12)	(9.98)	(9.93)
MARRSP	0.1413	0.1441	0.2947	0.2962	0.2275	0.2304	0.0879	0.0903
	(5.68)	(5.78)	(3.69)	(3.69)	(3.86)	(3.89)	(3.03)	(3.11)
RURAL	−0.1490	−0.1531	0.3456	0.3252	0.1459	0.1520	0.2357	−0.2430
	(−0.95)	(−0.97)	(0.52)	(0.49)	(0.39)	(0.40)	(−1.34)	(−1.38)
SOUTH	−0.0358	−0.0364	−0.0263	−0.0173	−0.0229	−0.0212	0.0403	−0.0422
	(−1.11)	(−1.13)	(−0.26)	(−0.17)	(−0.27)	(−0.25)	(−1.11)	(−1.16)
UKRAINE		−0.0206		0.0795		−0.0933		−0.0224
		(−0.73)		(0.94		(−1.33)		(−0.67)
RELIG		0.0451		0.2817		−0.0199		−0.0196
		(1.33)		(2.53)		(−0.28)		(0.48)
OTHANCS		0.0547		0.1122		−0.0174		0.0677
		(1.88)		(1.23)		(−0.26)		(1.96)
Sample size	6492	6492	856	856	1240	1240	4394	4394
Standard error	0.8294	0.8292	0.9326	0.9305	0.8392	0.8396	0.8038	0.8036
R^2	0.3436	0.3442	0.2736	0.2796	0.2319	0.2330	0.3622	0.3630
Adjusted R^2	0.3422	0.3425	0.2650	0.2685	0.2262	0.2255	0.3609	0.3612

Note: t-ratios in parentheses. Includes only immigrants who worked and had non-zero earnings in 1999.

Source: 2000 Census of Population, Public Use Microdata Sample, 5 percent sample.

Acknowledgements

Chiswick acknowledges the research support of the Institute of Government and Public Affairs, University of Illinois. Comments on earlier version from Carmel U. Chiswick, Allen Glicksman, and Mark Tolts are appreciated.

Notes

1 Analyses using a similar methodology have been conducted for the Hebrew language proficiency and labor market earnings of Jewish immigrants in Israel. See Chiswick (1988) and Chiswick and Repetto (2001) for analyses of the 1972 and 1983 Censuses of Israel. Unfortunately, the 1995 Census did not include any questions on language usage or language proficiency. The US and Israel studies are not strictly comparable because of differences in the Census questionnaires, the nature of immigration into these two countries, the relative magnitudes of the immigration flows after the collapse of the Soviet Union (small for the US, large for Israel), and the differences in the local (native) populations. Israel policy regarding intensive efforts to promote Hebrew language usage among immigrants was relaxed with regards to the Russian-speaking immigrants who arrived following the collapse of the FSU. For a discussion of this implicit change in policy see Glinert (1995).

2 In principle, data from the recently released National Jewish Population Survey (NJPS) 2000/2001 can be used to study the economic status of Soviet Jewish immigrants. The NJPS 2000/2001, however, provides a relatively smaller sample of Soviet Jews. Of the 5,148 respondents, both male and female aged 18 and over, only 281 were born in the FSU.

3 With the demise of the Soviet Union and the reunification with East Germany, Germany instituted a special immigration program to attract Soviet Jews to rebuild the German Jewish community (see Tress, 1995). In 2005, the German government was taking steps to effectively close this program (Bernstein, 2005).

4 According to the 2000 Census, the ethnic origins (ancestry) of the adult (aged 25–64) males born in the Soviet Union who immigrated in 1965 or later were 41 percent Russian, 20 percent Ukrainian, 11 percent Armenian, 10 percent response indicating a religion, 6 percent no ancestry reported, and 13 percent other responses. By languages spoken in the home, “only English” was reported by 4 percent, Russian 72 percent, Armenian 9 percent, Ukrainian 7 percent, Yiddish 0.2 percent, and all other languages 8 percent. There was little variation in the reported ancestry or language by sub-period of immigration. See Appendix Tables 13.A1 and 13.A2.

5 The very low proportion reporting Yiddish reflects the very rapid decline in the use of Yiddish by Russian/Soviet Jews during the 20th century. By the 1970s, “for the great majority of contemporary Soviet Jews (80 percent of our respondents), Russian is the native language”, with the proportion being greater for younger Jews. Yiddish was spoken primarily by older Jews or when younger Jews were speaking with their parents (Karklins, 1987, p. 29).

6 The schooling data cannot be decomposed into pre- and post-migration schooling, although given the age at migration there is likely to be little post-migration schooling among Soviet Jews.

7 The treatment of those with zero earnings differs across studies. As this appears to impact on the results, a careful examination of studies in this regard is recommended.

8 “Rural residence” is defined as living on a farm in the 2000 Census analysis and living in a rural area (farm or non-farm) in the 1980 and 1990 Census analyses.

9 Tolts (2004a) also finds a very low re-migration rate of Soviet Jewish immigrants who arrive in Israel.
10 The period of arrival categories used here are: 1996–2000, 1991–1995, 1987–1990, 1985–1986, 1980–1984, 1975–1979, 1970–1974, and 1965–1969. For the proportion of the sample who arrived in each interval, see Appendix Table 13.A4.
11 In the 2000 Census, unlike previous censuses, there is republic of birth codes for each of the 15 republics in the FSU, as well as a generic "USSR" code. Excluding those reporting Armenian by ancestry or language or that they speak Ukrainian at home, 46 percent reported the Russian Republic, 29 percent the Ukraine, 6 percent the USSR, 5 percent Belarus, and 14 percent reported having been born in the other 12 republics (Appendix Table 13.A3). In the post-World War II censuses until 2000 only the three Baltic Republics (Estonia, Latvia, and Lithuania) were separately identified from the rest of the Soviet Union because the US State Department did not recognize their incorporation into the Soviet Union.
12 A notable exception is the much larger positive effect of being married in the most recent cohort, 1990–2000.
13 For a discussion of the regional distribution of immigrants and their language skills, see Chiswick and Miller (2005).
14 Lower initial English proficiency and earnings and a speedier improvement appear to be a general refugee phenomenon, although not the larger payoff from schooling (see Chiswick, 1978, 1979;Chiswick & Miller, 1998).

References

Ahmed, B., & Robinson, J.G. (1994). Estimates of emigration of the foreign-born population: 1980–1990. US Bureau of the Census, Population Division Technical Working Paper no. 9.

Bernstein, R. (2005). Policy shifts in Germany trims Jewish migration. New York Times, February 20.

Chiswick, B. R. (1978). The effect of Americanization on the earnings of foreign-born men. *Journal of Political Economy, 86*(5), 897–922.

Chiswick, B. R. (1979). The economic progress of immigrants: Some apparently universal patterns. In: W. Fellner (Ed.), *Contemporary economic problems* (pp. 357–399). Washington: American Enterprise Institute.

Chiswick, B. R. (1988). Hebrew language usage: Determinants and effects on earnings among immigrants in Israel. *Journal of Population Economics, 11*(2), 253–271.

Chiswick, B. R. (1991). Jewish immigrant skill and occupational status at the turn of the century. *Explorations in Economic History, 28*(1), 64–86.

Chiswick, B. R. (1992). Jewish immigrant wages in America in 1909: An analysis of the Dillingham commission data. *Exploration in Economic History, 29*(3), 274–289.

Chiswick, B. R. (1993). Soviet Jews in the United States: An analysis of their linguistic and economic adjustment. *International Migration Review, 27*(2), 260–286.

Chiswick, B. R. (1997). Soviet Jews in the United States: Language and labor market adjustments revisited. In: N. L. Epstein, Y. Ro'i & P. Ritterband (Eds), *Russian Jews on three continents: Migration and resettlement* (pp. 233–260). London: Frank Cass Publishers.

Chiswick, B. R. (1999). The occupational attainment and earnings of American Jewry, 1890–1990. *Contemporary Jewry, 20*, 68–98.

Chiswick, B. R. (2000). Are immigrants favorably self-selected: An economic analysis. In: C. D. Brettell & J. F. Hollifield (Eds), *Migration theory: Talking across disciplines* (pp. 61–76). New York: Routledge.

Chiswick, B. R., & Miller, P. W. (1998). English language fluency among immigrants in the United States. *Research in Labor Economics, 17*, 151–200.

Chiswick, B. R., & Miller, P. W. (2005). Do enclaves matter in immigrant adjustment. *City and Community, 4*(1), 5–35.

Chiswick, B. R., & Repetto, G. (2001). Immigrant adjustment in Israel: Literacy and fluency in Hebrew and earnings. In: S. Djajic (Ed.), *International migration: Trends, policy and economic impact* (pp. 204–228). New York: Routledge.

Glinert, L. H. (1995). Inside the language planner's head: Tactical responses to a mass immigration. *Journal of Multilingual and Multicultural Development, 16*(5), 351–372.

Karklins, R. (1987). Determinants of ethnic identification in the USSR: The Soviet Jewish case. *Ethnic and Racial Studies, 10*(1), 27–47.

Mulder, T. J. (2003). Foreign-born emigration from the United States: 1990 to 2000. Paper presented at the Population Association of America, Annual Meeting, Minneapolis, May.

Tolts, M. (2004a). Demographic trends among the Jews of the FSU. Paper presented at the International Conference on Soviet and Post-Soviet Jewry in Honor of Professor Mordechai Altshuler, Hebrew University, Jerusalem, December 28–30, 2003. (Revised January 12).

Tolts, M. (2004b). The post-Soviet Jewish population in Russia and the world. *Jews in Russia and Eastern Europe, 1*(52), 37–63.

Tress, M. (1995). Soviet Jews in the Federal Republic of Germany: The rebuilding of a community. *The Jewish Journal of Sociology, 37*(1), 39–54.

U.S. Bureau of the Census. (2003). 2000 Census of population and housing, public use microdata sample, United States, Technical Documentation, Washington, DC.

Part IV

Language and earnings among the native born

14 The economic cost to native-born Americans of limited English language proficiency

I. Introduction

In recent years there has been much concern over the consequences of a variety of educational programs related to language issues, including bilingual education and English-as-a-second-language programs. These concerns have been legal and educational. The legal concerns are over the access students have to various types of schooling, and have been played out in the courts, legislative bodies and in referendums. The educational concerns have frequently focused on the extent to which various types of schooling either enhance or retard English language proficiency and proficiency in the students' origin or ancestral language. This chapter focuses on another dimension of the public policy issue that, ultimately, may prove to be a critical, but hitherto largely unexplored, aspect of the debate. Namely, to what extent do native-born Americans with limited English language proficiency, but with some degree of proficiency in another language, incur economic costs.

Section II develops a methodology to explore these issues and discusses the data that will provide the basis for the empirical estimates. Section III presents the results of the empirical analysis of the effects on earnings of limited English language proficiency among those born in the United States. Section IV is a summary and conclusion.

II. The model and the data

The model

Language skills are a form of human capital. They satisfy the three criteria that define human capital (Becker, 1964). First, language skills are created at a cost, where this cost includes not only the time of the person, but also the time of parents, teachers and others that is devoted to enhancing language skills, as well as purchased inputs, including books (for home use and school use), school rooms and other school supplies. Even early language development among young children is not without cost, as parental and other adult time in talking with and reading to young children is essential to their language development.

Second, there is a presumption, to be developed further below, that language skills are productive. Language skills are presumably productive in the individual's role as a consumer and in the role as a producer (supplier of labor services). Those deficient in language skills will find it more difficult (costly) to search for the lowest price, or the highest quality for a given price, of the goods and services they may wish to purchase. Searching for information regarding the price, quality and efficacy of a good or service requires a degree of speaking proficiency and/or literacy.

Language skills are presumably important in the labor market.[1] The language skills may be required for the job because of oral communication with others (*e.g.*, consumers, suppliers, supervisors, or co-workers) or because literacy (reading, writing or both) may be needed to do the job efficiently or to do it safely.

Finally, language skills are embodied in the person. Unlike physical capital (*e.g.*, a machine) which may be separated from its owner, a person's language proficiency cannot be alienated from the individual.

If language skills are a form of human capital, as are schooling, on-the-job training, and health status, individuals will invest in these skills and will use them in consumption and production activities so as to maximise the person's own economic well-being, subject to various constraints. These constraints include one's own time, resources to purchase inputs into the production of human capital, ability, home environment, school quality, etc.

The primary language in the U.S. labor market is English and it would be reasonable to hypothesize that, at least up to some point, greater proficiency in English is associated with greater productivity in the labor market, and hence will result in higher earnings. Someone with greater proficiency could, presumably, do a job that could be done by someone with lesser proficiency, but in addition would have job opportunities that would not be available to someone with a lower degree of proficiency.

A person who is fully proficient in English, but who also has some proficiency in another language may have an additional economic advantage. Knowledge of a second language may expand job opportunities. This could arise if that language was valuable in international trade or finance, or in economic interactions with suppliers, consumers, or others who use that language in the United States.

There are, however, circumstances in which a person who speaks a language other than English, as well as being able to speak English, incurs an economic disadvantage. This could arise, for example, if the speaking or studying of the other language detracts from the individual's full proficiency in English. That is, the person may be bilingual, but not fully proficient in English, and the returns, if any, from knowing the other language are more than offset by deficiencies in English. It may also arise if the bilingualism is associated with an accent or intonation, and if this serves as a basis of discrimination in the labor market. The discrimination may be pure prejudice against those with a particular accent or speech pattern, or it may be based on the false notion that this implies less education or less ability.

In addition, speaking a language other than English may, for some, reflect isolation from the mainstream American economy. This may arise from either growing up in, or currently living and working in, an immigrant/ethnic/indigenous group enclave. Having been raised in such an environment may result in less human capital acquired in the home as a child or in school relevant for the mainstream American economy. Those with limited English language proficiency but with knowledge of another language may optimize their labor market success by working in a partial or full language enclave in which the non-English language serves as a partial or primary medium of communication.

It is, therefore, an empirical question as to the extent to which limited English language proficiency penalizes individuals in the labor market and the extent to which full bilingualism, in the sense of being fully proficient in English but with a knowledge of another language, enhances labor market opportunities.

The data

The purpose of this chapter is to investigate the relation between language skills and labor market outcomes among those born in the United States.[2] The simplest and most direct measure of labor market outcomes is earnings, that is, income from wages, salaries and self-employment. The analysis of earnings will be limited to males age 25 to 64 years. The age limitation is imposed to develop more precise estimates by excluding the aged, many of whom are partly or fully retired, and youths, many of whom are in school or investing heavily in on-the-job training. Analyses of the earnings of women are more complex than that of men because of a far greater propensity on the part of women to move in and out of the labor market. At this stage, to keep the analyses straightforward, the study is limited to adult men.

The 1990 Census of Population and Housing of the United States is the most appropriate data set for the purpose of this analysis. It is the most recent large data set that provides information on a variety of characteristics of each respondent, including their earnings and information on language skills. In particular, the Census in April 1990 asked the respondents to report their wage, salary and self-employment income in 1989. The Census also asked respondents to identify the language, if any, other than English currently spoken in the home (other than just a few words). Those who reported a language other than English were also asked to identify this language and whether their speaking ability in English was: "Very Well", "Well", "Not Well" or "Not at All" (none). These latter four categories plus the "Only English" group constitute the five language categories.

These data are, of course, not ideal. There are no data on proficiency in

English among those who report English is the only language spoken at home. There are no data on languages that are known but are not spoken at home. The data on English language proficiency for those who speak another language are self-reported. Finally, the Census does not ask questions on literacy. Analyses show, however, that the Census questions are very productive for research purposes and that there is a strong relation between literacy and speaking proficiency among limited English proficient individuals (Chiswick and Miller, 1998). In spite of its limitations, however, there is no data set that is superior to the 1990 Census of Population for an analysis of earnings and language proficiencies for a large random sample of the entire native-born population. Since relatively few in the adult native-born population speak a language other than or in addition to English at home, the very large samples for analysis in the microdata files created from the decennial census are essential for generating sufficiently large samples for those who speak a language other than English.

III. Analysis

Speaking a language other than English at home is not common among adult native-born men in the United States.[3] According to the 1990 Census of Population, 5.5 percent of men reported speaking a non-English language at home, and were to some degree bilingual as nearly all reported speaking at least some English. Of those reporting a non-English language spoken, 60 percent (primarily Hispanics) reported Spanish. Another 9 percent each reported German and French. This was followed by other European, Asian and American Indian languages, representing the languages of old immigrants, new (post-1965) immigrants and indigenous peoples.[4]

Table 14.1 indicates the proportion of the racial/ethnic groups that report speaking a language other than English. While overall this is done by only 5.5 percent of the men, the range is from 3.3 percent for white non-Hispanic men to about two-thirds of all Hispanic and Mexican-origin men. About one-in-five American Indians and native-born men of Asian and Pacific Island origin speak a language other than English. While nearly all the Hispanic and Mexican-origin men who speak another language report Spanish, the other groups report much greater linguistic diversity.

Table 14.2 reports the mean earnings in 1989 of adult (age 25 to 64) men by language skills. Those who speak only English at home have the highest earnings, over $32,300, while those who speak Spanish or other languages have lower earnings. Even those who report they speak English "Very Well" have lower earnings than English-only speakers, just over $31,800 for languages other than Spanish and just under $26,000 for Spanish speakers. Earnings generally decline in Table 14.2 with a decline in self-reported proficiency in English.

The data in Table 14.2 are suggestive but cannot be taken as definitive because they represent simple relationships. Other factors associated with

Table 14.1 Non-English languages spoken by adult native-born men by race and Hispanic origin, 1990 Census

Racial/Hispanic Group (percent)	*Percent Speaking a Language Other than English*	*Major non-English Languages (as % of those speaking another language)*
Total (100.0)	5.5	Spanish (60.0), German (9.4), French (9.2), Italian (4.6), Polish (1.8), Japanese (1.5), Greek (1.4), Chinese (1.1), Navaho (0.8), Portuguese (0.7), Other (9.5)
White, Non-Hispanics (88.6)	3.3	Spanish (27.1), German (16.5), French (15.4), Italian (8.7), Polish (3.5), Greek (2.7), Japanese (2.6), Chinese (2.1), Navaho (1.6), Other (19.8)
Hispanics[(a)] (3.5)	63.2	Spanish (98.9), Other (1.1)
Mexican[(a)] (2.6)	67.7	Spanish (99.6), Other (0.4)
Blacks (7.9)	3.7	Spanish (54.3), French (17.6), German (11.6), Japanese (2.7), Arabic (2.1), Korean (2.0), Other (9.7)
American Indians (0.8)	21.8	Navaho (28.6), Spanish (16.3), Pima (4.2), Yupik (4.0), Cherokee (3.5), Apache (3.4), Dakota (3.3), "American Indian" (3.3), Choctaw (2.7), French (2.7), Other (28.0)
Asian & Pacific Islanders (0.7)	20.2	Chinese (34.8), Japanese (24.9), Tagalog (15.1), Spanish (9.7), Hindi (2.6), Hawaiian (2.2), Samoan (2.1), Other (8.6)

Note:
(a) = Persons of Hispanic or Mexican origins may be of any race.

Source: U.S. Bureau of the Census, *1990 Census of Population*, Public Use Microdata Sample, one percent sample.

speaking a language other than English at home may be influencing earnings. Those who speak a language other than English at home may have less schooling, may live in poorer regions of the country (rural areas and the South), may work fewer weeks in the year, may be more likely to be Hispanic or members of other disadvantaged racial and/or ethnic minorities, or may have had foreign-born parents. With the exception of parental nativity, it is possible to control statistically for each of these potentially confounding influences.[5]

The "human capital earnings function", which relates the natural logarithm of earnings to a set of explanatory variables, is estimated statistically using ordinary least squares regression techniques. The explanatory variables include years of schooling completed, potential labor market experience

Table 14.2 Mean earnings of adult native-born men who worked in 1989, by language skills, 1990 Census[(a)]

Speaks Only English at Home:	$32,315
Speaks Spanish and Speaks English:	
Very Well	25,987
Well	21,664
Not Well	23,258
Not at All	16,649
Speaks a Language Other Than Spanish and Speaks English:	
Very Well	31,833
Well	26,406
Not Well	27,675
Not at All	(b)

Notes:
(a) Earnings are wage, salary and self-employment income.
(b) Sample size is only 23 cases (unweighted).

Source: U.S. Bureau of the Census, *1990 Census of Population*, Public Use Microdata Sample, one percent sample.

(age-schooling-5) and its square, the natural logarithm of weeks worked, and dichotomous variables for whether the respondent is black by race, married and living with his spouse, living in a rural area, and living in a southern state. Two types of language variables are included separately in the analysis. One is whether the respondent speaks a language other than or in addition to English at home. The second set is the degree of English language fluency, that is, four dichotomous variables for speaking English "Very Well", "Well", "Not Well", or "Not at All", with speaking only English as the benchmark. The analysis is computed for all adult native-born men with earnings in 1989, and then separately for groups defined by race and Hispanic origin.

The results of the estimated effect of the language variables on earnings for the seven racial/ethnic groups under study are reported in Table 14.3.[6] The full set of regression results is presented in the Statistical Appendix. Among native-born men with earnings age 25 to 64 in 1990, only 5.5 percent reported speaking a language other than English. Of these 76.7 percent reported speaking English "Very Well", 16.3 percent reported "Well", 6.6 percent "Not Well", and a surprising 0.5 percent "Not at All". While it is difficult to understand the circumstances under which adult men born in the U.S. who have earnings would report speaking English "Not Well" or "Not at All", it is reassuring that these are very rare circumstances. In part because this group is very small, it may be dominated by measurement error.

The data in the first column of the top panel in Table 14.3 indicate that, other things being the same, adult men who speak a language other than English at home receive 8 percent lower earnings than those who speak only English at home.[7]

The dichotomous non-English language variable is replaced by the four

Table 14.3 Partial effects of language variables on the earnings of adult native-born men, 1990 Census[a]

Speaks	*All*	*White Non-Hispanic*	*Hispanic*	*Mexican Origin*	*Black*	*American Indian*	*Asian and Pacific Island Origin*
Foreign Language	–N0.081 (–15.6)	–0.041 (–5.5)	–0.107 (–8.8)	–0.093 (–6.4)	–0.012 (–0.5)	–0.190 (–5.2)	–0.059 (–1.7)
English Fluency							
Very Well	–0.075 (–12.9)	–0.031 (–3.7)	–0.094 (–7.6)	–0.074 (–5.0)	–0.025 (–1.0)	–0.178 (–4.2)	–0.022 (–0.6)
Well	–0.126 (–9.2)	–0.096 (–4.5)	–0.164 (–7.6)	–0.176 (–7.0)	+0.013 (+0.2)	–0.223 (–3.3)	–0.024 (–2.7)
Not Well	–0.056 (–2.7)	–0.046 (–1.6)	–0.153 (–3.9)	–0.152 (–3.3)	+0.032 (+0.5)	–0.226[b] (–1.8)	–0.018[c] (–0.2)
Not at All	–0.022 (–0.2)	–0.048[c] (–0.2)	–0.087[b] (–0.9)	–0.224[b] (–2.0)	–0.058[c] (–0.3)	+0.545[c] (–2.5)	–0.445[c] (–0.8)

Notes:
(a) Dependent variable is the natural logarithm of wage, salary and self-employment income. Other variables in the equation are: years of schooling, years of potential labor market experience and its square, the natural logarithm of weeks worked, and dichotomous variables for being married (spouse present) and living in a rural area or a southern state. A variable for Blacks is included in the overall, Hispanic, and Mexican-origin equations. The American Indian equation includes a dichotomous variable for Alaska, as does the Asian and Pacific Islander equation for California and for Hawaii.
(b) Sample size less than 100 and more than 50.
(c) Sample size 50 or fewer cases.
'*t*' statistics are in parentheses.

Source: U.S. Bureau of the Census, *1990 Census of Population*, Public Use Microdata Sample, one percent sample.

categories for English language proficiency in the second panel of Table 14.3. Compared to those who speak only English, those who speak English "Very Well" have about 7.5 percent lower earnings, those who speak it "Well" have 12.6 percent lower earnings, with the two very small groups with more limited English skills being associated with 5.6 percent lower earnings for the "Not Well" and no significant effect for the very few in the "Not at All" category.

The remaining columns in Table 14.3 look at the effect of language on earnings for adult native-born men, separately by racial/ethnic category. This is done to remove the possible confounding effects in the first column of interactions of language patterns with race/ethnicity. The race/ethnic categories are essentially mutually exclusive, except for those of Mexican origin who are a subset of Hispanics.

Among adult U.S.-born men, those who speak a language other than English at home earn less than those who do not by about 4 percent among white non-Hispanic men, 11 percent among Hispanic men (9.3 percent for those of Mexican origin), and about 19 percent among American Indians. Although the coefficients are negative, there is no significant earnings effect among the small samples of Black and Asian and Pacific Islander men who speak a non-English language.

Interesting results emerge when the English language fluency variables are entered into the analysis. Those who speak English "Very Well" have lower earnings than those who speak only English. The effect is largest for American Indian men (about 18 percent), followed by Hispanic (9 percent), Mexican origin (7 percent) and white non-Hispanic men (3 percent), but is not statistically significant for Black and Asian-origin men. Those whose self-reported English is only "Well" have a larger deficiency for each group. The earnings disadvantage associated with this level of English proficiency is now statistically significant for Asians, but is not significant among Blacks. The "Not Well" and "Not at All" categories become even smaller in size when divided into race/ethnic categories, and generally show negative patterns.

In summary, among adult native-born men with earnings, those who speak a language other than English at home receive lower earnings than those who speak only English at home. Even those who report that they speak English "Very Well" have lower earnings than those who speak only English at home.

Additional statistical tests were undertaken to determine why even those bilingual speakers who speak English "Very Well" earn less than monolingual English speakers. Among white *non-Hispanic* men, those who speak Spanish (27 percent of non-English language speakers) are at a smaller earnings disadvantage than those who speak other languages. The earnings of the Spanish speakers do not differ significantly from those of monolingual English speakers, although the earnings differential is significant for the other languages.[8]

This pattern did not change when the analyses for white non-Hispanic men were performed separately for the five states with a high concentration of Spanish speakers (Arizona, California, Florida, New Mexico and Texas) and all other states.

Among Hispanics, nearly all who speak a language other than English report it is Spanish. When the analysis is performed separately for the five high Spanish concentration states and all other states an important disparity arises. The labor market disadvantages from speaking Spanish at home are much larger in the high Spanish concentration states.[9] It is striking that among adult male native-born Hispanics, other things being the same, those who report they speak English "Very Well" receive about 12 percent lower earnings in the 5 high Spanish Concentration states and 4 percent lower earnings in other states, compared to those Hispanics who speak only English at home.

The results for white non-Hispanics and for Hispanics, the two numerically largest groups of bilingual speakers, suggest that ethnicity and ethnic concentrations matter. The earnings disadvantage appears to be larger among those who speak their ethnic-origin language, or do so in an area where many others speak the same language.

IV. Summary and conclusion

This study has been concerned with the effect on their earnings of bilingualism and English language ability among adult (age 25 to 64) native-born men. The data are from the very large microdata sample from the 1990 Census of Population. Among these men in the U.S., 5.5 percent report speaking a language other than English at home. Of those speaking another language, 60 percent report Spanish (primarily of Hispanic origin), followed by German and French (9 percent each) and other old immigrant, new immigrant and indigenous (American Indian) languages.

Overall and other variables the same, those who speak another language earn less than those who at home are monolingual English speakers. Earnings decline with a lower degree of proficiency in English, but even those who report they speak English "Very Well" receive lower earnings than those who speak only English. The earnings disadvantage among those who speak English "Very Well", compared to those who speak only English at home, ranges from about 3 percent for white non-Hispanic men, to about 9 percent for Hispanic men, to about 18 percent for American Indians.

There are several reasons why those who report they are bilingual and speak English "Very Well" might have lower earnings than those who speak only English. One is that their level of proficiency in English might be lower than the monolingual English speakers because speaking in childhood and/or as adults this other language competes with their obtaining full proficiency in English. That is, they may have less English language proficiency than monolingual English language speakers. Another is that they may experience discrimination because of an accent or speech pattern related to their other language. Living and working in an ethnic concentration area because of their language deficiencies may also be, in part, responsible for their lower earnings.

These hypotheses are supported by the finding that, among white non-Hispanic men, speaking Spanish has a smaller adverse effect than speaking other languages, and that among Hispanics living in a state with a high concentration of Spanish speakers has a larger negative effect on earnings. This suggests that it is not the Spanish language per se that is associated with lower earnings, but rather speaking the ancestral language and living among others who do so.

There appears to be no statistical support for the proposition that bilingualism, as measured in this study, enhances earnings in the U.S. It does provide support for the proposition that whatever detracts from full proficiency in English has an adverse effect on earnings.

Statistical Appendix

Table 14.A1 Regression estimates of earnings equations, total adult native-born men in paid employment, 1990 Census

Variable	*(i)*	*(ii)*	*(iii)*	*Mean/(SD)*
Constant	4.047 (199.09)	4.064 (199.48)	4.065 (199.50)	1.000 (0.00)
Education	0.093 (179.14)	0.092 (178.00)	0.092 (177.75)	13.468 (2.67)
Experience (EXP)	0.037 (77.73)	0.037 (77.54)	0.037 (77.55)	22.007 (11.13)
$EXP^2/100$	–0.055 (56.85)	–0.055 (56.68)	–0.055 (56.69)	6.082 (5.73)
Log Weeks Worked	1.092 (215.21)	1.091 (214.97)	1.091 (214.96)	3.799 (0.44)
Married	0.273 (103.11)	0.272 (102.66)	0.272 (102.66)	0.692 (0.46)
Race (Black)	–0.190 (48.85)	–0.193 (49.58)	–0.193 (49.60)	0.094 (0.29)
Rural Location	–0.129 (46.39)	–0.131 (47.21)	–0.131 (47.21)	0.272 (0.45)
South	–0.099 (40.79)	–0.099 (40.69)	–0.098 (40.65)	0.361 (0.48)
Speaks a Non-English Language	(a)	–0.081 (15.59)	(a)	0.055 (0.23)
Speaks English Very Well	(a)	(a)	–0.075 (12.95)	0.042 (0.20)
Speaks English Well	(a)	(a)	–0.126 (9.24)	0.009 (0.09)
Speaks English Not Well	(a)	(a)	–0.056 (2.66)	0.004 (0.06)
Speaks English Not at All	(a)	(a)	–0.022 (0.24)	0.0003 (0.02)
$\bar{R}^2$	0.3789	0.3792	0.3792	
Sample Size	501,021	501,021	501,021	501,021

Notes:
(a) = Variable not entered.
'*t*' statistics in parentheses in first four columns computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator, standard deviations in parentheses in final column.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 1 percent sample.

Table 14.A2 Regression estimates of earnings equations, adult white non-Hispanic native-born men in paid employment, 1990 Census

Variable	*(i)*	*(ii)*	*(iii)*	*Mean/(SD)*
Constant	4.054 (176.01)	4.057 (176.07)	4.058 (176.10)	1.000 (0.00)
Education	0.094 (169.63)	0.094 (169.64)	0.094 (169.57)	13.620 (2.61)
Experience (EXP)	0.039 (77.41)	0.039 (77.31)	0.039 (77.32)	22.087 (11.12)
$EXP^2/100$	–0.062 (58.66)	–0.062 (58.54)	–0.062 (58.54)	6.115 (5.70)
Log Weeks Worked	1.082 (187.09)	1.082 (187.03)	1.088 (187.02)	3.813 (0.42)
Married	0.270 (94.38)	0.270 (94.13)	0.270 (94.12)	0.716 (0.45)
Rural Location	–0.136 (46.90)	–0.136 (46.96)	–0.136 (46.95)	0.295 (0.46)
South	–0.088 (33.79)	–0.088 (33.89)	–0.088 (33.89)	0.341 (0.47)
Speaks a Non-English Language	(a)	–0.041 (5.49)	(a)	0.033 (0.18)
Speaks English Very Well	(a)	(a)	–0.031 (3.74)	0.026 (0.16)
Speaks English Well	(a)	(a)	–0.096 (4.47)	0.005 (0.07)
Speaks English Not Well	(a)	(a)	–0.046 (1.65)	0.003 (0.05)
Speaks English Not at All	(a)	(a)	–0.048 (0.21)	0.0001 (0.01)
$\bar{R}^2$	0.3462	0.3462	0.3462	
Sample Size	444,029	444,029	444,029	444,029

Notes:
(a) = Variable not entered.
'*t*' statistics in parentheses in first four columns computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator, standard deviations in parentheses in final column.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 1 percent sample.

Table 14.A3 Regression estimates of earnings equations, adult Hispanic native-born men in paid employment, 1990 Census

Variable	*(i)*	*(ii)*	*(iii)*	*Mean/(SD)*
Constant	4.184 (51.67)	4.276 (52.04)	4.299 (52.16)	1.000 (0.00)
Education	0.078 (33.44)	0.075 (32.08)	0.074 (30.79)	12.325 (3.12)
Experience (EXP)	0.025 (10.87)	0.026 (11.17)	0.026 (11.12)	20.423 (11.21)
$EXP^2/100$	–0.030 (6.52)	–0.031 (6.66)	–0.031 (6.57)	5.427 (5.89)
Log Weeks Worked	1.113 (54.61)	1.109 (54.37)	1.108 (54.31)	3.737 (0.51)
Married	0.257 (20.33)	0.258 (20.46)	0.258 (20.42)	0.613 (0.49)
Race (Black)	–0.066 (1.54)	–0.090 (2.11)	–0.089 (2.08)	0.017 (0.13)
Rural Location	–0.072 (3.94)	–0.073 (4.01)	–0.073 (4.00)	0.112 (0.32)
South	–0.187 (15.66)	–0.165 (13.52)	–0.162 (13.28)	0.349 (0.48)
Speaks a Non-English Language	(a)	–0.108 (8.92)	(a)	0.632 (0.48)
Speaks English Very Well	(a)	(a)	–0.096 (7.69)	0.482 (0.50)
Speaks English Well	(a)	(a)	–0.165 (7.66)	0.115 (0.32)
Speaks English Not Well	(a)	(a)	–0.154 (3.92)	0.030 (0.17)
Speaks English Not at All	(a)	(a)	–0.090 (0.88)	0.005 (0.07)
$\bar{R}^2$	0.4595	0.4619	0.4623	
Sample Size	17,630	17,630	17,630	17,630

Notes:
(a) = Variable not entered.
'*t*' statistics in parentheses in first four columns computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator, standard deviations in parentheses in final column.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 1 percent sample.

Table 14.A4 Regression estimates of earnings equations, adult Mexican origin native-born men in paid employment, 1990 Census

Variable	*(i)*	*(ii)*	*(iii)*	*Mean/(SD)*
Constant	4.309	4.376	4.420	1.000
	(46.05)	(46.26)	(46.63)	(0.00)
Education	0.075	0.073	0.071	12.063
	(28.78)	(28.09)	(26.27)	(3.20)
Experience (EXP)	0.026	0.027	0.027	21.083
	(9.70)	(10.08)	(10.04)	(11.35)
$EXP^2/100$	–0.032	–0.033	–0.032	5.733
	(5.90)	(6.09)	(6.00)	(6.07)
Log Weeks Worked	1.079	1.077	1.075	3.732
	(45.68)	(45.53)	(45.44)	(0.52)
Married	0.278	0.279	0.278	0.630
	(8.28)	(18.33)	(18.25)	(0.48)
Race (Black)	–0.137	–0.166	–0.164	0.007
	(1.89)	(2.26)	(2.23)	(0.08)
Rural Location	–0.054	–0.051	–0.051	0.111
	(2.61)	(2.48)	(2.51)	(0.31)
South	–0.207	–0.185	–0.180	0.392
	(15.23)	(13.26)	(12.89)	(0.49)
Speaks a Non-English Language	(a)	–0.095	(a)	0.677
		(6.48)		(0.47)
Speaks English Very Well	(a)	(a)	–0.076	0.510
			(5.08)	(0.50)
Speaks English Well	(a)	(a)	–0.178	0.128
			(7.02)	(0.33)
Speaks English Not Well	(a)	(a)	–0.154	0.033
			(3.30)	(0.18)
Speaks English Not at All	(a)	(a)	–0.227	0.005
			(2.05)	(0.07)
$\bar{R}^2$	0.4539	0.4555	0.4565	
Sample Size	12,907	12,907	12,907	12,907

Notes:
(a) = Variable not entered.
'*t*' statistics in parentheses in first four columns computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator, standard deviations in parentheses in final column.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 1 percent sample.

Table 14.A5 Regression estimates of earnings equations, adult black native-born men in paid employment, 1990 Census

Variable	*(i)*	*(ii)*	*(iii)*	*Mean/(SD)*
Constant	4.001	4.001	4.001	1.000
	(73.12)	(73.13)	(73.12)	(0.00)
Education	0.084	0.084	0.084	12.504
	(43.07)	(43.03)	(43.02)	(2.70)
Experience (EXP)	0.024	0.024	0.024	21.866
	(14.97)	(14.97)	(14.97)	(11.15)
$EXP^2/100$	–0.023	–0.023	–0.023	6.023
	(7.29)	(7.29)	(7.30)	(5.90)
Log Weeks Worked	1.114	1.114	1.114	3.694
	(87.41)	(87.41)	(87.44)	(0.58)
Married	0.289	0.289	0.289	0.505
	(35.08)	(35.05)	(35.05)	(0.50)
Rural Location	–0.072	–0.072	–0.072	0.128
	(6.18)	(6.19)	(6.19)	(0.33)
South	–0.174	–0.174	–0.174	0.554
	(20.69)	(20.69)	(20.69)	(0.50)
Speaks a Non-English Language	(a)	–0.012	(a)	0.037
		(0.55)		(0.19)
Speaks English Very Well	(a)	(a)	–0.025	0.026
			(0.95)	0.16
Speaks English Well	(a)	(a)	0.013	0.006
			(0.24)	(0.08)
Speaks English Not Well	(a)	(a)	0.032	0.004
			(0.51)	0.06
Speaks English Not at All	(a)	(a)	–0.058#	0.0002
			(0.33)	(0.01)
$\bar{R}^2$	0.4867	0.4867	0.4867	
Sample Size	39,598	39,598	39,598	39,598

Notes:
(a) = Variable not entered.
'*t*' statistics in parentheses in first four columns computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator, standard deviations in parentheses in final column.
= Based on fewer than 30 observations.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 1 percent sample.

Table 14.A6 Regression estimates of earnings equations, adult male American Indians in paid employment, 1990 Census

Variable	*(i)*	*(ii)*	*(iii)*	*Mean/(SD)*
Constant	4.308 (31.93)	4.391 (32.06)	4.387 (31.82)	1.000 (0.00)
Education	0.0691 (11.10)	0.067 (10.84)	0.067 (10.80)	12.467 (2.80)
Experience (EXP)	0.024 (4.02)	0.024 (4.03)	0.024 (4.14)	21.296 (10.54)
$EXP^2/100$	–0.028 (2.38)	–0.028 (2.35)	–0.030 (2.47)	5.645 (5.46)
Log Weeks Worked	1.066 (33.33)	1.060 (32.97)	1.060 (32.94)	3.595 (0.70)
Married	0.293 (9.91)	0.289 (9.87)	0.288 (9.80)	0.584 (0.49)
Rural Location	–0.132 (4.60)	–0.109 (3.77)	–0.110 (3.77)	0.408 (0.49)
South	–0.036 (1.26)	–0.063 (2.19)	–0.062 (2.16)	0.037 (0.47)
Alaska	0.691 (8.88)	0.708 (9.16)	0.707 (9.15)	0.046 (0.21)
Speaks a Non-English Language	(a)	–0.190 (5.21)	(a)	0.218 0.41
Speaks English Very Well	(a)	(a)	–0.178 (4.22)	0.153 (0.36)
Speaks English Well	(a)	(a)	–0.223 (3.25)	0.050 (0.22)
Speaks English Not Well	(a)	(a)	–0.226 (1.81)	0.014 (0.12)
Speaks English Not at All	(a)	(a)	0.545# (2.45)	0.001 (0.03)
$\bar{R}^2$	0.4797	0.4836	0.4835	
Sample Size	3,970	3,970	3,970	3,970

Notes:
(a) = Variable not entered.
'*t*' statistics in parentheses in first four columns computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator, standard deviations in parentheses in final column.
= Based on fewer than 30 observations.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 1 percent sample.

Table 14.A7 Regression estimates of earnings equations, adult Asian and Pacific Islander native-born American men in paid employment, 1990 Census

Variable	*(i)*	*(ii)*	*(iii)*	*Mean/(SD)*
Constant	3.995 (5.28)	4.028 (15.34)	4.031 (15.36)	1.000 (0.00)
Education	0.107 (16.17)	0.107 (16.10)	0.107 (16.34)	14.304 (2.43)
Experience (EXP)	0.045 (8.62)	0.044 (8.52)	0.044 (8.56)	19.772 (11.07)
$EXP^2/100$	–0.073 (6.61)	–0.072 (6.45)	–0.072 (6.45)	5.135 (5.46)
Log Weeks Worked	1.031 (15.64)	1.028 (15.59)	1.026 (15.50)	3.818 (0.39)
Married	0.234 (9.10)	0.233 (9.06)	0.233 (9.09)	0.564 (0.50)
Rural Location	–0.034 (0.71)	–0.036 (0.76)	–0.040 (0.84)	–0.063 (0.24)
South	–0.090 (1.78)	–0.088 (1.74)	–0.085 (1.70)	–0.080 (0.27)
California	0.094 (2.70)	–0.091 (2.60)	–0.092 (2.66)	0.381 ((0.49)
Hawaii	–0.010 (0.26)	–0.021 (0.57)	–0.017 (0.47)	–0.0341 (0.47)
Speaks a Non-English Language	(a)	–0.059 (1.73)	(a)	0.202 (0.40)
Speaks English Very Well	(a)	(a)	–0.022 (0.61)	0.147 (0.35)
Speaks English Well	(a)	(a)	–0.240 (2.68)	0.038 (0.19)
Speaks English Not Well	(a)	(a)	–0.018 (0.16)	0.016 (0.12)
Speaks English Not at All	(a)	(a)	0.445# (0.77)	0.001 (0.04)
$\bar{R}^2$	0.3682	0.3687	0.3703	
Sample Size	3,447	3,447	3,447	3,447

Notes:
(a) = Variable not entered.
'*t*' statistics in parentheses in first four columns computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator, standard deviations in parentheses in final column.
= Based on fewer than 30 observations.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 1 percent sample.

Acknowledgements

This chapter is based on a report prepared for the Center for Equal Opportunity, August 1998. The views expressed are those of the authors, and do not necessarily reflect the views of their affiliations or of the Center for Equal Opportunity. An abridged version of this chapter was published electronically by the Center for Equal Opportunity in 1998, and is available at: www.ceousa.org/earnings.html (Date accessed: 15 July 2006).

Notes

1 There have been several studies in recent years on the effects of destination language proficiency on the earnings of immigrants in the United States and several other immigrant receiving countries, including Australia, Canada, Israel, Germany, New Zealand and the Netherlands. (See for example, Chiswick and Miller (1995).) These studies consistently show that the lower the degree of proficiency in the destination language, whether measured by speaking skills or by literacy, the lower are the earnings of immigrants.

2 Those living in the U.S. but born in a U.S. territory, who are primarily from Puerto Rico, are excluded from this analysis.

3 For a statistical portrait and analysis of the languages spoken in the U.S. and the people who speak them, see Chiswick and Miller (1996).

4 The term American Indian is used here in preference to Native American because of the possible confusion of the latter term with native-born Americans.

5 Parental nativity was last asked in the decennial census in 1970 but the 1970 Census did not include information on proficiency in English.

6 Although the coefficients of other variables (see the Statistical Appendix) are not discussed here, they are consistent with other studies. Moreover, the coefficients of the other variables are essentially invariant with respect to the inclusion of the language variables.

7 It is likely that a significant proportion of these men are the U.S.-born children of immigrants. In the most recent census in which the earnings for the U.S.-born adult children of immigrants (second-generation Americans) could be studied, they earned about 5 percent more than those with native-born parents (Chiswick, 1977). This implies that if the 1990 Census data made it possible to control for parental nativity the earnings disadvantage of those who speak a language at home other than English would be somewhat greater (*i.e.*, more negative) than what is estimated in Table 14.3.

8 The partial effect on earnings of English-language fluency variables among white *non-Hispanic* men:

Speaks English	*Speaks*	
	Spanish	*Other Languages*
Very Well	−0.00048	−0.042
	(−0.03)	(−4.33)
Well	−0.037	−0.116
	(−0.92)	(−4.59)
Not Well or Not at All	+0.0056	−0.079
	(+0.13)	(−2.17)

Note: '*t*' statistics in parentheses.

9 The partial effect on earnings of English language fluency variables among Hispanic men:

Speaks English	*5 High Spanish Concentration States*[(a)]	*Other States*
Very Well	–0.119	–0.040
	(–7.86)	(–1.79)
Well	–0.213	–0.087[(b)]
	(–8.30)	(–0.24)
Not Well or Not at All	–0.183	–0.068[(b)]
	(–4.13)	(–1.06)

Notes: '*t*' statistics in parentheses. (a) Arizona, California, Florida, New Mexico, and Texas. (b) Very small samples.

References

Becker, G. S. (1964) *Human Capital*, New York: National Bureau of Economic Research.

Chiswick, Barry R. (1977) "Sons of Immigrants: Are They at an Earnings Disadvantage?", *American Economic Review*, Vol. 67, No. 1, pp. 376–380.

Chiswick, Barry R. and Paul W. Miller (1995) "The Endogeneity Between Language and Earnings: International Analyses", *Journal of Labor Economics*, Vol. 13, No. 2, pp. 245–287.

Chiswick, Barry R. and Paul W. Miller (1996) "The Language of the United States: What is Spoken and What it Means", *READ Perspectives*, Vol. 3, No. 2, pp. 5–41.

Chiswick, Barry R. and Paul W. Miller (1998) "Language Skill Definition: A Study of Legalized Aliens", *International Migration Review*, Vol. 32, No. 4, pp. 877–900.

15 Earnings in Canada

The roles of immigrant generation, French ethnicity, and language

I. Introduction

Canada has a heterogeneous population. Nearly a quarter of the population is foreign born; another quarter is native born with one or both parents born outside of Canada. About 30 percent of the population is of French ethnicity; while about three-quarters of those of French ethnicity live in the province of Quebec and constitute a majority of the Quebec population, the remainder live in all parts of the country. As a result, Canada is in a unique situation. It is a "dual language" country. English dominates in nine provinces, French dominates in one province, but there are considerable efforts devoted to developing bilingualism as an official national policy. Immigrant generation, French ethnicity, and language fluency are the three key ethnic-based issues relevant for Canada. While there has been some separate work on these issues, this chapter analyzes jointly their roles in the Canadian labor market.

Heterogeneity among the foreign born

Immigrants in Canada are a heterogeneous group, differing in terms of national origin, endowments of marketable skills, and other demographic characteristics.[1] About one-quarter of the immigrants were born in English-speaking developed countries (principally the British Isles), and another one-quarter are from Southern Europe. About one-sixth came from Northern Europe, and a further sixth from Asia. The Southern Europeans constitute the least skilled, with an average 8.8 years of schooling, which is 3 years less than the national average. The Asians, on the other hand, with an average educational attainment of 13 years, are relatively well educated. The Asians also represent one of the most recent waves of immigrants; in the 1981 Census their average duration of residence in Canada is 11 years, compared to the 20 year average for all the foreign born. Most immigrants report that they are fluent in one of the official languages. Of the male foreign born in the labor market, 97 percent report fluency in either French or English. The proportion is lowest for those from Southern Europe and East Asia (92 and 95 percent, respectively).

This heterogeneity among the immigrants is reflected in their labor market performances. The average earnings of immigrants from English-speaking developed countries (i.e., the United States, the United Kingdom, and the Republic of Ireland) in 1980 was $24,193, compared to the $17,568 average for those born in Southern Europe. To some degree this earnings difference may merely reflect differences in the marketable endowments (e.g., schooling, labor market experience, or language fluency) which immigrants bring to the labor market; but part may also reflect different degrees of adaptability to the Canadian labor market. Rates of adjustment may differ according to the average skill level of the immigrant group, as well as country of origin and motive for migrating.

In the aggregate, immigrants appear to do well compared to the native born. The average annual earnings of all foreign-born men in 1980 was $20,662, some 4 percent more than the $19,916 average earnings for the Canadian born. This comparative labor market success of the foreign born carries over to the nearly one-quarter of the adult male work force who are of foreign parentage (native born with one or both parents foreign born). Data from the 1971 Census indicate that on average these men have annual earnings 13 percent higher than men with native-born parents.

Heterogeneity among the native born

There is also considerable heterogeneity among the native born. Over one-third of native-born Canadians are of French ethnic origin, and nearly all of them have Canadian-born parents. For those of French ethnicity 40 percent speak French but not English, whereas for other native-born Canadians 90 percent speak English but not French. The French-Canadian native-born men in the labor market had 11.0 years of schooling and earned $18,789 in 1980, compared with 11.9 years of schooling and $20,546 of earnings for the native-born "English" (other) Canadians.

Purpose of this analysis

The effects of immigrant status, French ethnicity, and language fluency in Canada has received relatively little attention in the economics literature. Moreover, the coverage of the few studies available is highly selective, and the results often conflict.

This chapter seeks to correct these deficiencies by using data from the 1971 and 1981 Censuses of Canada to analyze the labor market adjustment of immigrants and French Canadians, as well as the role of language in the Canadian labor market. The analysis is conducted in a manner which will facilitate comparison with previous studies of immigrants and their children in Canada and in other countries. In addition, by comparing 1971 and 1981 Census data, the paper highlights the similarity and changes in patterns over the decade.

Section II provides a brief review of the Canadian literature. Section III discusses the framework used for the analysis, particularly the model of immigrant adjustment. The data are described in Section IV. The analyses of earnings for the Canadian born and the foreign born are presented and compared in Sections V and VI, respectively, using the 1981 Census. Section VII analyzes the earnings of second- and later-generation Canadians using the 1971 Census. Section VIII documents changes from the 1971 Census to the 1981 Census in immigrant adjustment patterns and the effects of French ethnicity and language. Section IX provides, by way of conclusion, a brief overview of the major findings relevant for public policy and directions for further research.

II. Review of existing studies for Canada

A number of researchers have analyzed the influence of language and French ethnicity on earnings and the economic position and progress of immigrants and their children in the Canadian labor market, either as the basis of a research project, or peripheral to other concerns.[2]

The effects of immigrant status on earnings

Based on analysis of the 1971 Census data on the earnings of male residents of Toronto who were in full-time civilian employment, Tandon (1978) reports that recent immigrants earn considerably less than comparable Canadian-born men. Immigrants adjust rapidly to the Canadian labor market, and after approximately 16 years of Canadian labor market experience the average foreign-born worker earns about the same as his native-born counterpart. Past this point the earnings of immigrants actually tend to exceed the earnings of the native-born, when other readily measured variables are the same.

In a study of the Ontario labor market, Tandon (1977) reports that there is considerable variation across countries of origin in the labor market performance of immigrants. Recent immigrants from the United States, for example, have earnings higher than the Canadian born, and they maintain this advantage as their duration of residence increases. In contrast, recent immigrants from the British Isles have earnings below those of the native born, but this negative gap narrows with duration of residence and eventually becomes positive. The earnings of immigrants from Western Europe, Southern Europe, Asia, Latin America, and the West Indies never attain equality with the earnings of the native born, even though they are characterized by a narrowing of the earnings differential with increases in duration of residence.

The study of the 1971 Census data has been extended to the entire Canadian labor market by Kuch and Haessel (1979), Richmond and Kalback (1980), and Carliner (1981).[3] Kuch and Haessel analyzed the variation in income across employed persons aged 15 or more years. One focus of their research was the identification of culturally imbued labor market behavior. Consequently, they present empirical estimates for earnings disaggregated by ethnic group but pooled across nativity and birthplace groups. As such their study does not lend itself toward informed analysis of aspects of the international transferability of skills and immigrant adjustment processes. Still, one result of their study is of considerable interest. Three birthplace dummies were included in the regression to distinguish the foreign born according to length of residence in Canada. On this basis, Kuch and Haessel (1979, p. 177) conclude: "after one has controlled for such things as ethnicity, occupation, industry, schooling, experience and hours and weeks worked, recent internal migration and the period of immigration into Canada have no significant residual effects on earnings."

Richmond and Kalback (1980) analyzed the income of males and females aged 15 years and over. They report that though their birthplace, foreign-parentage, ethnicity, and language variables were statistically significant, they contribute very little to the explanation of the variance of incomes. Moreover, from an economic viewpoint, the magnitudes of the partial effects on income of these characteristics are quite small. Comparison of a number of cross-tabulations suggested considerable income variation associated with period of immigration, but this relationship was not subjected to rigorous analysis.[4]

Carliner's (1981) study was restricted to men aged 18–64 years in full-time, full-year employment. Though he concentrated on wage differences among language groups, two important findings relating to immigrant adjustment are presented. First, while very recent immigrants earned far less than immigrants who have been in Canada for some time, immigrants who have been in Canada for more than a decade generally earned higher wages than native-born Canadians of native parents. These differences, however, were not always significant. Second, native-born children of immigrants earned significantly more (2.5 percent) than the children of natives, other things the same.

The results of the Kuch and Haessel, Richmond and Kalback, and Carliner studies contrast with Tomes's (1983) finding that, among the Canadian born, variables indicating foreign-born parents are rarely significant. The evidence they present relating to duration of residence is also inconsistent with Tandon's (1977, 1978) finding that this characteristic has a substantial influence on the earnings of foreign-born Canadians.

The effect of language on earnings

Analysis of the role of language skills has represented an important part of the study of earnings differentials in Canada. A large number of recent studies have examined this issue using 1971 Census data. These studies often differ sharply on the specification of the language variables. Kuch and

Haessel (1979), for example, include a single dummy variable for bilingualism, and Gunderson (1979b) uses a four-way classification (speaks French only; speaks English only; bilingual; unable to speak either of the official Canadian languages). Other authors refine the language variables by combining current facility in the official Canadian languages and "home language," that is, the language most frequently spoken in the current home (Carliner, 1981; Tomes 1983). "Mother tongue," the language spoken in the home when the respondent was a child, is combined with current language skills by Shapiro and Stelcner (1981). Detailed results differ across the various studies, but the general pattern is that bilingualism tends to be rewarded only among native French speakers, and monolingual English-speaking workers earn more than monolingual French speakers. Carliner (1981), however, shows that this latter result depends upon place of residence: earnings of the two monolingual groups do not differ significantly outside of Quebec. Veltman, et al. (1979) use information from all three 1971 Census language questions (current fluency in the official languages; home language; and mother tongue) to distinguish 15 language groups for their study of Montreal. Their emphasis is on different types of language shifts. The findings suggest an economic basis to language transfer and bilingualism in the Montreal labor market.

Carliner (1981) also reports that, holding other things constant, bilingual French-Canadian workers earn significantly less than monolingual English speakers. This suggests an important role for ethnic background in determining rewards in the Canadian labor market, although there is little agreement on this issue. For example, Kuch and Haessel (1979) report that French ethnicity lowers earnings by 8.8 percent, while Tomes (1983) reports that variables indicating ethnic origin were usually statistically insignificant.

Some language studies have been limited to adult men in Quebec using survey data collected during the 1970s. A review of a series of these studies by Vaillancourt and Lacroix (1985) concludes that monolingual French speakers have lower earnings than monolingual English speakers, but English–French bilinguals have no earnings advantage. However, using a 1978 sample for Quebec and a human capital earnings functions approach, Grenier (1985) finds that English–French bilinguals do have significantly higher earnings than English monolinguals or French monolinguals.

III. The statistical and analytical framework

This section discusses the statistical and analytical models that are the bases of performing and interpreting the empirical analysis.

The earnings function

The analysis of earnings among adult men in the Canadian Census is based on the now standard "human capital earnings function." Using this statistical technique, the natural logarithm of earnings is regressed on years of schooling, years of labor market experience, and other relevant labor market and demographic variables (e.g., sector of the economy, or marital status), as well as geographic variables (e.g., region of the country and degree of urbanization).

For the purpose of this study the human capital earnings function is augmented to incorporate explicitly the effects of the foreign origin (foreign born or foreign parentage) and, for the foreign born, the differential effects on earnings in the destination of human capital acquired in the countries of origin and destination. [For the development of this approach and its application to the United States, see Chiswick (1977, 1978).] The analyses of the effects of French ethnicity and language (fluent in English, French, both, or neither language) are performed by including in the earnings function dichotomous variables for these characteristics. It is implicitly assumed that all of the explanatory variables are exogenous, that is, that they are determined by factors other than the individual's earnings. For some purposes, and for some variables, this assumption is too stringent because the person's status with respect to location, economic sector, language, marriage, and labor supply may itself be a function of earnings opportunities. The longer the run considered, the greater the degree of endogeneity. In this study, however, the approach generally adopted in the literature is followed and the more simplifying assumption of exogeneity is maintained in performing the statistical analysis. Endogeneity is considered in interpreting the results.

To permit comparisons with previous Canadian studies and the literature for other countries, and for both immigrant and French-ethnicity analyses, several specifications of the regression equation are presented. In particular, the estimated equation which includes variables for a university degree, government employment, self-employment, and language usage is similar to that used in previous Canadian studies. When these variables are excluded from the model, the specification more closely resembles that used in research into immigrant labor market performance in other countries.

The immigrant adjustment model

There are several distinct but interrelated dimensions to the labor market success (status or adjustment) of immigrants. The analysis of immigrant earnings can, in principle, consider characteristics such as country of origin, motive for migration, selection (immigrant rationing) criteria, level of schooling, and duration of residence. The framework for this analysis is best described by two models of immigrant behavior: one is the degree of skill transferability, and the other is the motive for migrating (see Chiswick 1978, 1979).

Table 15.1 presents a schematic representation of the discussion. For simplicity of exposition, skill transferability is dichotomized between high and low levels; motive for migrating is split between economic and noneconomic factors. The first three entries in each cell in the table describe the level of

Table 15.1 Determinants of labor market adjustment in Canada of the foreign born relative to the native born

Motive for migrating	*Skill transferability*	
	High	*Low*
Economic	1. Easy adjustment 2. Large effect of preimmigration skills 3. Flattest gradient with duration of residence EXAMPLE: U.K. immigrants	1. Lower attainment 2. Lowest effect of preimmigration skills 3. Steeper gradient with duration of residence EXAMPLE: Greek immigrants
Noneconomic[a]	1. Lower attainment 2. Lower effect of preimmigration skills 3. Steeper gradient with duration of residence EXAMPLE: U.S. draft evaders	1. Lowest attainment 2. Lower effect of preimmigration skills 3. Steepest gradient with duration of residence EXAMPLE: Vietnamese refugees

Note
a Includes refugees, ideological migrants, and tied movers.

labor market attainment, the effect in the destination of preimmigration skills, and the speed of adjustment in the destination (i.e., the effect of post-migration experience). The fourth entry gives an example of the type of immigrant in Canada corresponding to that cell.

An important determinant of the earnings of immigrants is the degree of transferability to the destination of the skills acquired in the country of origin. Skill transferability needs to be defined broadly to include not merely the more obvious job or occupation-specific skills but also the degree to which the language and labor market information (including information about labor market characteristics and institutions) obtained in the country of origin are productive in the destination. The level of skill includes schooling, apprenticeship training, less formal on-the-job training, certifications, and licenses. It also includes language skills, such as fluency in the language of the destination—or in the "ethnic enclave" language in certain circumstances.

Transferability can differ sharply by type of skill for workers with similar levels of training. Lawyers and judges, for example, generally have little transferability of their skills because legal systems are country specific. Physicians, on the other hand, have much greater international transferability to their skills.

The smaller the degree of skill transferability, the greater would be the decline in occupational status from the last permanent occupation in the country of origin to the early occupation in the destination. Recent arrivals with low skill transferability would have low occupational attainment, low earnings, and low employment levels, other things the same. Preimmigration skills (such as schooling and labor market experience) have less value in the destination labor market for these immigrants than for immigrants with highly transferable skills.

Because incentives to increase skill transferability would have a large payoff for immigrants with low skill transferability, these immigrants are more likely to be making skill-transferable investments, including language training, schooling, and special preparatory training for certification (or licensing) examinations. In addition, "learning by living" is an important source of improvement in the transferability of previously acquired skills.

As a result, immigrants with low skill transferability may seem to be doing less well during their early years in the destination, but improvements in their economic position are far more rapid with duration in the destination than are the improvements for those with greater skill transferability. Yet, it is unlikely that these disadvantages are ever fully dissipated, other things the same, and although the earnings gap narrows it may never disappear.

Motives for migration can be primarily economic or noneconomic. The economic motive focuses on wage differences between the origin and destination. The noneconomic motives are far broader in scope. Noneconomic migrants include refugees, that is, persons who flee the origin because of a fear of persecution or harm to their person or property. They also include ideological migrants, that is, those who are in no immediate or personal danger but voluntarily move for ideological or nationalistic reasons. A large category of noneconomic immigrants comprises *tied movers*. Tied movers are disproportionately children and married women, persons whose migration decision is not determined primarily by their own labor market opportunities but rather by the migration decision of another family member.

Economic migrants are expected to have a higher attainment and an easier adjustment than are noneconomic migrants. The former base their decision on the economic incentives, and they are more likely to be favorably self-selected for labor market success. This favorable self-selection is expected to be more intense the greater are the relative out-of-pocket costs of migration. For noneconomic migrants other factors have intervened to determine the migration decision. Since labor market success is not a primary determinant of the decision, the favorable self-selection for the labor market is less intense and labor market success is less likely to occur. Among refugees in particular there may be little time to prepare for the transferability of country-specific skills.

Application of the immigrant adjustment model

The skill transferability and motive for migrating models are useful in providing a framework for the empirical application and interpretation of the parameters of an earnings function for immigrants. The empirical implementation of the immigrant adjustment model is, of course, limited by the data that are available. It also should be recognized that the application of the framework may require modification on the basis of the historical experiences and institutional features of a particular country of destination and its immigration.

Data are not available in the Canadian Census on specific skills or labor market experiences in the country of origin. It is assumed, however, that immigrants from developed English-speaking countries have greater transferability for their preimmigration schooling and labor market experience than immigrants from non-English-speaking countries.

Motives for international migration are complex, and many motives, with varying degrees of intensity, may be operative for each of the individuals in a cohort of seemingly identical individuals from a given country of origin at a point of time. The type of visa obtained to gain entry (skill based, kinship based, or refugee) may be an extremely poor guide. Economic migrants may use a kinship (relation) or refugee visa to gain entry. Refugees not eligible for a refugee visa may seek a skill-based or kinship visa.

For the Canadian analysis of immigrant men the tied-mover motivation is likely to be of little importance, except for those brought to Canada by their parents. Refugee flows are not always clearly defined by country of origin and period of migration. For example, during the period of the Vietnam War some migrants from the United States were strongly motivated by political and personal considerations, including draft evasion, but many others were primarily motivated by job opportunities. However, the post-World War II Eastern European immigrants and the post-1975 Indochinese immigrants would be disproportionately refugees.

Two problems with ambiguous implications may affect analyses of immigrant adjustment: one is the problem of return migration; the other is that of changes over time in unmeasured dimensions of immigrant skill or quality. It is known that for some countries of origin and under some circumstances the reemigration of the foreign born is substantial, either back to their country of origin or to some third country (primarily the United States). In other instances it is apparently trivial. In the Canadian context, return migration of U.S. immigrants has been substantial, but the return migration of Eastern European immigrants has been very small. This adds complexities to the interpretation of immigrant adjustment patterns. For example, adjustment profiles estimated for Canada from cross-sectional data are biased upward (downward) if emigration is more (less) common among those who make a poorer labor market adjustment.

Changes in the unmeasured dimensions of immigrant quality may also affect interpretations from immigrant adjustment research. These may arise from changes in the criteria for rationing visas, the stringency of enforcement of immigration law, or the incentives that influence the extent of favorable self-selection (skill transferability and motive for migrating) in the supply of immigrants. (See Jasso and Rosenzweig, 1988.)

French ethnicity and language

Economic theory can be used to generate hypotheses regarding the effects on labor market earnings of ethnicity and language fluency (i.e., types of languages spoken).

According to Becker's (1957) discrimination model, "effective discrimination" against a group (N) is said to exist when the group experiences a larger decline in its income than the other group (W) as a consequence of labor market discrimination. Effective discrimination against N occurs when N is a numerical minority and an economic minority (i.e., a smaller aggregate income). Effective discrimination against a numerical majority may arise only if the group is more of an economic minority than a numerical majority.

Applying this framework to Canada, French Canadians would be a numerical and economic minority in nine provinces, and hence may be subject to effective discrimination. In Quebec the French Canadians are a numerical majority, but it is less clear to what extent they are an "economic minority" in Becker's terminology. It is not clear that they would be subject to effective discrimination in Quebec.

This simple approach needs to be augmented by consideration of government policy. While one may question whether provincial and national government policy favored French Canadians relative to other Canadians in the 1970s, it is clear that the policy trend over the past few decades, including the 1970's was in the direction of reducing French-Canadian disadvantages, particularly in certain sectors of the economy (e.g., government employment) and certain areas of the country (e.g., Quebec). Furthermore, whatever discriminatory differentials existed in 1970 would have presumably diminished over the decade as a result of competitive market forces. A smaller disadvantage or a larger advantage from being French-Canadian, other things the same, would be anticipated for the 1981 Census compared to the 1971 Census.

Even in a world without discrimination, ethnic groups may differ in their stock of human capital and income. Group differences in the trade-offs of quantity for quality of children in one generation can result in systematic group differences in levels of human capital and income in the next generation (Chiswick, 1983b). For example, consider two groups, E and F, that have the same level of wealth in the first generation, but group F faces a lower price of quantity of children (e.g., because it is more rural or the women have a lower value of time in the labor market because of a lower level of schooling) or a higher cost of fertility control (e.g., because of more stringent religious prohibitions of birth control). Then group F would have more children, but each child would receive less investment in its human capital—prior to schooling, during schooling, and concurrent with schooling. The lower human capital would appear in data as a lower level of schooling, a smaller effect of a year of schooling on earnings, lower overall earnings, and lower earnings when other readily measured variables (i.e., years of schooling and experience) are held constant.

In the Canadian context, group F may well be represented by French Canadians. Although the distribution by size of place is similar for French and other Canadians, French Canadians have historically had a higher fertility rate and a lower level of schooling. If so, lower earnings might be expected for French Canadians, even when other measured variables are held constant.

The labor market discrimination, government intervention and quantity/quality fertility models are not mutually exclusive. To one degree or another all may be operative and helpful in understanding differences in demographic, human capital, and labor market outcomes correlated with French ethnicity in Canada. The statistical procedures implemented below estimate the net effect of these and perhaps other factors that determine earnings differences between French and other native-born Canadians.

To the extent that bilingualism means fluency in both English and French it would be anticipated that, other things the same, bilinguals could not be at any earnings disadvantage compared to equally fluent monolingual English or French speakers. If anything, having the additional skill of fluency in a second language may enhance earnings since bilinguals can operate effectively in either the English or French labor markets, serve as a bridge between them, or be employed in jobs that require bilingualism.

Bilingualism may, however, serve as a proxy for other variables. Native-born bilinguals outside of Quebec are more likely to be of French origin, whereas those in Quebec are more likely to be of English origin. Since there may be ethnic group differences in labor market outcomes unrelated to language fluency, it is important to control statistically for the effects of French ethnicity when evaluating the effects of bilingualism. Moreover, since Quebec is predominantly French speaking and the other provinces are predominantly English speaking, the effects of English, French, and bilingual fluency may vary by province.

IV. The data

The analysis in this chapter is based largely on the microdata file of the 1 in 50 Public Use Sample Tapes from the 1981 Census of Canada.[5] The study is limited to male members of the civilian work force between the ages of 25 and 64 who worked for one or more weeks during 1980 and who reported income from either wages and salaries or self-employment activities during that year.[6] In the Statistical Appendix, Table 15.A1 presents the definitions and mnemonic names for the variables used in the analysis; Tables 15.A2 and 15.A3 present means and standard deviations for a number of characteristics by nativity and French ethnicity. Regression estimates are discussed and interpreted in the text and reported fully in the Statistical Appendix tables.

A. Characteristics by nativity

The 1980 average annual earnings of foreign-born adult men in Canada was \$20,662, some 4 percent more than the \$19,916 average earnings for the Canadian born (Table 15.2). Part of this higher average annual earnings may derive from the foreign born being better educated than the native born: on average they have 0.7 years of schooling more than the native born (12.3 years compared to 11.6 years), and a slightly higher proportion have a university degree. With total labor market experience measured by a proxy (the number of years since age 6 that the person was not in school) the foreign born also have 1.7 more years of labor market experience than the native born. There is little difference by nativity in the employment variable, the number of weeks worked in 1980. The foreign born worked 46.7 weeks compared with 46.0 weeks for the Canadian born. Similarly, there is little difference between the percentage of the two groups who are not married: 20 percent of the Canadian born and 16 percent of the foreign born. Self-employment is also equally prevalent (13.7 percent of the native born are self-employed, compared to 14.9 percent of the foreign born); but proportionately fewer of the foreign born are employed in the government sector (5.4 percent, compared to 9.1 percent of the native born).

There are more pronounced regional differences by nativity. Only 25 percent of the foreign born live outside of the major metropolitan centers, compared to 55 percent of the Canadian born. The distribution of the foreign-born population across provinces is quite dissimilar to the distribution of the native-born population, with immigrants having a relatively greater concentration in Ontario and British Columbia.

Finally, there are a number of differences by nativity in the language

Table 15.2 Means and standard deviations of selected variables: adult men, 1981[a]

Nativity and Ethnic Group	*Earnings (\$)*	*Schooling (years)*	*Labor Market Experience (years)*	*French Ethnicity (%)*
All	20,092	11.7	23.0	28.7
	(12,696)	(3.9)	(12.7)	—
Native born	19,916	11.6	22.6	35.8
	(12,620)	(3.8)	(12.8)	—
"English"[b]	20,546	11.9	22.3	0.0
	(13,094)	(3.6)	(12.8)	—
French	18,789	11.0	23.1	100.0
	(11,639)	(4.0)	(12.8)	—
Foreign born	20,622	12.3	24.3	3.6
	(12,953)	(4.4)	(12.3)	—

Notes

a Men aged 25–64 who worked and had nonzero earnings in 1980. Standard deviations in parentheses.

b Persons not of French ethnic origin.

Source: Statistical Appendix Tables 15.A2 and 15.A3.

variables. Only 2.7 percent of the foreign born reported that they spoke neither of the official languages. The foreign born are more likely than the native born to speak English but not French. Four-fifths of the foreign born speak English but not French, compared to three-fifths of the native born.

Characteristics by French ethnicity

Among the native born, there are a number of striking differences between the one-third of the population who are of French ethnicity and other ethnic groups. Native-born "English" (i.e., non-French) Canadians have, on average, earnings 5 percent higher than native-born French Canadians. In part this reflects their having almost one year of education more than the French Canadians and a greater proportion with a university degree (15 percent compared to 11 percent). Part of the difference also arises from their working on average 1.3 weeks per year more than those of French ethnicity (46.5 weeks compared to 45.2 weeks).

It is expected that the language and location characteristics of the two populations may impact on the average earnings. Three-quarters of French Canadians live in Quebec; around one-half are English–French bilinguals, but 40 percent speak only French. Among other Canadian-born men, however, only 5 percent are residents of Quebec and only 8 percent reported they are bilingual. The direction of the impact on average earnings of these differences is difficult to determine a priori.

V. Earnings of the Canadian born

The regression analyses of earnings for Canadian-born men using the 1981 Census largely accord with findings derived from analyses of earlier data. This section first discusses the human capital and demographic variables, then considers language differences, and closes with analyses of the effects of French ethnicity.[7]

Human capital and demographic variables

Human capital acquired through schooling and on-the-job training is an important determinant of earnings. Additional years of schooling which do not result in the award of a university degree raise earnings by a modest 3.9 percent.[8] In contrast, possession of a university degree raises earnings by a highly significant 18 percent. As argued by Kuch and Haessel (1979), this rather substantial earnings differential may reflect the credential effect associated with possession of a degree, or it could derive from degree holders being relatively more able. In part, however, it also reflects nonlinearities in the partial effect of years of schooling on the natural logarithm of earnings.[9] The sizable impact of a university degree is lower than was generally established using the 1971 Census data. Part of the explanation for this change may be found in differences in the coding of the schooling data, rather than in a decline in the reward to tertiary-level education.[10]

The earnings function was also computed with explanatory variables similar to those used in recent studies of labor markets in other countries. In particular, the employment type, university degree, and language variables have been omitted, but a French-ethnic origin variable is added [see Statistical Appendix Table 15.A4, columns (2) and (3)]. The coefficients of most of the explanatory variables are not affected by the change in the specification of the function. However, when the university degree variable is deleted, the partial effect of years of schooling increases from 3.9 to 5.2 percent. It still falls short of the average coefficient of around 7 percent estimated for the United States (Chiswick, 1977, 1978) and the 8 percent estimated for Australia (Chiswick and Miller, 1985).

Earnings also increase sharply with years of labor market activity: an additional year of labor market experience is associated with a 2.3 percent increase in earnings (evaluated at 10 years of labor market experience). After around 30 years of labor market activity the earnings profiles level off, however, and then decline.

As with most previous studies of annual earnings, weeks worked is the most important explanatory variable. The elasticity of annual earnings with respect to weeks worked is 0.918 and is significantly less than unity. That is, the percent increase in earnings is less then the percent increase in weeks worked. Thus, individuals who work full-year earn a lower weekly wage than individuals who work part-year. As noted by previous researchers (Chiswick, 1974; Kuch and Haessel, 1979), this may reflect a weekly wage premium for workers in seasonal employment. In the United States, where seasonality is less important, the coefficient is unity.

Consistent with expectations, men who are not currently married are at a substantial earnings disadvantage relative to the married men. The regression coefficient of –0.28 on the marital status variable implies about one-quarter lower weekly earnings for men who are not married. Although large, this coefficient is in the same order of magnitude as was found in other Canadian studies (e.g., Tomes, 1983) and in studies for other countries. Earnings levels also differ according to employment type: self-employment is associated with 22 percent lower earnings, whereas employment in the government sector is associated with an earnings premium of 5 percent.[11]

There are substantial interregional and interprovincial earnings differentials. The earnings gap between residents of metropolitan and nonmetropolitan areas is 8 percent. The interprovincial earnings structure indicates that residents of British Columbia and the Prairie provinces experience higher earnings than residents of Ontario, ceteris paribus. Residents of Newfoundland and the Maritime provinces have lower earnings when other things are the same, while the earnings of residents of Quebec and Ontario do not differ significantly. It is suggested that these patterns reflect either substantial costs of labor mobility that prevent internal migration from equalizing nominal and real earnings across the provinces (Robinson and Tomes, 1982a) or substantial differences across provinces in the cost of living (Daniel, 1982).

Language variables

Among the native born there is a wage premium for bilingualism. In the 1981 Census data, Quebec residents who speak only French have 1.6 percent lower earnings than those who are English-French bilingual [Statistical Appendix Table 15.A4, column (1)]. Non-Quebec-dwelling English speakers have 3.9 percent lower earnings than those who are bilingual.[12] A notable feature of the language results is that the earnings of monolingual English-speaking workers residing in Quebec do not differ significantly from the earnings of the bilingual reference group.

Analyses of the 1971 Census found that monolingual English-speaking workers generally earned more than the otherwise comparable monolingual or bilingual French-Canadian workers (see Carliner, 1978). Thus, the data suggest that the earnings differences between the language groups has been diminishing over time.[13]

French ethnicity

Men of French ethnic origin are found to have earnings 3.3 percent higher than other Canadian-born men, other things the same, and the coefficient is highly significant (t = 4.6). This contrasts with the 8.8 percent earnings disadvantage for this group reported by Kuch and Haessel (1979), and the 1 percent lower earnings reported by Richmond and Kalback (1980) using 1971 Census data. Chiswick and Miller (1984), applying the same methodology to the 1971 Census, have found a significant negative differential of 2.6 percent. Thus, if we use the same statistical methodology, over the decade of the 1970s French Canadians have shifted from being at about a 3 percent earnings disadvantage to being at about a 3 percent earnings advantage. Thus, whether due to explicit public policy or other factors, the labor market situation of French Canadians has shown dramatic improvement.[14]

A Chow test revealed there is a significant difference in the set of regression coefficients between native-born French Canadians and English Canadians.[15] The sign, magnitude, and statistical significance of most variables are very similar across the earnings equations. The two most noticeable differences concern the language variables and the employment variable, *weeks worked*.

Among French Canadians, those who live in Quebec and only speak French have lower earnings than other French Canadians. But among other Canadians, it is those who speak only English and live outside of Quebec who have lower earnings. Thus, there is an interaction effect. For *both* French and

English Canadians being monolingual and living in a province in which this is the dominant language results in about 4 percent lower earnings than being bilingual.

The coefficient on weeks worked is significantly different ($t = 6.27$) between the groups; the elasticity of earnings with respect to weeks worked is 0.87 for French Canadians and 0.94 for other Canadians. This suggests greater seasonality in employment or a lower-elasticity (or more background-bending) labor supply curve among the French Canadians.

A similar pattern emerges from an examination of earnings differences among the French ethnics according to whether they are residents of Quebec [columns (3) and (4) of Appendix Table 15.A5]. A Chow test on the equality of the coefficients for this disaggregation resulted in a statistically significant F statistic of 2.68. Again this structural difference appears to derive mainly from the language and weeks worked variables. Monolingual French-speaking French-Canadian residents of Quebec have lower earnings than other French Canadians. The lower coefficient on weeks worked for French Canadians in Quebec is consistent with a greater seasonality of employment in the province.

VI. Earnings of the foreign born

A major focus of this study is the earnings performance of the foreign born. The section begins with a comparison of native- and foreign-born Canadians, and then considers differences among the foreign born. (The detailed data on means of variables are presented in Statistical Appendix Tables 15.A2 and 15.A7, and the regression equations are reported in Tables 15.A4, 15.A6, and 15.A8.)

Native-born vs. foreign-born comparison

Schooling has a statistically significant effect on the earnings of immigrants in Canada, but the partial effect is smaller for them than for the Canadian born.[16] An additional year of schooling that does not lead to the award of the university degree raises earnings among the foreign born by 2.5 percent, around one and a half percentage points less than the effect schooling has on the earnings of the native born. When the university degree variable is excluded from the specification the partial effect of an extra year of schooling is somewhat greater, 3.9 percent, but this is still more than 1 percentage point less than the 5.2 percent partial effect for the Canadian born (Table 15.3).

Interestingly, for the foreign born the partial effect on earnings of a university degree is 22.7 percent, which is 4.4 percentage points higher than for the native born, and this difference is statistically significant. Without a better understanding of the reasons for the large effect of the university degree

Table 15.3 Partial effect on earnings of schooling, labor market experience, and duration in Canada, 1981 and 1971[a]

Census Year/ Country of Origin	*Schooling*		*Labor Market Experience in Origin*		*Duration in the Destination*
	Native Born	*Foreign Born*	*Native Born*	*Foreign Born*	*Foreign Born*
1981 Census:					
All	5.2	3.9	2.30	1.48	1.54
ESDC[b]	—	5.2	—	2.24	0.60
Other	—	3.3	—	1.14	2.00
1971 Census:					
All	5.6	4.1	2.02	1.42	1.39
ESDC[b]	—	5.1	—	2.06	0.32
Other	—	3.1	—	1.04	1.82

Notes

a Regression analysis of annual earnings for men aged 25–64. Separate regressions by nativity. Controlling for schooling, experience, marital status, weeks worked, metropolitan area, province, and duration of residence for the foreign born. Experience and duration evaluated at 10 years.

b ESDC means immigrants from the English-speaking developed countries (i.e., the United States, the United Kingdom, and the Republic of Ireland).

Source: Chiswick and Miller (1984); Statistical Appendix Tables 15.A4 and 15.A8.

variable among the native born, it is difficult to interpret the economic relevance of the difference by nativity.

There are rather sizable differences between the partial effects on earnings of preimmigration and postimmigration experience. Additional years of preimmigration experience raise earnings by 1.5 percent (evaluated at 10 years of experience), and this is less than the 2.3 percent earnings growth experienced by the native born (Table 15.3). However, an additional year of postimmigration experience (holding constant total experience) raises earnings by 3.2 percent (evaluated at 10 years), and this exceeds the effect of experience on earnings for the native born. The lower impact of preimmigration experience presumably reflects the less-than-perfect international transferability of skills. The much higher growth of earnings with postimmigration experience presumably reflects the effects of adjustment to the Canadian labor market—the learning of the language and culture, acquiring Canadian-specific labor market information, the modification of existing skills, and the acquisition of new skills rewarded in the Canadian labor market.

In large part because schooling and labor market experience acquired in the country of origin is of less value in the Canadian labor market, the average newly arrived immigrant has earnings about one-quarter lower (coefficient of –0.31) than a comparable native-born worker.[17] But as the partial effect on earnings of a year of Canadian labor market experience is greater for the foreign born than for the native born, this earnings gap narrows with increases in duration of residence. In the 1981 Census data, the earnings of immigrants reach equality with that of the native born after around 22 years of residence, and thenceforth exceed the earnings of the native born.

Recall that among the native born, residents of nonmetropolitan areas received earnings 7.5 percent lower than their metropolitan area counterparts. For the foreign born, however, there is no evidence of earnings differences on the basis of size of place. This suggests that the earnings differences among the native born are more likely to be reflecting monetary and psychic costs of moving from smaller to larger places, rather than merely reflecting a lower real cost of living in nonmetropolitan areas.

Another contrast to the results for the native born is the different ranking of provinces by earnings. Among the foreign born, the earnings of residents of the Maritime provinces do not differ significantly from those of the reference province, Ontario, while residents of Newfoundland, the Prairie provinces, and British Columbia experience significantly higher earnings, and residence in Quebec is associated with statistically significant (7.0 percent) lower earnings. Residence in Quebec was not associated with a significant earnings difference compared to Ontario among the native born. This suggests that relative to natives the labor market experience of Quebec-dwelling immigrants is inferior to that of immigrants in other provinces. This may indicate why immigrants tend to avoid Quebec.

Employment in government, a high-wage sector for the native born, is not associated with a statistically significant earnings premium among the foreign born. This result is consistent with finding a lower private sector-government sector wage differential for the foreign born in Australia (Chiswick and Miller, 1985).

Self-employment is associated with lower earnings. While the Canadian-born self-employed also receive lower earnings, the magnitude of the earnings gap is somewhat smaller among the foreign born, 17.6 percent compared to the 21.7 percent relative earnings differential among the native born [compare columns (1) and (4) in Appendix Table 15.A4].

Analyses among the foreign born

Among the foreign born, earnings vary systematically by language group even when other variables, including duration of residence in Canada, are the same [Appendix Table 15.A4, column (4)]. Among immigrants, English–French bilinguals have the highest earnings. Significantly lower earnings are received by French-speaking immigrants in Quebec (13.5 percent), English-speaking immigrants in Quebec (10.2 percent), and English-speaking immigrants in other provinces (5.9 percent). There is a highly significant 18 percent earnings penalty among the foreign born, associated with an inadequate command of both of the official languages.

The transferability of immigrant skills can be expected to vary with the similarity of the country of origin to Canada. A high degree of transferability would be expected for immigrants from the English-speaking developed countries, primarily the United States, the United Kingdom, and the Republic of Ireland. A similar transferability might be expected among French-origin immigrants in Quebec. However, only 4 percent of the foreign born reported a French origin, and a separate statistical analysis for such a small sample would not be meaningful.

Earnings vary considerably among the foreign born by country of origin. Average annual earnings of immigrants are highest for those from the British Isles ($24,495), with successively lower earnings received by immigrants from the United States ($23,060), Western Europe ($22,335), Eastern Europe ($20,546), Asia ($18.705), and Southern Europe ($17,568).[18] The earnings gap is narrowed when other variables are held constant (see Table 15.4). Compared to immigrants from the United States, those born in the British Isles have a significantly higher earnings (by 7 percent), while the earnings of immigrants from Southern and Western Europe do not differ significantly from U.S.-born immigrants. Significantly lower earnings are received by immigrants from Eastern Europe (5 percent) and Asia (12 percent).

Among immigrants from developed English-speaking countries, duration of residence is highly significant and indicates an earnings growth of 0.6 percent with each additional year of Canadian labor market experience, ceteris paribus (Table 15.3).[19] For immigrants from non-English-speaking countries, however, duration of residence exerts a more important influence. Evaluated at 10 years, the partial effect is 2.0 percent. Earnings increase at a decreasing rate with duration of residence for the non-English-speaking aggregate and for three of the five component birthplaces. In the case of those born in Southern and Western Europe the linear duration of residence variable is statistically significant, although the squared term is not.

The partial effects of schooling and preimmigration experience are greater the greater is the international transferability of skills (Table 15.3). The coefficient of education is higher among immigrants from English-speaking developed countries, 5.2 percent (t = 19.26), compared to 3.3 percent (t = 23.03) for other countries. The partial effect of education is lowest among the Southern and Eastern Europeans, 1.2 percent (t = 3.83) and 2.1 percent (t = 2.48), respectively. Similarly, returns to labor market experience in the country of origin are higher for immigrants from the English-speaking developed countries, 2.2 percent, compared to 1.1 percent for immigrants from other countries (evaluated at 10 years of experience). The partial effects of preimmigration experience are smallest for those from Southern and Eastern Europe.

Table 15.4 Annual earnings and relative difference in earnings among the foreign born by country of origin, 1971 and 1981

Country of Birth	*1981 Census*				*1971 Census*	
			Relative Difference in Earnings		*Relative Difference in Earnings*	
	Sample Size (Percent of Foreign Born)	*Annual Earnings*	*Overall*[a]	*Other Things the Same*[b]	*Overall*[a]	*Other Things the Same*[b]
United States	1,263 (5.5)	23,060	Benchmark		Benchmark	
British Isles[c]	4,735 (20.6)	24,495	0.148 (5.99)[d]	0.069 (3.31)	0.112 (3.51)	–0.018 (0.68)
Western Europe	3,174 (13.8)	22,335	0.042 (1.61)	–0.035 (1.54)	–0.017 (0.52)	–0.089 (3.00)
Southern Europe	5,352 (23.3)	17,568	–0.192 (7.80)	0.029 (0.13)	–0.185 (5.74)	–0.151 (4.94)
Eastern Europe	2,125 (9.3)	20,546	–0.065 (2.36)	–0.050 (2.10)	–0.069 (2.03)	–0.091 (3.04)
Asia	3,053 (13.3)	18,708	–0.213 (8.21)	–0.116 (5.13)	–0.195 (4.56)	–0.195 (5.29)
Remainder	3,267 (14.2)	19,530	–0.128 (4.99)	–0.047 (2.11)	–0.082 (1.89)	–0.110 (2.91)
Total	22,969 (100.0)	20,669				

Notes
a Difference in the mean of the natural logarithm of earnings.
b Controlling for education, total experience, weeks worked, size of place and province, marital status, and duration of residence in Canada.
c British Isles includes the United Kingdom and the Republic of Ireland.
d The *t* statistics are in parentheses under the regression coefficients.

Source: 1971 Census—Chiswick and Miller (1984); 1981 Census—Statistical Appendix tables.

Summary

The pattern of the effects on earnings of schooling, preimmigration labor market experience, and postimmigration experience found in Canada is consistent with the hypotheses developed in Section III and with analyses for the other major immigrant-receiving countries, namely, the United States, Australia, the United Kingdom, and Israel (see Chiswick, 1979, 1980; Chiswick and Miller, 1985). In Canada, as in these other countries, schooling and preimmigration labor market experience have a smaller partial effect than does the training of the native born, but postmigration experience has a larger effect. Furthermore, among the foreign born, schooling and premigration experience have a larger effect and postmigration experience a smaller effect the greater the transferability of skills (that is, for immigrants from other English-speaking developed countries). Opposite effects are found for immigrants from countries which include a disproportionate number of refugees.

VII. Second-generation vs. later-generation Canadians

There are a number of reasons why foreign parentage may affect the earnings of native-born men (Chiswick, 1977). Some factors would tend to lower earnings, such as discrimination against first- and second-generation immigrants in wages and access to jobs, and disadvantages from being raised in homes where the native language is not spoken (or not spoken well) or which are less familiar with the local labor market institutions. On the other hand, the ability and motivation characteristics associated with self-selection in migration of the parents may be acquired in the home or inherited, and these may raise earnings. Furthermore, it is noted that the effects may vary according to which parent is foreign born.

While the information required for analysis of the effects of foreign parentage are not available from the 1981 Canadian Census, the 1971 Census does allow this issue to be investigated. Individuals with at least one foreign-born parent have earnings 13 percent higher than those with native parentage. Part of this substantial earnings advantage may derive from their having almost one year of schooling and two years of experience more than the native-parentage men. Part may also derive from a smaller percentage being nonmarried (15.5 percent compared to 17.6 percent). Differences in location, however, may either increase or decrease the relative earnings of the foreign-parentage group; fewer of the children of immigrants live in towns and rural nonfarm areas, a pattern of settlement which would tend to increase their earnings. However, they have a slightly greater proportional representation in the relatively low-income farm areas. There is also a greater concentration of the children of immigrants in the Prairie provinces (a relatively low-income region in 1971) and British Columbia (a relatively high-income region).

Other things the same, the Canadian-born children of immigrants have earnings 2.0 percent higher than those of native-parentage Canadians (t = 2.24).[20] Compared with native-parentage men, those with only one foreign-born parent have higher earnings—by 2.7 percent if the mother was foreign born and 1.0 percent if it was the father. In the more frequent case where both parents are foreign born (55 percent of foreign-parentage men), earnings are higher by a statistically significant 2.3 percent (t = 2.1).

The finding that the native-born children of immigrants earn around 2 percent more than the children of natives is broadly consistent with the 2.5 percent income advantage reported by Carliner (1981). It is also consistent with the U.S. and Australian patterns, where the native-born sons of immigrants earn about 5 and 1 percent, respectively, more than native-parentage men, other things the same (Chiswick, 1977; Chiswick and Miller, 1985).[21]

VIII. Changes over time: 1971–1981

Has the comparative success of French Canadians and immigrants in the labor market changed over the decade of the 1970s? Regressions were computed using the 1971 Census that, as far as was possible, correspond to the definitions and specifications employed in this analysis of the 1981 Census.[22]

French ethnicity and language

The most striking difference between the 1981 and 1971 Censuses is the effects on earnings of the French ethnic origin and language variables. Among the native born, the 2.6 percent earnings disadvantage (t = –2.40) associated with a French ethnic origin in 1971 has been reversed to a 3.3 percent income advantage in 1981 (t = 4.63). The 6 percent relative improvement in the earnings of French Canadians, other things the same, is statistically significant.

Moreover, the premier position in the Quebec wage structure of native-born monolingual English workers has apparently been eroded. In 1981 their earnings did not differ significantly from the bilingual reference group (coefficient = –0.031; t = –1.20), whereas in 1971 they were at a 6.5 percent earnings advantage (t = 2.38).

Compared to those who are bilingual, the monolingual English speakers residing outside of Quebec experienced a statistically significant 2–3 percentage point decline in their earnings among both the native and foreign born. On the other hand, native-born monolingual French speakers in Quebec experienced a dramatic reduction in their disadvantage relative to bilinguals (from 8.1 percent lower earnings in 1971 to 1.6 percent lower earnings in 1981).

Thus, the disadvantage from being of French ethnic origin or of being monolingual French speaking clearly declined or disappeared over the decade, while the advantage of being a monolingual English speaker also declined and disappeared during this period.

Immigrants and natives

The coefficients of the other parameters of the earnings function show remarkably little change from the 1971 Census to the 1981 Census for both the Canadian and foreign born (see, for example, Table 15.3). The coefficients of the weeks worked variable were higher and the self-employment variable were lower in 1981 than in 1971. These changes may be reflecting the greater recessionary stage of the economy in the later time period; the unemployment rate was 5.9 percent in 1970 and 7.5 percent in 1980.

There is, however, a major change in the intercept. It appears that the height of the earnings profile of immigrants relative to natives declined by about 7–9 percentage points at nearly all levels of duration of residence in

Canada. The decline was 8 percentage points at one year of residence, 9 percentage points at 10 years, and 7 percentage points at 20 years. The decline in the height of the profile is not solely attributed to a change in the immigrant mix by country of origin. The decline is found for all countries of origin, ranging from a small decline for immigrants from the British Isles and Southern Europe (0–4 percentage points) to a large decline for those from the United States (more than 20 percentage points) and Asia (about 30 percentage points).[23]

The decline in the relative earnings of immigrants may be caused by the same factors that resulted in the relative increase in the returns from labor market experience for both the native and foreign born (Table 15.3). These factors may also depress the relative earnings of workers with fewer skills relevant in the labor market when controlling for years of schooling and years in the labor market. Immigrants have fewer firm-specific and country-specific skills than do the native born, ceteris paribus. The increase in the return to training in the 1981 Census may have been the result of a combination of the recession and the recent influx into the labor market of the low-experience post-World War II baby-boom generation.

IX. Summary and conclusions

Immigrants constitute one-quarter of the Canadian adult male work force, and the children of immigrants a further one-quarter. French Canadians, nearly all of whom were born in Canada, are about three-tenths of the work force. Immigrant generation, French ethnicity, and fluency in one or both of the official languages are the primary focus in this study of earnings in Canada.

Immigrant generation

The average newly arrived immigrant in the 1981 Census had earnings that were only about three-quarters of the earnings of a comparable native-born worker. In large part this earnings disadvantage reflects the effects of less-than-perfect international transferability of human capital. However, as the partial effect on earnings of Canadian labor market experience is greater for the foreign born than for the Canadian born, this negative income gap closes with duration of residence. The earnings of immigrants reach equality with those of the native born after around 22 years of residence, and thenceforth exceed the earnings of the native born, according to these cross-sectional regressions that do not control for sample attrition (re-emigration) or vintage effects (cohort quality) across immigrant cohorts.

An extra year of schooling increases the earnings of the native born by 5.2 percent and the earnings of immigrants by 3.9 percent, other things the same. The partial effect of schooling on earnings is greater for immigrants from the English-speaking developed countries (5.2 percent) than it is for other immigrants (3.3 percent). The analyses indicate a substantial earnings disadvantage for immigrants who do not possess facility in either of the official languages.

The earnings profile of immigrants relative to the native born appears to be 7–9 percentage points lower in the 1981 Census than in the 1971 Census, at all durations of residence. The earnings profile appears to have declined for all countries of origin, with the decline being greatest for immigrants from the United States and Asia. The general decline may be due to the same factors responsible for the relative decline in earnings for workers with less experience in the Canadian labor market.

There are major differences in the distributions of the Canadian-born and foreign-born work forces across the provinces of Canada, with the high-income provinces of Ontario and British Columbia having a proportionately greater representation of the foreign born. The proportion of immigrants in Quebec is very small. Immigrants in Quebec have the lowest earnings relative to the native born and have low earnings relative to immigrants in other provinces.

The labor market performances of the native-born children of immigrants and the children of natives are very similar. One notable difference, however, is that the former have earnings a significant 2 percent higher than the native-parentage group, when other things are the same. This may reflect favorable characteristics acquired from their parents, who were presumably self-selected for migration.

What is perhaps most striking about the general pattern of immigrant adjustment in Canada is the similarity to the patterns found in the other major immigrant-receiving developed countries that have been studied in the past decade.

French ethnicity and language

Among the countries for which there have been systematic econometric analyses of earnings, Canada is unique in being a bilingual country in which one language is dominant in one area and another language is dominant in the rest of the country. Government efforts to promote bilingualism as a national policy are also unique. The analysis indicates that disadvantaged ethnic and language groups can experience considerable improvement in their economic well-being in a relatively short period of time, although it is not yet possible to determine whether these changes have been influenced by government policies or other factors.

The 1981 Census data suggest that French Canadians have 3 percent higher earnings than other Canadian-born men, other things the same, a reversal of the position (3 percent lower earnings) that was evident in the 1971 Census. There has also been a decline in the disadvantage from being monolingual French as well as a decline in the advantage from being monolingual English. Among the native born there is evidence of an earnings premium for bilingualism for French Canadians living in Quebec who learned English and for other Canadians living outside of Quebec who learned French. That is, among both French and English Canadians, being monolingual and living in a province in which this is the dominant language results in lower earnings (by about 4 percent) than are received by those who are bilingual. With such an economic reward for bilingualism in Canada, its incidence is likely to increase.

Unresolved issues

This study has expanded our understanding of the role of immigrant generation, French ethnicity, and language in the Canadian labor market. It has also demonstrated patterns that are similar to and different from the experiences of other countries. However, several important issues warrant further analysis. These include the reasons for the decline in the relative earnings of immigrants and the improvement for French Canadians over the decade of the 1970s. More research is also needed on the determinants of language fluency, particularly bilingualism in English and French, and the mechanisms through which language fluency impacts on labor market earnings in Canada.

Statistical appendix

Table 15.A1 Definitions of variables and mnemonic names (Analysis of the 1981 Canadian Census)

Mnemonic	*Definition*
Dependent Variable	
LNY	Natural logarithm of earnings (income from wages and salaries plus income from self-employment) in 1980
Human Capital and Demographic	
EDUC	Years of schooling
EDUCA	Years of schooling acquired by immigrants after arrival in Canada
EXP	Labor force experience = (age – EDUC – 6)
DEGREE	Possesses a university degree
NONMAR	Single, widowed, divorced, or separated
Employment Status	
LNWEEKS	Natural logarithm of weeks worked in 1980
GOVT	Employed in the public sector (industry coded as public administration or defense)
SELF	Self-employed in incorporated or unincorporated enterprises
Size of Place	
NONCMA	Not a resident of a Census Metropolitan area (a Census Metropolitan area is defined as a place having 100,000 or more population)
Province	
ONT	Ontario (reference group)
NFL	Newfoundland
MARIT	Maritime provinces (Nova Scotia, New Brunswick, Prince Edward Island)
QUE	Quebec
PR	Prairie provinces (Manitoba, Saskatchewan, Alberta)
BC	British Columbia
Language and Ethnicity	
BILING	Fluent in both English and French (reference group)
ENGONLY(NON-QUE)	Fluent in English only and does not live in Quebec
ENGONLY (QUE)	Fluent in English only and lives in Quebec
FRONLY(NON-QUE)	Fluent in French only and does not live in Quebec
FRONLY(QUE)	Fluent in French only and lives in Quebec
NENF	Fluent in neither English nor French
FRETH	French ethnic origin (single or multiple origins)
Nativity	
FOR	Individual born overseas
RESID	Years of residence in Canada, defined to equal zero for the native born

Table 15.A2 Means and standard deviations of variables for adult males in Canada by nativity, 1981

Variable	*Born in Canada*		*Foreign Born*		*All Persons*	
	Mean	*Standard Deviation*	*Mean*	*Standard Deviation*	*Mean*	*Standard Deviation*
Dependent Variable						
Earnings ($)	19916.	12620.	20662.	12943.	20082.	12696.
LNY	9.663	0.825	9.714	0.789	9.674	0.817
Human Capital and Demographic						
EDUC (years)	11.566	3.764	12.262	4.369	11.720	3.916
EDUCA (years)	—[a]	—[a]	2.124	4.488	—[a]	—[a]
EXP (years)	22.601	12.802	24.328	12.269	22.983	12.706
DEGREE (%)	13.9	34.6	17.5	38.0	14.7	35.4
NONMAR (%)	19.9	40.0	15.5	36.2	19.0	39.2
NONCMA (%)	54.6	49.8	25.1	43.4	48.1	49.9
Employment Status						
WEEKS	46.025	11.159	46.677	10.194	46.170	10.956
LNWEEKS	3.771	0.420	3.796	0.378	3.777	0.411
GOVT (%)	9.1	28.8	5.4	22.5	8.3	27.6
SELF (%)	13.7	34.6	14.9	35.6	14.2	34.9
Language and Ethnicity						
BILING (%)	22.6	41.8	14.2	34.9	20.6	40.4
FRONLY(NON-QUE)(%)	0.3	5.9	0.1	3.4	0.3	5.4
FRONLY(QUE) (%)	14.7	35.4	2.7	16.3	12.1	32.6
ENGONLY(NON-QUE)(%)	61.5	48.7	76.7	42.3	64.9	47.7
ENGONLY(QUEB) (%)	0.9	9.3	3.6	18.5	1.5	12.0
NENF (%)	0.0	1.9	2.7	16.2	0.6	7.9

(Continued Overleaf)

Table 15.A2 Continued

Variable	*Born in Canada*		*Foreign Born*		*All Persons*	
	Mean	*Standard Deviation*	*Mean*	*Standard Deviation*	*Mean*	*Standard Deviation*
FRETH (%)	35.8	48.0	3.6	18.7	28.7	45.2
Nativity						
FOR (%)	0.0	0.0	100.0	0.0	22.1	41.5
RESID (years)	—[a]	—[a]	19.217	10.441	—[a]	—[a]
Province						
ONT (%)	31.7	46.5	53.9	49.8	36.6	48.2
NFL (%)	2.5	15.7	0.3	5.9	2.1	14.2
MARIT (%)	6.9	25.4	1.5	12.1	5.7	23.2
QUE (%)	29.9	45.8	14.3	35.0	26.4	44.1
PE (%)	18.3	38.7	14.0	34.7	17.4	37.9
BC (%)	10.7	30.9	15.9	36.6	11.8	32.3
Sample size	80,795		22,969		103,764	

Note

a Variable not applicable or not calculated.

Source: 1981 Census of Canada.

Table 15.A3 Means and standard deviations of variables for adult Canadian-born males by French ethnicity, 1981

Variable	*French Ethnicity*		*Not of French Ethnicity*		*All Native Born*	
	Mean	*Standard Deviation*	*Mean*	*Standard Deviation*	*Mean*	*Standard Deviation*
Dependent Variable						
Earnings ($)	18789.	11639.	20546.	13094.	19916.	12620.
LNY	9.621	0.778	9.686	0.848	9.663	0.825
Human Capital and Demographic						
EDUC (years)	10.991	3.952	11.887	3.615	11.566	3.764
EXP (years)	23.063	12.815	22.343	12.788	22.601	12.802
DEGREE (%)	11.3	31.7	15.3	36.0	13.9	34.6
NONMAR (%)	19.2	39.4	20.4	40.3	19.9	40.0
NONCMA (%)	54.0	49.8	54.9	49.8	54.6	49.8
Employment Status						
WEEKS	45.155	11.775	46.511	10.770	46.025	11.159
LNWEEKS	3.745	0.441	3.786	0.407	3.771	0.420
GOVT (%)	9.5	29.3	8.9	28.5	9.1	28.8
SELF (%)	18.8	32.3	15.1	35.8	13.7	34.6
Language and Ethnicity						
BILING (%)	48.7	50.0	7.9	27.0	22.6	41.8
FRONLY(NON-QUE)(%)	0.9	9.5	0.0	1.5	0.3	5.9
FRONLY(QUE) (%)	39.9	49.0	0.6	8.0	14.7	35.4
ENGONLY(NON-QUE)(%)	10.3	30.4	90.1	29.9	61.5	48.7
ENGONLY(QUE) (%)	0.2	4.2	1.3	11.1	0.9	9.3
NENF (%)	0.0	0.0	0.1	2.4	0.0	1.9
FRETH (%)	100.0	0.0	0.0	0.0	35.8	48.0

(Continued Overleaf)

Table 15.A3 Continued

Variable	*French Ethnicity*		*Not of French Ethnicity*		*All Native Born*	
	Mean	*Standard Deviation*	*Mean*	*Standard Deviation*	*Mean*	*Standard Deviation*
Province						
QUE (%)	74.5	43.6	5.0	21.8	29.9	45.8
NON-QUE (%)	25.5	43.6	95.0	21.8	70.1	45.8
NFL (%)	0.3	5.7	3.8	19.0	2.5	15.7
MARIT (%)	4.9	21.6	8.0	27.2	6.9	25.4
ONT (%)	13.2	33.9	42.0	49.3	31.7	46.5
PR (%)	4.7	21.2	25.9	43.8	18.3	38.7
BC (%)	2.4	15.2	15.3	36.0	10.7	30.9
Sample size	28,953		51,842		80,795	

Source: 1981 Census of Canada.

Table 15.A4 Regression analysis of earnings for adult males in Canada by nativity, 1981[a,b] (Dependent variable: natural logarithm of earnings)

Variable	*Canadian Born*			*Foreign Born*		
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Constant	5.445 (200.12)	5.320 (203.16)	5.332 (204.50)	5.429 (100.13)	5.207 (103.75)	5.207 (103.75)
EDUC	0.039 (40.24)	0.052 (66.00)	0.051 (65.83)	0.025 (17.34)	0.039 (32.46)	0.039 (32.54)
DEGREE	0.183 (20.55)	—	—	0.227 (15.79)	—	—
EXP	0.036 (44.68)	0.034 (41.42)	0.034 (41.38)	0.023 (12.78)	0.023 (14.49)	0.023 (14.50)
EXP^2	–0.00061 (39.70)	–0.00055 (36.19)	–0.00055 (36.21)	–0.00045 (14.74)	–.00041 (13.61)	–0.00041 (13.63)
LNWEEKS	0.918 (161.75)	0.912 (159.44)	0.911 (159.35)	0.924 (79.80)	0.926 (79.19)	0.926 (79.19)
NONMAR	–0.276 (45.91)	–0.268 (44.16)	–0.269 (44.23)	–0.188 (15.50)	–0.184 (15.03)	–0.184 (15.02)
GOVT	0.053 (6.50)	—	—	0.027 (1.41)	—	—
SELF	–0.217 (31.56)	—	—	–0.176 (14.38)	—	—
NONCMA	–0.075 (15.06)	–0.096 (19.40)	–0.095 (19.18)	–0.013 (1.22)	–0.024 (2.32)	–0.024 (2.27)
NFL	–0.042 (2.69)	–0.027 (1.72)	–0.032 (2.03)	0.156 (2.13)	0.219 (2.95)	0.219 (2.94)
MARIT	–0.109 (10.94)	–0.093 (9.32)	–0.090 (9.07)	–0.021 (0.57)	0.001 (0.03)	0.002 (0.05)
QUE	–0.010 (0.99)	–0.008 (0.99)	0.016 (2.77)	–0.070 (3.16)	–0.060 (4.44)	–0.056 (4.32)
PR	0.051 (7.39)	0.027 (3.96)	0.025 (3.65)	0.082 (6.31)	0.080 (6.12)	0.080 (6.12)
BC	0.152 (18.43)	0.139 (16.74)	0.137 (16.48)	0.097 (7.80)	0.087 (6.96)	0.087 (6.96)
FRETH	—	0.033 (4.63)	—	—	0.027 (1.08)	—
RESID	—	—	—	0.025 (13.74)	0.022 (12.86)	0.022 (12.83)
$RESID^2$	—	—	—	–0.00034 (8.07)	–0.00033 (7.84)	–0.00033 (7.81)
FRONLY(NON-QUE)	0.027 (0.67)	— —	— —	–0.001 (0.01)	— —	— —
FRONLY(QUE)	–0.016 (1.84)	—	—	–0.135 (4.43)	—	—
ENGONLY(NON-QUE)	–0.039 (4.41)	—	—	–0.059 (3.35)	—	—
ENGONLY(QUE)	–0.031 (1.20)	—	—	–0.102 (3.71)	—	—

(Continued Overleaf)

Table 15.A4 Continued

Variable	*Canadian Born*			*Foreign Born*		
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
NENF	–1.363 (11.31)	—	—	–0.187 (5.89)	—	—
EDUCA	—	—	—	–0.009 (5.54)	—	—
R^2	.3622	.3493	.3491	.3242	.3085	.3085
Sample size	80,795	80,795	80,795	22,969	22,969	22,969

Notes
a Dashes indicate "variable not entered."
b The *t* statistics are in parentheses.

Source: 1981 Census of Canada.

Table 15.A5 Regression analysis of earnings for adult males born in Canada by French ethnicity, 1981[a,b] (Dependent variable: natural logarithm of earnings)

Variable	*French Ethnic Origin*				*Non-French Ethnic Origin*	
	Canada		*Quebec*	*Non-Quebec*		
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Constant	5.578 (139.08)	5.478 (140.05)	5.513 (127.22)	5.754 (67.09)	5.344 (142.17)	5.212 (150.28)
EDUC	0.036 (25.03)	0.051 (43.42)	0.036 (22.88)	0.037 (11.22)	0.042 (31.90)	0.053 (51.08)
DEGREE	0.226 (15.82)	—	0.254 (16.33)	0.141 (4.28)	0.153 (13.43)	—
EXP	0.038 (30.63)	0.036 (29.20)	0.039 (28.92)	0.034 (11.82)	0.036 (33.57)	0.033 (30.99)
EXP^2	–0.00062 (27.28)	–0.00057 (25.08)	–0.00065 (25.96)	–0.00053 (10.14)	–0.00060 (29.83)	–0.00055 (27.41)
LNWEEKS	0.876 (105.92)	0.869 (104.26)	0.893 (97.08)	0.830 (46.02)	0.945 (124.10)	0.941 (122.60)
NONMAR	–0.301 (32.48)	–0.289 (30.94)	–0.307 (30.10)	–0.276 (13.28)	–0.263 (33.62)	–0.257 (32.58)
GOVT	0.074 (6.02)	—	0.070 (5.21)	0.085 (3.07)	0.043 (4.01)	—
SELF	–0.179 (16.04)	—	–0.175 (14.23)	–0.189 (7.58)	–0.232 (26.76)	—
NONCMA	–0.039 (5.16)	–0.065 (8.69)	–0.057 (6.85)	0.017 (0.95)	–0.094 (14.43)	–0.115 (17.56)
NFL	0.024 (0.38)	0.048 (0.76)	—	–0.007 (0.09)	–0.024 (1.40)	–0.008 (0.49)
MARIT	–0.138 (7.18)	–0.126 (6.62)	—	–0.163 (7.28)	–0.094 (7.97)	–0.079 (6.60)
QUE	0.009 (0.66)	–0.019 (1.74)	— —	— —	–0.058 (2.62)	–0.012 (0.82)

PR	0.029 (1.49)	0.001 (0.02)	—	0.032 (1.48)	0.059 (7.70)	0.035 (4.61)
BC	0.113 (4.47)	0.090 (3.54)	—	0.107 (3.73)	0.158 (17.48)	0.148 (16.17)
FRONLY(NON-QUE)	0.011 (0.28)	—	—	–0.008 (0.18)	0.074 (0.37)	—
FRONLY(QUE)	–0.040 (4.48)	—	–0.030 (3.38)	—	0.003 (0.08)	—
ENGONLY(NON-QUE)	–0.003 (0.21)	—	—	–0.001 (0.08)	–0.050 (3.48)	—
ENGONLY(QUE)	–0.008 (0.09)	—	–0.002 (0.02)	—	0.002 (0.06)	—
NENF	—	—	—	—	–1.325 (10.51)	—
R^2	.4017	.3897	.4323	.3308	.3343	.3308
Sample size	28,953	28,953	21,567	7,386	51,842	51,842

Notes
a Dashes indicate "variable not entered."
b The *t* statistics are in parentheses.

Source: 1981 Census of Canada.

Table 15.A6 Pooled regression analysis of earnings for adult males in Canada, 1981[a,b] (Dependent variable: natural logarithm of earnings)

Variable	*Canadian and Overseas Born*			
	(1)	*(2)*	*(3)*	*(4)*
Constant	5.373 (232.85)	5.361 (231.49)	5.498 (227.81)	5.422 (213.79)
EDUC	0.048 (74.19)	0.048 (74.42)	0.035 (43.87)	0.040 (41.14)
DEGREE	—	—	0.196 (26.01)	0.181 (20.41)
EXP	0.031 (43.32)	0.031 (43.33)	0.034 (47.41)	0.037 (45.34)
EXP^2	–0.00052 (38.16)	–0.00052 (38.10)	–0.00058 (42.54)	–0.00061 (40.15)
LNWEEKS	0.916 (178.36)	0.917 (178.48)	0.922 (180.89)	0.920 (180.54)
NONMAR	–0.253 (46.49)	–0.253 (46.43)	–0.260 (48.25)	–0.259 (48.01)
GOVT	—	—	0.051 (6.82)	0.050 (6.67)
SELF	—	—	–0.206 (34.40)	–0.208 (34.65)
FRETH	—	0.039 (5.92)	—	—
FRONLY(NON-QUE)	—	—	0.010 (0.26)	0.019 (0.48)

(Continued Overleaf)

Table 15.A6 Continued

Variable	*Canadian and Overseas Born*			
	(1)	*(2)*	*(3)*	*(4)*
FRONLY(QUE)	—	—	–0.028 (3.31)	–0.021 (2.43)
ENGONLY(NON-QUE)	—	—	–0.040 (5.17)	–0.043 (5.51)
ENGONLY(QUE)	—	—	–0.087 (4.85)	–0.085 (4.72)
NENF	—	—	–0.232 (8.50)	–0.238 (8.42)
FOR	–0.315 (20.63)	–0.308 (20.13)	–0.300 (19.64)	0.041 (1.20)
RESID	0.020 (11.51)	0.020 (11.62)	0.020 (11.70)	0.025 (13.38)
$RESID^2$	–0.00027 (6.58)	–0.00028 (6.71)	–0.00028 (6.71)	–0.00031 (7.37)
FOR EDUC	—	—	—	–0.015 (8.53)
FOR DEGREE	—	—	—	0.047 (2.77)
FOR EXP	—	—	—	–0.015 (7.46)
FOR EXP^2	—	—	—	0.00017 (4.96)
EDUCA	—	—	—	–0.009 (5.40)
NONCMA	–0.083 (18.65)	–0.085 (19.02)	–0.065 (14.60)	–0.063 (14.11)
NFL	–0.032 (2.09)	–0.025 (1.68)	–0.045 (2.98)	–0.039 (2.62)
MARIT	–0.092 (9.63)	–0.094 (9.87)	–0.111 (11.66)	–0.108 (11.40)
QUE	0.005 (0.87)	–0.021 (3.03)	–0.0128 (1.85)	–0.022 (2.27)
PR	0.033 (5.36)	0.035 (5.76)	0.053 (8.66)	0.054 (8.85)
BC	0.123 (17.82)	0.126 (18.11)	0.137 (19.89)	0.137 (19.97)
R^2	.3395	.3397	.3524	.3532
Sample size	103,764	103,764	103,764	103,764

Notes
a Dashes indicate "variable not entered."
b The *t* statistics are in parentheses.

Source: 1981 Census of Canada.

Table 15.A7 Means and standard deviations of variables for foreign-born[a] males in Canada by country of origin, 1981

Variable	*English-Speaking*	*U.S.*	*U.K.*	*Non-English-Speaking*
Earnings ($)	24193.	23060.	24495.	19415.
	(14258)	(15900)	(13773)	(12204)
LNY	9.892	9.775	9.923	9.652
	(0.746)	(0.854)	(0.711)	(0.794)
EDUC (years)	13.870	14.293	13.757	11.693
	(3.432)	(3.811)	(3.315)	(4.520)
EXP (years)	23.659	22.172	24.055	24.565
	(12.071)	(12.779)	(11.845)	(12.329)
DEGREE (%)	22.0	39.9	17.2	15.9
	(41.4)	(49.0)	(37.8)	(36.5)
NONMAR (%)	16.4	18.6	15.8	15.2
	(37.0)	(38.9)	(36.5)	(36.0)
WEEKS	47.803	46.520	48.145	46.279
	(9.244)	(10.695)	(8.787)	(10.480)
LNWEEKS	3.830	3.788	3.841	3.784
	(0.337)	(0.392)	(0.320)	(0.391)
GOVT (%)	8.2	5.5	9.0	4.3
	(27.5)	(22.7)	(28.6)	(20.4)
SELF (%)	12.2	15.8	11.2	15.9
	(32.7)	(36.4)	(31.5)	(36.6)
NONCMA (%)	34.0	44.1	31.3	21.9
	(47.4)	(49.7)	(46.4)	(41.4)
FRONLY(NON-QUE) (%)	0.0	0.0	0.0	0.2
	(0.0)	(0.0)	(0.0)	(3.9)
FRONLY(QUE) (%)	0.5	2.1	0.1	3.5
	(6.9)	(14.2)	(2.5)	(18.4)
ENGONLY(NON-QUE) (%)	85.3	79.1	87.0	73.7
	(35.4)	(40.7)	(33.7)	(44.0)
ENGONLY(QUE) (%)	2.8	3.6	2.6	3.8
	(16.6)	(18.7)	(16.0)	(19.2)
NENF (%)	0.0	0.0	0.0	3.7
	(0.0)	(0.0)	(0.0)	(18.8)
FRETH (%)	3.3	12.7	0.7	3.7
	(17.8)	(33.4)	(8.6)	(19.0)
RESID (years)	20.614	19.470	20.919	18.723
	(11.212)	(13.444)	(10.517)	(10.109)
NFL (%)	0.9	1.4	0.8	0.1
	(9.5)	(11.9)	(8.8)	(3.8)
MARIT (%)	3.3	6.5	2.4	0.9
	(17.7)	(24.6)	(15.3)	(9.3)

(Continued Overleaf)

Table 15.A7 Continued

Variable	*English-Speaking*	*U.S.*	*U.K.*	*Non-English-Speaking*
QUE (%)	6.1 (24.4)	12.6 (33.2)	4.7 (21.1)	17.2 (37.7)
PR (%)	14.7 (35.4)	18.1 (38.5)	13.7 (34.4)	13.7 (34.4)
BC (%)	21.3 (40.9)	24.2 (42.9)	20.5 (40.4)	14.0 (34.7)
Sample size	5,998	1,263	4,735	16,971

Note
a Standard deviations are in parentheses.

Source: 1981 Census of Canada.

Table 15.A8 Regression analysis of earnings for foreign-born adult males in Canada by country of origin, 1981[a,b] (Dependent variable: natural logarithm of earnings)

	All English-Speaking		*U.S.*	*U.K.*	*Total Non-English Speaking*
	(1)	*(2)*	*(3)*	*(4)*	*(5)*
Constant	4.981 (48.61)	4.982 (48.46)	4.696 (20.69)	5.127 (44.45)	5.342 (91.87)
EDUC	0.052 (19.26)	0.052 (19.26)	0.052 (7.96)	0.053 (17.98)	0.033 (23.03)
EXP	0.035 (11.72)	0.035 (11.47)	0.040 (5.93)	0.032 (9.59)	0.018 (9.73)
EXP^2	–0.00063 (10.82)	–0.00064 (10.51)	–0.00073 (5.31)	–0.00057 (8.83)	–0.00033 (9.72)
LNWEEKS	0.978 (41.01)	0.978 (40.92)	1.008 (20.09)	0.955 (35.16)	0.901 (67.47)
NONMAR	–0.211 (9.67)	–0.210 (9.65)	–0.159 (3.14)	–0.222 (9.27)	–0.174 (11.87)
NONCMA	–0.129 (7.46)	–0.130 (7.46)	–0.200 (4.72)	–0.102 (5.37)	0.013 (0.98)
NFL	0.213 (2.53)	0.213 (2.53)	0.031 (0.19)	0.319 (3.23)	0.214 (1.59)
ATL	–0.026 (0.57)	–0.027 (0.59)	–0.096 (1.15)	0.059 (1.04)	–0.004 (0.07)
QUE	0.073 (2.18)	0.073 (2.16)	0.103 (1.61)	0.066 (1.59)	–0.053 (3.69)
PR	0.089 (3.80)	0.089 (3.77)	0.095 (1.71)	0.095 (3.69)	0.083 (5.32)
BC	0.048 (2.31)	0.048 (2.31)	0.029 (0.57)	0.064 (2.86)	0.104 (6.68)

RESID	0.006 (6.71)	0.005 (1.74)	0.010 (5.48)	0.004 (3.73)	0.030 (13.68)
$RESID^2$	(a)	0.00001 (0.18)	(a)	(a)-	–0.00050 (9.13)
R^2	.3290	.3290	.3682	.3163	.2974
Sample size	5,998	5,998	1,263	4,735	16,971

Notes
a Dashes indicate "variable not entered."
b The *t* statistics are in parentheses.

Source: 1981 Census of Canada.

Acknowledgements

The research for this paper was financed in part by a grant from the Rockefeller Foundation for research on immigrants. Much of the research for this study was undertaken while Barry Chiswick was a visiting scholar at the Hoover Institution, Stanford University, and Paul Miller was a visiting assistant professor at the University of Western Ontario, and the financial support of these institutions is appreciated.

Notes

1 Except where noted, the data in this section are for adult men (aged 25–64) in the labor force as reported in the 1981 Census of Canada. Earnings refer to wage, salary, and self-employment income in Canadian dollars.
2 The main focus in Tomes's (1983, 1985) studies is earnings differences by religion, which is another dimension of ethnicity. He finds relatively little difference in earnings and rates of return from schooling between Protestants and Catholics, but the small sample of Jews has higher earnings. Research on American Jews finds similar patterns, that is, higher earnings and higher rates of return from schooling (Chiswick, 1983a). Religion is not explicitly analyzed in this study because of its lesser relevance for policy analysis.
3 Richmond and Kalbach (1980) analyze the earnings of males and females aged 15 years and over, while Kuch and Haessel (1979) study the earnings of males aged 15 or more years. As reported by Tomes (1983), 34.9 percent of 15- to 24-year-old males, 65.3 percent aged 65 or more years, but only 6.9 percent of 25- to 64-year-old males are excluded from the sample owing to zero earnings. The requirement of positive earnings implies, therefore, that selectivity bias may be an important consideration in the younger and older age groups. Carliner's (1981) study was limited to men between ages 18 and 64 who worked more than 34 hours for more than 26 weeks per year, were not fulltime students, and were not self-employed. These selection criteria exclude 25.2 percent of the prime-age male wage and salary work force.
4 On the basis of limited controls introduced by way of cross-tabulatation, Richmond and Kalbach (1980, p. 53) conclude that "Immigrants who have been established in Canada for five years or more were earning above-average incomes largely because of their location in the large metropolitan areas and in the economically more prosperous regions of the country. However, within these locations they were not always earning as much as might have been expected given their high level of

education and qualifications. This was particularly true of the most recently arrived immigrants."

5 Due to the absence of information on birthplace of parents in the 1981 Census file, the analysis of the earnings of second- and later-generation Canadians presented in Section VII is based on data from the 1971 Census.

6 Using these criteria, a potential sample in excess of 100,000 observations is available. The size of this data set suggests that reliable statistical estimates would be derived at far less computer cost by using a stratified random sample. A 1 in 3 random sample of the native born in the two most populous provinces (Ontario and Quebec) was selected. The entire 1 in 50 sample was used for the native born in the other eight provinces and for the foreign born in all ten provinces. All results presented are weighted so that they reflect actual population proportions. Differences between weighted and unweighed results were, however, quite minor.

7 The detailed regression equation for the analysis of the 1980 earnings of the Canadian-born adult men are reported in columns (1)–(3) of Appendix Table 15.A4 and by French ethnicity in Table 15.A5. The Table 15.A4, column (1) specification is similar to that used in many previous analyses of variations in earnings or income in the Canadian labor market (see, e.g., Tomes, 1983).

8 The total effect of education is

$$\frac{\partial \ell n Y}{\partial EDUC} = \beta_1 + \beta_2 \frac{\partial DEGREE}{\partial EDUC}$$

where β_1 and β_2 are the partial effects on income of EDUC and DEGREE. Since β_2 is positive and ∂DEGREE/∂EDUC is positive, β_1 is a downward-biased estimate of the effect of schooling on income. As the Table 15.A4, column (1) specification includes a university degree variable, the coefficient of 0.039 should not be compared to estimates presented in studies adopting more conventional specifications of the human capital earnings function which do not include a degree variable.

9 Adding the square of educational attainment to the estimating equation results in a lowering of the coefficient on DEGREE by 2 percentage points. There is some evidence that the partial effect of education increases slightly with educational level. The linear education specification yields a very good approximation to the quadratic, however, and the simpler specification is retained to yield a greater degree of comparability with past research.

10 Statistics Canada indicates a high degree of comparability of the 1981 Census with previous censuses, with one of the main changes in procedures being the coding of "other non-university education." The change in procedures resulted in a shift from university to non-university education (see Statistics Canada, 1981a, p. 19). This may have changed the effect of DEGREE.

11 Smith (1976) reports that the wage differential between public and private sector employees in the United States varies with the level of government. Federal government employees enjoy a premium in wages, the earnings of state government employees do not differ significantly from those of the privately employed, while employment at the local government level is associated with a slight earnings disadvantage. Unfortunately, the level of government employment cannot be determined in the Canadian Census. For an analysis of earnings differentials between the private and public sectors in Canada using the 1971 Census data, see Gunderson (1979a).

12 Less than 0.1 percent of native-born Canadians reported that they did not speak either French or English. The regression coefficient of –1.36 implies 74 percent lower earnings.

13 Such a trend is compatible with conjectures made by both Carliner (1981) and

Fenwick (1982) that the policies of the Parti Québecois may have an important influence on language differences in wage rates. There has been little research on the determinants of bilingualism. In a study of French Canadians living in Quebec, Grenier and Vaillancourt (1983) find that the incidence of bilingualism is greater among males, among those born in an "English-speaking" area, and among those whose parents had more schooling, an English mother tongue, and higher occupational attainment. See also Veltman et al. (1979).

14 In the 1971 Census, only the respondent's paternal ancestry was to be reported, theoretically resulting in one ethnic origin per respondent. Multiple responses were reduced to single entries during coding (see Statistics Canada, 1981b, p. 111, for details). A broader definition of the term *ethnic group* which allows for multiple ethnic origins was used in the 1981 Census. The rules which Statistics Canada implemented in 1971 cannot be imitated using the 1981 Census data. The approach adopted has been to reduce all multiple ethnic origins involving a French response to French. When multiple origins are recoded to English, the apparent earnings advantage of men of French ethnic origin in 1981 is increased by about 1 percentage point.

15 The regressions are reported in the Statistical Appendix in Table 15.A5. The F ratio for the Chow test for the specification in columns (1) and (5) is 7.82, and for columns (2) and (6) it is 8.67.

16 For a sample of married men between the ages of 16–65, Robinson and Tomes (1982b) estimated that the partial effect of schooling on the earnings of immigrants was 2.8 percentage points less than the corresponding effect among the native born. The Robinson and Tomes regression employed the hourly wage as the dependent variable and controlled for education, experience, birthplace (dichotomous variable), and language.

17 The partial effect and *t* ratios for the foreign origin and duration of residence variables in a pooled Canadian-born/foreign-born regression, controlling for schooling, labor market experience, weeks worked, marital status, size of place, and province, are as shown in the tabulation.

	(1)	*(2)*[a]	*(3)*[b]
FOR	−0.315	−0.308	−0.300
	(20.63)[c]	(20.13)	(19.64)
RESID	0.0200	0.020	0.020
	(11.51)	(11.62)	(11.70)
$RESID^2$	−0.00027	−0.00028	−0.00028
	(6.58)	(6.71)	(6.71)

Notes

a Column 2: Also controlling for French ethnicity.

b Column 3: Also controlling for language and employment type.

c *t* ratios in parentheses.

Source: 1981 Census of Canada (see Statistical Appendix Table 15.A6).

18 The birthplace regions are defined as follows: *British Isles* includes the United Kingdom and the Republic of Ireland. *Western Europe* includes Belgium, Luxembourg, France, West Germany, the Netherlands, and Austria. *Southern Europe* comprises Greece, Italy, Portugal, and Yugoslavia. *Eastern Europe* includes Hungary, Poland, the USSR, and Czechoslovakia. *Asia* includes South and East Asia. The *Remainder* group includes the European born who could not be allocated to the above regions, Africa, South and Central America, and other regions not elsewhere identified.

19 Among immigrants from the United States and the British Isles, the coefficients and *t* statistics on RESID and its square are as follows:

$$0.005 \text{ RESID} - 0.00001 \text{ RESID}^2$$
$$(1.74) \qquad (0.18)$$

Omitting the squared term from the estimating equation yields a coefficient of 0.006 on RESID with a *t* statistic of 6.71. For immigrants from countries other than the United States and the British Isles, the coefficients and t statistics on RESID and its square are as follows:

$$0.030 \text{ RESID} - 0.00050 \text{ RESID}^2$$
$$(13.68) \qquad (9.13)$$

20 In general, there are only slight differences between the regression coefficients estimated for the children of natives and the children of immigrants.

21 Tomes (1983) did not find an earnings advantage for the foreign-parentage men. However, his regressions are disaggregated by major religious group, include a dummy variable for "mother tongue" (first language learned that is still understood), and include dummy variable for ethnic or cultural background traced through the father's side. The ethnic variables consist of the following categories: English; French; Western and Central Europe; Eastern and Central Europe; Southern Europe; Russian or Ukrainian; Scandinavian or Finnish; other backgrounds. Given Canadian immigration history, these variables largely control for immigrant generation among those born in Canada.

22 For the detailed 1971 Census analyses, see Chiswick and Miller (1984). A greater degree of detail on size of place is available from the 1971 Census file, with four regions of residence being distinguished: *Metropolitan* (population > 30,000); *Major Urban* (population < 30,000); *Rural Nonfarm*; and *Rural Farm*. Tests indicate that this difference does not have a material impact on the results.

23 A similar analysis for the United States showed a declining profile for Cuban immigrants but a rising profile for some others, particularly Asian immigrants (Chiswick, 1986). As in Canada, the U.S. data show a steepening in the effects of experience on earnings for the native and foreign born from the 1970 to the 1980 Census.

References

Becker, Gary S. (1957), *The Economics of Discrimination*. Chicago: University of Chicago Press.

Carliner, Geoffrey (1981), "Wage Differences by Language Group and the Market for Language Skills in Canada." *Journal of Human Resources*. 16(3): 384–399.

Chiswick, Barry R. (1974), *Income Inequality: Regional Analyses Within a Human Capital Framework*. New York: National Bureau of Economic Research.

Chiswick, Barry R. (1977), "Sons of Immigrants: Are They at an Earnings Disadvantage?" *American Economic Review* 67(1): 376–380.

Chiswick, Barry R. (1978), "The Effect of Americanization on the Earnings of Foreign-Born Men." *Journal of Political Economy* 86(5): 897–921.

Chiswick, Barry R. (1979), "The Economic Progress of Immigrants: Some Apparently Universal Patterns." In William Fellner (ed.), *Contemporary Economic Problems, 1979*. Washington, DC: American Enterprise Institute.

Chiswick, Barry R. (1980), "The Earnings of White and Coloured Male Immigrants in Britain." *Economica* 47(February): 81–87.

Chiswick, Barry R. (1983a), "The Earnings and Human Capital of American Jews." *Journal of Human Resources* (Summer): 313–336.

Chiswick, Barry R. (1983b), "Differences in Educational Attainment Among Racial and Ethnic Groups: Patterns and Preliminary Hypotheses." Paper presented at National Academy of Education Conference on the State of Education, Chicago.

Chiswick, Barry R. (1986), "Is the New Immigration Less Skilled Than the Old?" *Journal of Labor Economics* 4(2): 168–192.

Chiswick, Barry R. and Paul W. Miller (1984), "Immigrant Earnings in Canada: 1971 Census Analysis." University of Illinois at Chicago (mimeo).

Chiswick, Barry R. and Paul W. Miller (1985), "Immigrant Generation and Income in Australia." *Economic Record* 61(173): 540–553.

Daniel, Mark (1982), "Comparative Living Costs: Who's Ahead?" *The Canadian Business Review* (Spring): 44–48.

Fenwick, Rudy (1982), "Ethnic Culture and Economic Structure: Determinants of French–English Earnings Inequality in Quebec." *Social Forces* 61(1): 1–23.

Grenier, Gilles (1985), "Bilinguisme, transferts linguistiques et revenus du travail au Québec: Quelques éléments d'interaction." Pp. 243–288 in François Vaillancourt (ed.), *Economie et langue*. Québec: Conseil de la Langue Française.

Grenier, Gilles and François Vaillancourt (1983), "An Economic Perspective on Learning a Second Language." *Journal of Multicultural and Multilingual Development* 4(6), 471–83.

Gunderson, Morley (1979a), "Earnings Differentials Between the Public and Private Sectors." *Canadian Journal of Economics* 12(2): 228–242.

Gunderson, Morley (1979b), "Decomposition of the Male/Female Earnings Differential: Canada 1970." *Canadian Journal of Economics* 12(3): 479–485.

Jasso, Guillermina and Mark R. Rosenzweig (1988), "How Well do U.S. Immigrants Do? Vintage Effects, Emigration Selectivity, and Occupational Mobility", pp. 229–253 in *Research in Population Economics*, Vol. 6, Greenwich, Connecticut: JAI Press.

Kuch, P. and W. Haessel (1979), *An Analysis of Earnings in Canada*. Ottawa: Statistics Canada.

Richmond, A. H. and W. E. Kalback (1980), *Factors in the Adjustment of Immigrants and their Descendants*. Ottawa: Statistics Canada.

Robinson, Christopher and Nigel Tomes (1982a), "Self-Selection and Interprovincial Migration in Canada." *Canadian Journal of Economics* 15(3): 474–501.

Robinson, Christopher and Nigel Tomes (1982b), "Family Labour Supply and Fertility: A Two-Regime Model." *Canadian Journal of Economics* 15(4): 706–734.

Shapiro, Daniel M. and Morton Stelcner (1981), "Male–Female Earnings Differentials and the Role of Language in Canada: Ontario and Quebec, 1970." *Canadian Journal of Economics* 14(2): 341–348.

Smith, Sharon P. (1976), "Government Wage Differentials by Sex." *Journal of Human Resources* 11(2): 185–199.

Statistics Canada (1981a), "School Attendance and Level of Schooling; 1981 Census of Canada." Catalogue No. 92–914.

Statistics Canada (1981b), "1981 Census Dictionary." Catalogue No. 99–101.

Tandon, Bankey B. (1977), "An Empirical Analysis of Earnings of Foreign-Born and Native-Born Canadians." Unpublished Ph.D. thesis, Queen's University at Kingston (Ontario).

Tandon, Bankey B. (1978), "Earnings Differentials Among Native-Born and Foreign-Born Residents of Toronto." *International Migration Review* 12(3): 405–410.

Tomes, Nigel (1983), "Religion and the Rate of Return on Human Capital: Evidence for Canada." *Canadian Journal of Economics* 16(1): 122–138.

Tomes, Nigel (1985), "Religion and the Earnings Function." *American Economic Review* 75(2): 245–250.

Vaillancourt, François and Robert Lacroix (1985), "Revenus et langue au Québec, 1970–1980: Une revue des écrits." Pp. 221–242 in François Vaillancourt (ed.), *Economie et langue*. Québec: Conseil de la Langue Française.

Veltman, Calvin, Jac-André Boulet, and Charles Castonguay (1979), "The Economic Context of Bilingualism and Language Transfer in the Montreal Metropolitan Area." *Canadian Journal of Economics* 13(2): 468–479.

16 Indigenous language skills and the labor market in a developing economy: Bolivia

With Harry A. Patrinos and Michael E. Hurst

I. Introduction

A little understood but widespread feature of many developing countries is the existence of population groups whose members do not speak the dominant language (or languages) of the country, that is, the language that predominates in the modern or formal sector of the economy. This characteristic exists in each of the major regions of developing economies.[1] In Latin America, for example, indigenous or native languages are spoken by segments of the population in nearly all of the countries, although the languages brought over by the European conquest, in particular Spanish and Portuguese, dominate political and economic life.[2] In Africa, the designation of country boundaries by the European colonial powers, with little or no regard for ethnic or language groups, combined with the migration of indigenous ethnic groups, has resulted in linguistic heterogeneity within individual countries. In Asia as well, linguistic heterogeneity in large countries (e.g., India and Indonesia), as well as in small ones (e.g., Laos, Papua New Guinea), has emerged as a frequently perplexing issue.

This linguistic heterogeneity has often been hidden. Government authorities frequently focused on promoting linguistic homogeneity, or pretending that it exists, as a way of creating a sense of national unity. Outside observers generally do not venture from the modern sector in developing countries' major cities, thereby missing exposure to the degree of linguistic heterogeneity. In the past, linguistic minorities have not forcefully expressed their linguistic and cultural uniqueness, although recent events suggest that this is changing.

In recent years there has been a growing interest in the role of indigenous peoples and linguistic heterogeneity, as the new nation states feel more secure, democratic principles and institutions are more firmly established, and the development process is spread from selected centers to the broader population. The research in this area, however, remains quite thin in large part because of the scarcity of data on language use. The purpose of this chapter is to expand the research base on linguistic minorities in developing countries through an analysis of data from a 1993 household survey conducted in Bolivia.

Section II provides a brief review of the literature on the labor-market implications of limited dominant language proficiency in developing economies, focusing on the research in Latin America. The data set analyzed for this study is described in Section III. Section IV presents the analyses of the determinants of dominant language skills, and the determinants of labor-force participation and labor-market earnings among men and women, with an emphasis on the effects of language skills on these labor-market outcomes. The chapter closes with a summary and conclusion.

II. Review of the literature

There has been remarkably little systematic research on the skills and economic attainment of indigenous peoples, that is, the modern-day descendants of the original populations of an area that has also been settled by later arrivals. For example, in spite of the extensive economics research conducted in the United States on African Americans, Hispanic Americans, and immigrants, and in spite of the extensive studies of immigrants in other countries of overseas settlement, such as Canada, Australia, and New Zealand, the descendants of the original populations of these countries have received slight attention. Whether called Native Americans, First Nation, Aboriginals, or Maoris, these groups now constitute numerically very small segments of the population (except in New Zealand) and suffer disadvantages of geographic isolation, low levels of education, linguistic difficulties, and low earnings compared to the majority, primarily white, population.[3]

While a few studies have examined these issues in North America and Oceania, even less research has been carried out for indigenous and non-indigenous differences in Latin America. One exception is an analysis by Jonathan Kelley of the "cost of being Indian" in rural Bolivia.[4] Using a 1966 survey of about 1,000 male household heads in rural Bolivia, Kelley decomposed the earnings differential between indigenous and nonindigenous men into components based on education, occupation, and income. With background data on the respondents' education and occupations, and the respondents' fathers' education and occupations, Kelley concluded that nearly all (between 95% and 100%) of the overall differentials were due to class components (family background, education, and occupation) rather than ethnic differences per se. In other words, equalizing human capital and family backgrounds of individuals would result in the virtual elimination of ethnic inequalities. Kelly suggests that the main reason for the change from ethnic to class inequality in Bolivia was the 1952 revolution, which resulted in a considerable increase in the power and opportunities available to Bolivia's indigenous population.

The indigenous and nonindigenous education and earnings differences in Bolivia during the late 1980s were analyzed by George Psacharopoulos, using the 1989 national (urban) household survey that included about 10,000 employed individuals.[5] Indigenous workers were found to receive much lower rewards for schooling and labor-market experience, although data on family background were not available. Younger cohorts were found to be more educated and earned more in the labor market. The overall earnings differential was not decomposed into individual characteristics components.

Much of the overall differential in earnings between monolingual Spanish speakers and monolingual Guarani speakers in Paraguay is explained by human capital differences.[6] Thus, in Paraguay, narrowing differences in educational attainment would go a long way toward equalizing labor-market outcomes. In Peru, however, a large portion of the indigenous-nonindigenous wage gap is unexplained by human capital and other observable differences.[7] A study of education and earnings in Guatemala finds that the earnings of Spanish speakers are higher than any of the indigenous language groups, both overall and when other variables are the same.[8]

III. The survey data

The data used in this study are from the Encuesta Integrada de Hogares (Integrated Household Survey), a 1993 household survey of Bolivia conducted by the Instituto Nacional de Estadistica (National Institute of Statistics). The survey was conducted in the capital cities of the nine "departments" (states) of Bolivia, including the national capital, La Paz. The respondents of particular interest are those age 15 and older.

The survey includes the question: "What languages do you usually speak?" in answer to which the respondent could report more than one language. The six language categories were Spanish, three indigenous languages (Quechua, Aymara, and Guarani), "other native languages," and "other non-native languages." For the purpose of this study, these were reclassified into three language groups: "Spanish only" (including those who spoke Spanish and a nonnative language), "indigenous only" (those who spoke only one or more native languages), and "bilingual" (those who spoke Spanish and one or more native languages). These language variables serve as the dependent variable in the language analysis and as explanatory variables in the analysis of labor-force participation and earnings.

There are no questions in the survey on the respondent's ethnicity, that is, on whether the respondent is of indigenous origin. There is, however, a high degree of correspondence in Latin America between being of an indigenous origin and speaking an indigenous language. Usually few nonindigenous individuals speak an indigenous language, although there are individuals of indigenous origins who do not speak their ancestral language. Most household surveys use the language question to identify indigenous people.[9]

The other variables used in the analyses are straightforward. The variable "Age" is measured in years. The gender variable is "Male," which equals unity for males and is zero for females. Data were collected on the number of live births for women 13–49 years of age. The variable "Number of live births" is the number of live births for women age 15–49 years and is zero for men and for women over age 49. A dichotomous variable "Age greater than 49" is created that equals unity for individuals who were not asked the live births question, that is, women older than 49 and men. Controlling for sex and age, the number of live births variable reflects the effect of fertility of women ages 15–49 on the dependent variable.

Since this is a survey conducted only in department capital cities, and many of the respondents, in particular indigenous language speakers, are from rural areas, several geographic variables are created. "La Paz" is unity for those who live in the Bolivian capital city and is zero otherwise. "Rural birthplace" is unity for those born in a rural area, otherwise it is zero. In rural areas about 90% of the population is of indigenous origin. Three internal migration variables refer to the time when the respondent moved to the present city of residence: "Migrated—over 5 years ago" is unity for those who moved more than 5 years ago, "Migrated—1–5 years ago" is unity for those who moved more than 1 but less than 5 years ago, and "Migrated—less than 1 year ago" is unity for those who moved to their present city less than 1 year before the interview was conducted. The benchmark of these region variables is a person living in the department capital, other than La Paz, in which the person was born.

The analyses of earnings and labor supply include additional variables. Among the variables is years of schooling completed, obtained by converting the categorical schooling data into a continuous years measure. The variable for labor-market experience is the number of years since age 15 that the respondent has not been in school (i.e, experience = age −15 or experience = age–schooling–6, whichever is smaller). This measure of experience is based on the assumption that years of labor-market activity prior to age 15 are not relevant for skill formation in the adult labor market. "Second job" is unity if the respondent has one or more jobs in addition to his or her primary employment. This also serves as a proxy for working more hours in a week. "Self-employed" is unity if the person is self-employed.

The earnings variable is monthly earnings from the primary job, second jobs, and self-employment, if any. Since the dependent variable in the earnings analysis is in logarithmic form, only individuals with positive earnings are included in the analysis. A measure of the total income of all household members other than the earnings of the respondent is also included in the analysis of labor supply.

The means and standard deviations of the variables are reported in Table 16.1.

IV. Analysis of language skills and labor-market activities

Language skills

The model adopted for the analysis of dominant language skills among Bolivia's population has been developed and tested for immigrant linguistic

Table 16.1 Means and standard deviations of variables, overall and by sex, urban Bolivia, 1993

Variable	*Pooled*		*Males*		*Females*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Age	35.0	16.0	35.1	15.8	34.9	16.2
Age squared	1,482.6	1,373.2	1,480.7	1,336.9	1,484.3	1,403.8
Migrated—more than 5 years ago	.454	.498	.457	.498	.451	.498
Migrated—1–5 years ago	.069	.253	.067	.251	.070	.255
Migrated—less than 1 year ago	.018	.132	.017	.128	.019	.136
Number of live births	1.09	2.05	N.A.	N.A.	3.40	2.29
Older than 49	.188	.390	.186	.389	.189	.391
La Paz	.234	.424	.227	.419	.241	.428
Rural birth place	.258	.437	.261	.439	.255	.436
Labor force participation rate	.529	.499	.659	.474	.417	.493
Schooling	9.68	4.70	10.17	4.44	9.22	4.89
Experience	17.87	16.37	17.69	16.01	18.03	16.67
Experience squared	587.39	888.19	569.15	845.91	603.14	922.89
Single	.361	.480	.375	.484	.349	.477
Widowed, divorced, separated	.094	.292	.043	.203	.138	.345
Married	.545	.498	.582	.493	.513	.500
Monolingual Spanish	.536	.499	.542	.498	.531	.499
Monolingual indigenous	.016	.126	.005	.069	.026	.159
Bilingual	.448	.497	.453	.498	.443	.497
Yearly earnings	382.12	791.21	576.93	971.11	213.96	539.80
Yearly earnings, with positive earnings	722.23	968.30	875.31	1,081.52	513.28	738.56
Ln yearly earnings, with positive earnings	6.14	.91	6.38	.84	5.81	.88
Second job	.032	.176	.041	.199	.024	.154
Self-employed	.171	.377	.159	.366	.182	.385
Household income (nonearner)	1,116.97	1,687.10	896.78	1,279.72	1,307.04	1,952.45
Sample size	12,480		5,782		6,698	

Note: Variables are defined in the text; adults are age 15 and older. Means in Table 16.1 are from the sample used to compute the language estimates in Tables 16.2 and 16.3. Numbers of live births in the "pooled" column is for all observations. The number of live births in the female column is for women who are younger than 49 years. N.A. = Not applicable.

Source: Encuesta Integrada de Hogares (La Paz: Instituto Nacional de Estadistica, 1993).

minorities in several developed countries.[10] The model is based on three conceptual variables—exposure to the dominant language, efficiency in the acquisition of dominant language skills, and economic incentives to acquire these skills. It is hypothesized that in Bolivia Spanish language proficiency would be less among those who have less exposure to Spanish, who are less efficient in learning language skills, and who have less of an economic incentive to acquire Spanish-language proficiency.

Exposure would be less among those born in rural areas that are populated almost exclusively by people who speak the indigenous languages. It would also be less among those who have more recently migrated to a department capital city.

Women have lower labor-force participation rates than do men, and participation for women is more episodic than it is for men. In addition, women in the labor market are less likely to be in the modern, formal sector. It is therefore expected that women would have less of an economic incentive to learn Spanish, the language in the dominant, modern labor-market economy. Other things being equal among women, those with more children are expected to have lower labor-force participation rates and hence a weaker incentive to learn Spanish.

Those who were older are less likely to have attended school and are less likely to be literate in any language. Moreover, for the same duration in the city older migrants would have come at an older age, and the efficiency in new language skill acquisition declines with age.[11] Thus, age would be associated with a lesser use of Spanish.

The data on language skills in the survey refer only to the languages usually spoken. The questionnaire provides for six language categories, but respondents may indicate using more than one language. There are no data on levels of proficiency in the languages identified or on proficiency in other languages that are not usually spoken. The data are, therefore, qualitative (categorical). Proficiency in Spanish presumably increases with the level of schooling since this is the language of instruction in schools.[12]

To make the analysis more manageable, we divided the information on languages usually spoken into three mutually exclusive categories: Spanish only (including those who speak a non-native language), indigenous language(s) only, and bilingual (Spanish and at least one indigenous language). Multinomial logit analysis is the preferred statistical technique.

Table 16.2 reports the multinomial logit analysis coefficients and *t*-ratios. Because of the difficulty of interpreting multinomial logit coefficients, Table 16.3 reports separately by gender the effects of changes in the values of the explanatory variables on the probability of being in each of the three language categories. The reference person used for the base group is an individual with the mean age (35 years) and for a female with the mean number of live births (3.4 live births), residing in the place of birth which is a department capital other than La Paz. For this reference group, 70% of the men speak Spanish only, 30% are bilingual, and 0.1% are monolingual native

Table 16.2 Multinomial logit estimates of language usage, overall and by sex, urban Bolivia, 1993

	Female			*Pooled*			*Male*		
	Spanish Only vs. Indigenous	*Spanish Only vs. Bilingual*	*Bilingual vs. Indigenous*	*Spanish Only vs. Indigenous*	*Spanish Only vs. Bilingual*	*Bilingual vs. Indigenous*	*Spanish Only vs. Indigenous*	*Spanish Only vs. Bilingual*	*Bilingual vs. Indigenous*
Age	−.1917	−.0952	−.0965	−.1713	−.0290	−.1422	−.1570	−.1150	−.0420
	(−4.50)	(−13.64)	(−2.27)	(−3.47)	(−2.30)	(−2.91)	(−2.34)	(−12.81)	(−.63)
Age squared	.0010	.0008	.0002	.0009	.0002	.0006	.0006	.0009	−.0004
	(3.06)	(9.17)	(.76)	(2.31)	(1.89)	(1.71)	(1.00)	(8.97)	(−.66)
Male	1.5970	−.1835	1.7804	*	*	*	*	*	*
	(7.06)	(−3.11)	(8.01)						
Migrated—more than 5 years ago	−.0527	−.2854	.2327	−.0763	−.1897	.1134	.3334	−.3479	.6814
	(−.18)	(−5.42)	(.80)	(−.24)	(−2.49)	(.36)	(.43)	(−4.76)	(.88)
Migrated—1–5 years ago	−1.0726	−.5495	−.5232	−1.2524	−.4059	−.8465	.5191	−.6133	1.1324
	(−2.82)	(−5.79)	(−1.39)	(−3.03)	(−2.80)	(−2.10)	(.41)	(−4.87)	(.90)
Migrated—Less than 1 year ago	−1.6350	−.4139	−1.2211	−1.0833	.1150	−1.1983	−1.9147	−.7353	−1.1794
	(−3.18)	(−2.29)	(−2.44)	(−1.70)	(.40)	(−1.92)	*	*	*
Number of live births	−.2091	−.0915	−.1176	−.2652	−.1637	−.1016			
	(−4.23)	(−5.75)	(−2.46)	(−4.59)	(−8.06)	(−1.82)	*	*	*
Older than 49	−.8068	−.1145	−.6924	−1.4014	−.8730	−.5284			
	(−1.83)	(−1.12)	(−1.59)	(−2.63)	(−5.16)	(−1.01)	(−1.86)	(−3.14)	(−1.16)
La Paz	.4939	−.4854	.9794	.3941	−.6136	1.0077	.7086	−.4063	1.1149
	(2.16)	(−9.25)	(4.35)	(1.56)	(−7.86)	(−4.10)	(1.19)	(−5.66)	(1.88)
Rural birth place	−2.8264	−1.2020	−1.6244	−2.7215	−1.1004	−1.6211	−2.9259	−1.2564	−1.6694
	(−12.68)	(−20.37)	(−7.40)	(−11.27)	(−12.68)	(−6.89)	(−4.64)	(−15.57)	(−2.66)
Constant	11.3371	3.2180	8.1190	11.1207	1.7907	9.3300	11.5353	3.4828	8.0525
	(9.71)	(21.83)	(6.96)	(8.29)	(6.77)	(7.00)	(5.85)	(19.98)	(4.09)

Note: Asymptotic *t*-ratios in parentheses; adults ages 15 and older.
* Variable not entered.

Source: Encuesta Integrada de Hogares (La Paz: Instituto Nacional de Estadistica, 1993).

Table 16.3 Estimated probabilities of language usage, by sex, urban Bolivia, 1993

Variable	*Male*			*Female*		
	Spanish	*Bilingual*	*Indigenous*	*Spanish*	*Bilingual*	*Indigenous*
Reference person	.699	.300	.0007	.638	.359	.0026
Age = 25	.807	.193	.0002	.672	.327	.0008
Age = 45	.608	.391	.0019	.613	.380	.0068
Migrated—more than 5 years ago	.621	.378	.0005	.594	.404	.0026
Migrated—1–5 years ago	.557	.442	.0003	.538	.454	.0075
Migrated—less than 1 year ago	.525	.471	.0037	.661	.331	.0078
Number of live births + 1	*	*	*	.599	.397	.0031
Age older than 49	*	*	*	.423	.570	.0069
La Paz	.608	.392	.0003	.490	.509	.0013
Rural birth place	.395	.597	.0076	.363	.615	.0221

Note: The reference person is a 35-year-old adult who was born in the current department capital city of residence and who did not live in La Paz in 1993. The reference woman also had 3.4 live births, which is the average for women younger than age 49.
* Variable not entered.

Source: Means from Table 16.1, coefficients from Table 16.2.

language speakers; for women the proportions are 64%, 36%, and 0.3%, respectively.

In the pooled analyses, as well as in the analyses done separately by gender, age is associated with language usage. Older persons are less likely to be monolingual Spanish speakers and more likely to speak an indigenous language. Among indigenous language speakers, older women are associated with speaking only the indigenous language. Among men there is no statistically significant effect of age on speaking Spanish.

Compared to the base or reference person, an additional 10 years of age lowers the probability of being a monolingual Spanish speaker from 70% to 61% for men, and from 64% to 61% for women. Those 10 years increase for men the probability of being bilingual from 30% to 39%, and from being solely an indigenous language speaker from 0.1% to 0.2%. For women the respective proportions increase from 36% to 38%, and from 0.3% to 0.7%.

Among women, the larger the number of live births, the less likely women are to speak Spanish, relative to being bilingual or relative to speaking only an indigenous language. For the reference woman an additional live birth beyond the mean level lowers the probability of speaking only Spanish from 64% to 60% and raises the proportion of women who are bilingual from 36% to 40%.

Place of birth and years since moving to the current department capital city also matter, partly because the rural population is largely indigenous. Those born in rural areas are less likely to be monolingual Spanish speakers, and among those who can speak an indigenous language a larger proportion speaks only an indigenous language. While for the reference person (as defined in Table 16.3) born in the place of current residence 70% of the men speak only Spanish, and only 40% of those who were born in rural areas and who now live in the department capitals speak only Spanish. Among men born in rural areas, 60% are bilingual and 0.8% speak only an indigenous language. Among women born in rural areas, 36% speak only Spanish, 62% are bilingual, and 2% speak only an indigenous language.

There is a lower level of Spanish usage and a higher rate of bilingualism among those who migrate from urban areas to the department capital cities than among those who were born there. The use of Spanish tends to increase the longer the duration of residence in the department capital. An exception are women who migrated less than 1 year prior to the survey. They are more likely to be monolingual Spanish speakers and less likely to be bilingual than men or women born in the city of residence.

In the Bolivian capital, La Paz, Spanish monolingualism is less common. In La Paz, only 61% of the men are monolingual Spanish speakers, compared with 70% of the men in other department capital cities. Among women, 49% in La Paz are monolingual Spanish speakers, compared with 64% in other department capital cities. Spanish–indigenous language bilingualism is more common in La Paz for both men and women.

Thus, Spanish language usage appears to be greater among younger persons, those living in the department capital city in which they were born (other than La Paz), males, and females with fewer children. Those born in rural areas have very low levels of Spanish language skills.

Labor-force participation

In this section, we examine the labor-force participation behavior of men and women age 15 and older. Those with earnings are defined as being labor-force participants, and those without earnings (zero earnings) are defined as non-participants. With the dichotomous dependent variable, logit analysis is the statistical technique employed. The explanatory variables are as defined above.

Labor-market participation is expected to increase the higher the potential labor-market earnings are and is expected to decrease with greater home-sector productivity. Hence, labor-market participation is expected to increase with schooling level and with years of potential labor-market experience, until labor-market participation declines with the reduction in productivity associated with aging. Because of the income effect, it is expected that participation would be lower the greater the income is of other members of the household.

Because of a division of labor in the household associated with marriage, married women are expected to be less likely to work for earnings than never married (single) women and those who are widowed, divorced, or separated. Married men, on the other hand, are expected to be more likely to work than men in other marital statuses.

Table 16.4 reports separately for men and women the logit equations for whether the respondent had positive earnings. Among women the likelihood

Table 16.4 Analysis of labor force participation, by sex, urban Bolivia, 1993

Variable	*Males*		*Females*	
	Coefficients (t-ratios)	*Marginal effects*	*Coefficients (t-ratios)*	*Marginal effects*
Schooling	−.0325 (−3.556)	−.0066	.0598 (7.438)	.0149
Experience	.1591 (16.368)	.0324	.099 (9.612)	.0247
Experience squared	−.00354 (−20.114)	−.0007	−.00201 (−10.361)	−.0005
Single	−1.5221 (−13.713)	−.3102	1.2661 (9.463)	.3164
Widowed, divorced, separated	−.9638 (−5.623)	−.1964	.9549 (9.307)	.2386
Migrated—more than 5 years ago	−.0281 (−.310)	−.0057	.0178 (.217)	.0044
Migrated—1–5 years ago	.053 (.353)	.0108	−.0657 (−.442)	−.0164
Migrated—Less than 1 year ago	−.1602 (−.589)	−.0327	−.6375 (−2.026)	−.1593
La Paz	.1229 (1.453)	.0250	.2581 (3.084)	.0645
Monolingual Spanish	−.0396 (−.498)	−.0081	−.264 (−3.525)	−.0660
Monolingual indigenous	1.2425 (2.023)	.2532	.4379 (2.191)	.1094
Older than 49	*	*	−.3168 (−1.648)	−.0792
Number of live births	*	*	.0306 (1.385)	.0076
Rural birth place	.0421 (.407)	.0086	.0769 (.814)	.0192
Household income	−.00019 (−6.552)	−.00004	−.00005 (−2.174)	−.00001
Constant	1.2217 (6.772)	.2490	−1.6626 (−10.213)	−.4155
Chi square	1,942.2		457.2	
Sample size	5,574		3,941	

Note: Dependent variable: unity for persons with positive earnings, zero otherwise. Asymptotic *t*-ratios are in parentheses; adults are ages 15 and older. The marginal effect for an independent (regressor) variable is the percentage point change in the labor-force participation rate attributed to a 1-unit change in the value of the variable, or to a change from 0 to 1 for binary variables.
* Variable not entered.

Source: Encuesta Integrada de Hogares (La Paz: Instituto Nacional de Estadistica, 1993).

of working increases with the schooling level. Among men, however, schooling has a significant negative effect ($t = -3.5$), although the marginal effect for men is very small (less than 1% per year of schooling). For men, the negative effect of higher levels of schooling on working does not appear to be strongly related to current school enrollment. If a school enrollment variable is included in the equation (not shown here) the magnitude of the negative effect of schooling diminishes but remains negative and significant ($t = -2.2$). When the analysis is computed for men not currently enrolled in school, the magnitude of the negative schooling effect is similar to the effect when enrollment is not held constant.

The marital status effects are in accord with expectations. That is, married women are less likely and married men are more likely to work for earnings than are men and women in other marital statuses.

The income effect is shown in the equations for both men and women. The higher the household income from sources other than the respondent's earnings, the less likely is the respondent to participate in the labor market.

The effect of geography on participation rates varies by gender. Among men, geography does not matter. The participation rate does not vary significantly by whether the man was born in a rural area, when he moved to the current department capital of residence, or whether he lives in La Paz.

Among women, however, geography does matter. Labor-market participation is about 6 percentage points higher in La Paz than in the other cities ($t = 3.1$), perhaps reflecting the higher earnings women receive in La Paz and the greater employment opportunities for women in government service occupations in the capital city. Women who are relative newcomers to the department capital are less likely to work. The participation rates are about 16% lower among women who migrated to the city less than 1 year before the survey. This may arise from their being "tied movers." That is, they moved to accompany a husband or father rather than on behalf of their own employment opportunities. Studies in developed countries have found that female tied movers have more difficulty finding employment than women who did not move or were unattached movers.[13]

The effect of language skills on working also varies by gender. Among men, those who speak only an indigenous language are more likely to be working than are bilingual or Spanish-only speakers. This may reflect unmeasured dimensions of low family income and large family size among indigenous language speakers. Among women, Spanish-only speakers work about 7 percentage points less ($t = -3.5$) than bilingual speakers, and monolingual indigenous language speakers work about 11 percentage points more ($t = 2.2$) than bilingual speakers.

Earnings

Analysis of the earnings of the adult respondents in the department capitals in Bolivia is based on the human capital earnings function, now a standard statistical technique.[14] In this procedure, the natural logarithm of earnings is regressed as a linear function of years of schooling, years of potential labor-market experience and its square, and a set of variables describing other relevant demographic and human capital characteristics of the individual. For this study, these characteristics include dichotomous variables for marital status (single, widowed, divorced, separated), a second job, self-employed, rural birthplace, time of moving to the current city of residence (migrated more than 5 years ago, within 1–5 years, less than 1 year ago), and current residence in La Paz. The language variables indicate whether the respondent is a monolingual Spanish speaker or speaks only indigenous languages, with bilingual speakers as the benchmark.

For women, two additional variables are included in the analysis to reflect the interruption in work experience that would accompany raising children. These variables are the number of live births for women ages 15–49 (zero for those over age 49), and a dichotomous variable for a woman over age 49 for whom data on the number of live births are not available.

In Table 16.5 the regression analysis of the natural logarithm of earnings is reported separately for men and women who show positive earnings. Earnings increase significantly with both schooling and years of potential labor-market experience for both men and women. The increase is 6.5% per year of schooling for men and 6.7% for women, but the difference between the genders is not statistically significant.

Men have a somewhat steeper experience-earnings profile. When evaluated at 10 years of potential labor-market experience, earnings increase by 3.3% per year for men and by 2.8% per year for women. This difference by gender may reflect greater investments in on-the-job training by men or greater measurement error for women because of the weaker correspondence between the proxy measure of potential labor-market experience and women's actual experience arising from labor-force withdrawal because of children and other home production activities.

Marital status has a different effect for men and women. Among men, those who never married earn about 25% less (coefficient –0.28) than those currently married, while widowed, divorced, and separated men earn about 13% less (coefficient –0.12) than their married counterparts. Among women, ceteris paribus, those who never married earn about 19% (coefficient –0.21) less than married women with earnings, but widowed, divorced, and separated women earn about 10% more (coefficient 0.10) than married women.

Among women 15–49 years old, earnings are lower by 2.6% per live birth. In Bolivia, for the average number of live births per woman (about 3.4), earnings are lower by about 9% compared to childless married women. The negative effect of children on earnings may reflect less time currently in the labor market the larger the number of children women have had in the past. Among older women for whom data on the number of live births are not available, earnings are lower by about 17% compared to younger women (ages 15–49 years) without children for whom live birth data are available.

Table 16.5 Regression analysis of earnings, by sex, urban Bolivia, 1993

Variable	*Males*	*Females*
Schooling	.0633	.0650
	(23.302)	(15.500)
Experience	.0488	.0403
	(14.685)	(7.431)
Experience squared	−.0008	−.0006
	(−12.611)	(−5.371)
Single	−.2835	−.2098
	(−7.847)	(−3.455)
Widowed, divorced, separated	−.1249	.0976
	(−1.987)	(2.071)
Migrated—more than 5 years ago	.1005	.1080
	(3.476)	(2.513)
Migrated—1–5 years ago	.0102	.1434
	(.208)	(1.774)
Migrated—less than 1 year ago	.0862	−.0929
	(.934)	(−.495)
La Paz	.0431	.1516
	(1.550)	(3.559)
Second job	.4807	.2935
	(10.207)	(4.109)
Self-employed	.4777	.3338
	(17.413)	(7.772)
Monolingual Spanish	.2067	.2487
	(8.314)	(6.265)
Monolingual indigenous	.0504	−.2951
	(.296)	(−2.693)
Older than 49	*	−.1892
		(−1.936)
Number of live births	*	−.0257
		(−2.326)
Rural birth place	−.2150	−.2320
	(−7.006)	(−4.617)
Constant	5.0491	4.6074
	(86.508)	(48.276)
Adjusted R-square	.351	.246
Sample size	3,674	1,800

Note: Dependent variable: natural log of earnings for persons with positive earnings. t-ratios are in parentheses; adults are ages 15 and older.
* Variables not entered.

Source: Encuesta Integrada de Hogares (La Paz: Instituto Nacional de Estadistica, 1993).

Geographic origins and mobility matter. Among both men and women, those born in rural areas earn about 20% less than their urban-born counterparts. This may be a reflection of discrimination against those from rural origins or an indication of lower human capital because of the lower quality of schooling in rural areas and job training that is less relevant for the urban sector.

The longer an individual has lived in the department capital, the higher are the earnings. Among both men and women, there is no significant earnings difference between those born in urban areas who have lived in the current city of residence less than 1 year and those born in the city of residence. Among those who migrated 1–5 years ago there is no significant effect for men and a marginally significant positive effect for women. However, those who migrated more than 5 years ago earn about 10% more than those born and raised in the city, and the difference is highly statistically significant. This is consistent with the hypothesis that migrants tend to be favorably self-selected and that the full effects of this selectivity are muted in the first few years because they have less knowledge of the labor market in the destination and less firm-specific and city-specific job training. Once these handicaps have been overcome, the favorable selectivity of migrants tends to show higher earnings. This is similar to the patterns observed for immigrants in the United States and elsewhere.[15]

For men, the earnings in La Paz do not differ significantly from those in the other department capitals, but they are significantly higher (coefficient 0.15 or about 16% higher earnings) for women. Probably this reflects the greater employment opportunities in the female-intensive government sector in the nation's capital and the higher pay schedule for these jobs.

Earnings differ significantly by language skills. Among men, other variables being equal, Spanish speakers earn about 23% (coefficient 0.21) more than bilingual speakers. There is no significant difference in earnings between bilingual speakers and those who speak only an indigenous language, perhaps because so few men speak only an indigenous language.

The differences in earnings by language skills are greater for women. Monolingual Spanish speakers earn about 28% (coefficient 0.25) more than bilingual speakers, who earn about 25% (coefficient −.030) more than women who speak only an indigenous language.

These differences in earnings by language skills reflect the value of speaking Spanish in the department capitals' labor markets. The lower earnings of the bilingual speakers also reflect their poor Spanish language skills, the lower quality of schooling received by indigenous peoples, or discrimination against indigenous language speakers in the modern (as distinct from traditional) labor market.

V. Summary and conclusions

In this article, we use the Encuesta Integrada de Hogares (Integrated Household Survey) that was conducted in 1993 in the department capital cities in Bolivia to analyze the determinants of Spanish and indigenous language usage in Bolivia, and the effects of language usage on labor-force participation and earnings. Three language groups are considered: those who speak Spanish but not an indigenous language, those who speak an indigenous language but do not speak Spanish, and bilingual Spanish–indigenous language speakers. The analysis is conducted for men and women age 15 and older.

Language patterns in Bolivia are found to be consistent with the model of language proficiency based on exposure, efficiency, and economic incentives. Spanish is more likely to be the only language used by those who are more active in the labor market (men and women with fewer children), those with more exposure to Spanish as distinct from indigenous languages (born in an urban area or having resided in the city for a longer duration), and those more efficient in acquiring language skills (moved to the city at a younger age). Surprisingly, Spanish monolingualism is less common in La Paz than in the other department capital cities, ceteris paribus. Indigenous language monolingualism is most common among older women, those born in rural areas, those with several children, and those who have recently arrived in the department capital city, especially if it is La Paz.

The determinants of the propensity to work for wages differ between men and women. As expected, women not currently married are more likely to work than currently married women, but there is no significant effect of the number of children on work propensity. Among men, however, marriage is associated with greater labor-force participation. Schooling has a positive effect on labor-market participation for women, but a surprising negative effect partly related to school attendance for men, although the magnitude of the latter effect is small. Women who have lived in the city less than a year are less likely to work, perhaps because they are recent tied movers. Among both men and women, greater household income from sources other than their own labor supply has a negative effect on labor-market participation.

Labor-force participation varies by language usage. Among men the relatively few who speak only an indigenous language have a higher participation rate. Among women, monolingual indigenous language speakers have a higher participation rate than the bilingual speakers, but monolingual Spanish speakers have a lower propensity to work, other variables being the same.

The analysis indicates that earnings increase with human capital among both men and women. Earnings rise with years of schooling (implying about a 6.5% rate of return for both genders), potential labor-market experience, being born in an urban area, and among migrants with a longer duration in the city. Migrants who have resided 5 or more years in the city have about 10% higher earnings than nonmigrants born there, other variables being the same, suggesting favorable selectivity in migration.

Earnings in La Paz are about 16% higher for women than they are in the other department capital cities, but there is no similar effect among men. The greater earnings and labor-force participation of women in La Paz may be due to the greater female intensity of employment in the central government sector of the economy.

Language skills are important. Monolingual Spanish speakers earn about 25% more than those who speak both Spanish and an indigenous language, while women who speak only an indigenous language earn about 25% less than the bilingual speakers. Bilingual speakers may be penalized in the labor market because of a poorer proficiency in Spanish.

The analysis indicates that indigenous and modern-sector language skills can be modeled successfully for a developing economy and that these language skills have an impact on labor market participation and on earnings. It also suggests that there may be large benefits from programs designed to improve Spanish language proficiency, for example, through bilingual education, among people of indigenous origins.

Acknowledgements

This chapter is an outgrowth of Barry Chiswick's Visiting Scholar appointment in the Human Development Network at the World Bank, and was completed while he was John M. Olin Visiting Professor, Center for the Study of the Economy and State, Graduate School of Business, University of Chicago. The support of these institutions is appreciated. The views expressed in this chapter are solely those of the authors and do not necessarily express the views of the sponsoring organizations.

Notes

1 See, e.g., the detailed statistics on language usage around the world in Joseph E. Grimes and Barbara F. Grimes, *Ethnologue: Languages of the World*, 13th ed. (Dallas: Summer Institute of Linguistics, 1993).

2 See, e.g., George Psacharopoulos and Harry Anthony Patrinos, eds., *Indigenous People and Poverty in Latin America: An Empirical Analysis* (Washington, D.C.: World Bank, 1994).

3 See, e.g., for the United States a series of studies by Gary D. Sandefur and his colleagues: Gary D. Sandefur, "American Indian Migration and Economic Opportunities," *International Migration Review* 20 (1986): 55–68; Gary D. Sandefur and A. Sakamoto, "American Indian Household Structure and Income" *Demography* 25 (1988): 71–80; Gary D. Sandefur and A. Pahari, "Racial and Ethnic Inequality in Earnings and Educational Attainment," *Social Service Review* 63 (1989): 199–221; Gary D. Sandefur and Wilbur J. Scott, "Minority Group Status and the Wages of Indian and Black Males," *Social Service Research* 12 (1983): 44–68; Gary D. Sandefur, S. McLanahan, and R. A. Wojtkiewicz, "Race and Ethnicity, Family Structure, and High School Graduation," Discussion Paper no. 893–89 (University of Wisconsin—Madison, Institute for Research on Poverty, 1989); and C. M. Snipp and Gary D. Sandefur, "Earnings of American Indians and Alaskan Natives: The Effects of Residence and Migration," *Social Forces* 66 (1988): 994–1008. See also Barry R. Chiswick, "Differences in Education and Earnings across Racial and Ethnic Groups: Tastes, Discrimination, and Investments in Child Quality," *Quarterly Journal of Economics* 103, no. 3 (August 1988): 571–97. Robert G. Gregory, Annie C. Abello, and Jamie Johnson, "The Individual Economic Well Being of Native American Men and Women during the 1980's: A Decade of Moving Backwards," *Population Research and Development Review* 16, nos. 1–2 (April 1997): 115–45; James D. Gwartney and James E. Long. "The Relative Earnings of Blacks and Other Minorities," *Industrial and Labor Relations Review* 31, no. 3 (1978): 336–46; and Michael E. Hurst, "The Determinants of Earnings Differentials for Indigenous Americans: Human

Capital, Location or Discrimination?" *Quarterly Review of Economics and Finance* 37, no. 4 (Fall 1997): 787–807. For Canada, see, e.g., C. Y. Kuo, "The Effect of Education on the Earnings of Indian, Eskimo, Metis, and White Workers in the Mackenzie District of Northern Canada," *Economic Development and Cultural Change* 24, no. 2 (1976): 387–98; E. H. Lautard, "Occupational Segregation and Inequality between Native and Non-native Canadians, 1971," *Canadian Journal of Native Studies* 2 (1982): 303–20; Harry A. Patrinos and Chris N. Sakellariou, "North American Indians in the Canadian Labour Market: A Decomposition of Wage Differentials," *Economics of Education Review* 11, no. 3 (1992): 257–66; and Jack C. Stabler, "Dualism and Development in the Northwest Territories," *Economic Development and Cultural Change* 3 (1989): 805–39. For Australia and New Zealand, see, e.g., Peter Brosnan, "Age, Education and Maori-Pakeha Income Differences," *New Zealand Economic Papers* 18 (1984): 49–61; and Paul W. Miller, "The Structure of Aboriginal and Non-Aboriginal Youth Unemployment," *Australian Economic Papers* 28 (1989): 39–56.

4 Jonathan Kelley, "Class Conflict or Ethnic Oppression? The Cost of Being Indian in Rural Bolivia," *Rural Sociology* 53, no. 4 (1988): 399–420.

5 George Psacharopoulos, "Ethnicity, Education, and Earnings in Bolivia and Guatemala," *Comparative Education Review* 37, no. 1 (1993): 9–20.

6 Harry A. Patrinos, Eduardo Velez, and George Psacharopoulos, "Language, Education and Earnings in Asuncion, Paraguay," *Journal of Developing Areas* 29, no. 1 (1994): 57–68.

7 Donna J. MacIsaac and Harry Anthony Patrinos, "Labor Market Discrimination against Indigenous People in Peru," *Journal of Development Studies* 32, no. 2 (1995): 218–33.

8 Harry Anthony Patrinos, "Difference in Education and Earnings across Ethnic Groups in Guatemala," *Quarterly Review of Economics and Finance* 37, no. 4 (Fall 1997): 809–22.

9 See, e.g., MacIsaac and Patrinos; Psacharopoulos and Patrinos; Patrinos; Martonez Plaza, "Towards Standardization of Language Teaching in the Andean Countries," *Prospects* 20, no. 3 (1990): 377–84; and Juan Chackiel and Alexia Peyser, "Indigenous Population from Latin American National Censuses" (photocopy, Latin American Demographic Centre, CELADE, August 1993).

10 See Barry R. Chiswick and Paul W. Miller, "The Endogeneity between Language and Earnings: International Analyses," *Journal of Labor Economics* 13, no. 2 (April 1995): 245–87, and "English Language Fluency among Immigrants in the United States," *Research in Labor Economics* 17 (1998): 151–200.

11 Michael H. Long, "Maturational Constraints on Language Development," *Studies in Second Language Acquisition* 12, no. 3 (1990): 251–85.

12 The language of instruction in the schools is Spanish even in the lowest grades in the rural areas in which most people speak an indigenous language and do not know Spanish. There have been only a few small experiments with Spanish-native language instruction in schools with indigenous-speaking children. See Lambros Comitas, "Education and Social Stratification in Contemporary Bolivia," in *Education and Development: Latin America and the Caribbean*, ed. Thomas J. LaBelle (Los Angeles: University of California Press, 1972), pp. 363–78; and Xavier Albo and Lucia d'Emilio, "Indigenous Languages and Intercultural Bilingual Education in Bolivia," *Prospect* 20, no. 3 (1990): 321–29.

13 For an analysis of this issue, see Jacob Mincer, "Family Migration Decisions," *Journal of Political Economy* 85, no. 5 (October 1978): 749–74.

14 Jacob Mincer, *Schooling, Experience and Earnings* (New York: National Bureau of Economic Research, 1974).

15 Barry R. Chiswick, "The Effect of Americanization on the Earnings of Foreign-Born Men," *Journal of Political Economy* 86, no. 5 (October 1978): 897–922.

Part V

Language, networks and enclaves

17 Do enclaves matter in immigrant adjustment?

Introduction

This chapter is concerned with the issue of immigrant/ethnic concentrations, that is, the tendency of immigrants to concentrate geographically by ethnicity or country of origin within the host country.[1] In particular, it is concerned with the consequences of enclaves or concentrations for two characteristics of immigrant adjustment—destination language proficiency and labor market earnings. Other aspects of immigrant life influenced by concentrations, including political participation and influence, are beyond the scope of this chapter.

There are two basic research questions of interest. One is the effect of immigrant concentrations on proficiency in destination language skills. The other is the direct effect of the immigrant's proficiency in the destination language and the effects of these immigrant concentrations on their labor market earnings. In particular, this study separates the direct effects and indirect effects via language proficiency of immigrant concentrations on earnings. The application is to the United States. The methodology developed, however, could be applied to any immigrant-receiving country for which there is appropriate census or survey data.

The section "Immigrant Concentrations: Hypotheses" provides a brief introduction to the broader setting within which the issue of immigrant concentrations arises. Testable hypotheses are developed, with a particular emphasis on ethnic goods. "The Data" section discusses the data used in the empirical analysis. In the "Analysis of Language" section a model of dominant language acquisition is presented and estimated, with a particular focus on the effects of immigrant/linguistic concentrations on dominant language proficiency. "Analysis of the Earnings" section is the analysis of the earnings of immigrants with a particular focus on the effects of the immigrant's destination language skills and living in a linguistic concentration area on the respondent's labor market earnings. The chapter closes with a summary and conclusion, with implications for public policy.

Immigrant concentrations: hypotheses

Immigrant flows

A characteristic of the late 20th century that is surely to continue into at least the early 21st century is an increase in the movement of people across international borders (Chiswick and Hatton, 2003). International migration has increased into the traditional immigrant-receiving countries, such as the United States, Canada, and Australia. Yet, international migration into traditional countries of emigration has also become commonplace. Italy, Ireland, Germany, and Japan, among others, are now experiencing large net in-migration, or where restricted by law, as in Japan, pressures for in-migration as evidenced by illegal flows.

These migration flows have, in part, been "East" to "West," that is, from the former Soviet Union and the Eastern block countries to the United States, Canada, Germany, and Israel. More pronounced, however, are the migration flows from the "South" to the "North," more precisely, from less developed countries to highly developed economies. Unprecedented immigration flows have been experienced from Latin America to North America, from Africa to Western Europe, and from Asia to North America, Western Europe, Australia, and Japan (Chiswick and Hatton, 2003).

An important characteristic of these international migration flows is that the immigrants are "different" from the natives. As was true of the immigration flows from Southern and Eastern Europe to North America at the turn of the 20th century, the immigrants to the developed countries at the turn of the 21st century "sound" and "look" different.

In spite of the world becoming a smaller place with the ease (falling cost) of the transmission of information and ideas, and the falling cost of transportation for people and goods, and hence the "Westernization" of much of the world, new immigrants are frequently distinctive. Although distinctive clothing, especially for men, is less common than in the past, immigrants as a group frequently differ from natives as a group in appearance, religion, customs, belief-systems, language, and other characteristics associated with ethnicity.

Immigrant concentrations

The immigrant groups typically have a spatial distribution in their host countries that differ sharply from that of the native born. For obvious reasons, new immigrants typically settle in areas based on three characteristics (Bartel, 1989).[2] The first is "ports" of entry, near seaports in the past, near airports in the current era. The second is where family and friends (co-ethnics) from earlier migrations have settled. Even if the location choice of the first settler from the ethnic group is purely random among a set of equally attractive locations in a destination country, once that first settler is established, future settlers are no longer indifferent among destination sites. The third is where the jobs are, that is, where the immigrants are most able to gain employment that makes best use of their skills, or lack thereof. With the passage of time "ports of entry" and "family and friends" become less central and economic factors relatively more central in deciding where to live in the host country, and immigrants tend to disperse to some extent.

Some interpret the "family and friends" or chain migration effect on immigrant formations of concentrations as "clannishness." Yet to say it is clannishness is to beg the question as "clannishness" per se has no content. An alternative interpretation, however, is that settling in areas with others from the same origin provides for economies in communication, information, consumption, and in the labor market.

Where new immigrants differ from the host population in terms of language skills, communication in all spheres of life are that much more difficult. These communication costs can be reduced if the host population were to learn the immigrant's language. Yet, it is not cost effective for a majority host population speaking the dominant language to learn the myriad of new languages that minority immigrants bring with them from various linguistic backgrounds.

These communication costs are reduced when immigrants learn the dominant language of the destination country. Yet, this learning can be costly and cannot be done instantaneously in the destination. Thus, to varying degrees, new immigrants from a different linguistic origin tend to lack complete proficiency in the dominant language of the host economy, unless dominant language proficiency is a requirement for entry. Moreover, as with the production of other forms of human capital and of market goods and services, beyond some point, costs per unit of improved proficiency increase with a faster speed of language acquisition. Thus, the optimal acquisition of dominant language proficiency among immigrants takes time and, for some, full proficiency may never be obtained in their lifetime.

Finally, these communication costs for the immigrants can be reduced by living and/or working in a linguistic concentration area (Bauer, Epstein, and Gang, 2002). Not all members of the group need dominant language proficiency, and the earlier arrivals and those more efficient in language acquisition are more likely to become proficient. They can serve as either direct or indirect translators for communication between the enclave and the host society. The demand for this specialized function increases with the size of the linguistic minority group and decreases as the members of the group learn the dominant language or as the native population learns the immigrant language.

Even aside from issues of language skills, immigrant/ethnic concentrations provide information networks that can be very valuable in social interaction, consumption, and employment activities. Natives of an area have acquired location-specific human capital, which includes information obtained directly and indirectly through established networks. Not being connected to host country information networks when they arrive, immigrants have an incentive to create or "import" information networks through living in geographic concentrations with other new and longer term immigrants from the same origin.

Ethnic goods

Immigrants tend to differ from the native or host population in many dimensions related to ethnicity. They may differ in the foods they eat, the clothing they wear, the holidays they celebrate, the religion they practice, the media they read or hear (e.g., newspapers and radio), their social organizations, and the languages they speak, among other characteristics.[3] There is frequently a tension among immigrants between preserving the culture of the "old country" in the new setting and adopting the culture of the host country.

Let us call "ethnic goods" the consumption characteristics of an immigrant/ethnic group not shared with the host population, or with other immigrant groups, broadly defined to include market and nonmarket goods and services, including social interactions for themselves and their children with people of their same origin.[4] To the extent that "ethnic goods" are distinctive and are important in the market basket, immigrants from a particular origin have a different market basket than the native born and immigrants from other origins. The full cost of consumption of these ethnic goods varies with the price of purchased market goods and services and the value of time, but also with the importance and distinctiveness of the ethnic goods and the size of the group.[5]

There are certain fixed costs and economies of scale in the production and distribution of ethnic goods. Social interaction with others of the same origin (including finding an appropriate marriage partner) may involve little in the way of conventional market goods and services, but importantly involves the number of other individuals in the group. The cost would decrease (presumably at a decreasing rate) the larger the size of the group.[6] Up to a point, an ethnic religious institution (e.g., church, mosque, temple, or synagogue) or an ethnic school for the children of immigrants has a lower per capita cost for members for the same type of facility providing the same level of services to the congregants or students if it is in a larger rather than in a smaller ethnic community.[7] There are fixed costs for buildings and hiring religious officials, among other items, including the probability that enough individuals will show up on a given occasion for the religious service.

The cost of "importing" into the community ethnic-specific goods (e.g., saris, Chinese vegetables, kosher meats) also varies with the size of the market because of economies of scale. Indeed, as the size of the community increases, the manner of "importation" may change from a family making a trip to a larger nearby community, to collective/cooperative efforts to place periodic bulk orders, to the establishment of a single (monopoly) outlet, to many competitive outlets selling the product. The full price declines, the larger the size of the community.

The cost of living in an area then depends on the relative cost of ethnic goods, broadly defined, and the importance and distinctiveness of ethnic goods in the person's market basket. The cost of ethnic goods is lower, the larger is the size of the particular ethnic/immigrant community. The share of ethnic goods in the market basket is likely to be lower, the closer culturally the group is in the origin to the host society, the greater the extent of assimilation into the host society, the longer the immigrant's duration of residence in the destination, and among the native-born descendants of immigrants.

Ethnic goods have implications for living in an ethnic concentration area as well as for geographic differences in earnings. If ethnic goods, defined broadly, are an important part of the market basket, the person faces a higher real cost of living where ethnic goods are more expensive (an area where fewer co-ethnics live) than where they are less expensive (a high ethnic concentration area). Then the ethnic immigrant would be indifferent between a similar job in a high-concentration area and a low-concentration area only if the latter provided a higher nominal wage that was just sufficient to compensate for the higher cost of living.[8]

Thus, ethnic goods can result in different geographic concentrations of various immigrant groups and differences in the pattern of regional wage differentials across immigrant groups and between immigrants and natives. The general observation would be lower nominal wages, the larger the size of the concentration, other variables being the same. Note that the "ethnic goods" hypothesis regarding the negative relation between the concentration measure and earnings is an equilibrium situation based on differences in the real (ethnic-specific) cost of living. It reflects compensating wage differentials.[9]

When a new immigrant group initially arrives in a destination it may be indifferent among alternative regions in the destination that are equally attractive in terms of job opportunities and ports of entry. The initial settlers would tend to be immigrants with a lower demand for ethnic goods. Subsequent immigrants from this ethnic group will not be indifferent among the alternative destinations as ethnic goods will be cheaper where their co-ethnics have already settled. With the ethnic community established, those with a higher demand for ethnic goods would find immigration much more attractive.

New ethnic concentrations away from the original center in the destination can be formed under any one of several scenarios. An individual with a very low demand for ethnic goods may settle elsewhere and gradually (and perhaps inadvertently) serve as a nucleus for others to follow. An individual with a high demand for ethnic goods may randomly receive a very high wage offer from the distribution of wage offers and settle in a new area. This person may serve as a nucleus and may even have an economic incentive to subsidize ethnic goods to encourage others to join him or her in the new location.

Moreover, if a very "large" number of immigrants settle in the initial location and they are less than perfect substitutes in production for native workers, under the crowding hypothesis their wages decline relative to what they could earn in alternate locations with fewer (perhaps none) of their group. If the wage gap compensates for the higher cost of living because of ethnic goods, a second enclave can be established. Thus, the number of enclaves or areas of concentration will vary systematically with the size of the immigrant/ethnic group and the distinctiveness and intensity of the demand for ethnic goods.

Immigrant "crowding"

An alternative to the "ethnic goods" hypothesis is a labor supply or "crowding" hypothesis. If there are a large number of immigrants with a given skill level, and if they are not good substitutes in production for others with the same skill level, their earnings would be depressed, as indicated in the previous paragraph. This is, however, a disequilibrium situation as immigrant workers with a given level of skill could receive higher real wages outside the enclave. The internal mobility of immigrant and native-born labor, and other factors of production, as well as goods and services, would bring about factor price equalization, eliminating the negative relation between concentration and earnings.

The "crowding hypothesis" is not likely to be compelling for the United States. The United States has highly fluid labor, capital, and product markets where inter-regional mobility is the norm rather than the exception. The largest single group of immigrants is from Mexico, and they tend to have low levels of skill, without a high degree of specialized skills.[10] As such, they are good substitutes in production for other low-skilled labor, whether native born or foreign born. Among the non-Mexican immigrants, the countries and languages of origin are numerous and skill levels are more highly varied. It is difficult to think of any groups in the U.S. that are sufficiently large and specialized with a low substitutability with native-born and other foreign-born workers. To the extent that a sudden exogenous infusion of immigrant labor with specialized skills impacts a local labor market, disequilibrium earnings differentials would emerge, but would be dissipated over time with internal mobility of factors of production (including immigrant labor) and tradable goods. The persistence over time of immigrant concentrations is not consistent with the implications of the crowding hypothesis effect on wages.

Consequences of concentrations

Limited destination language proficiency is likely to reduce the earnings potential of immigrants (Chiswick and Miller, 1992, 1995). It raises the cost or lowers the efficiency of job search and in many jobs may restrict access (e.g., if there is a need to pass a test that requires proficiency) or merely lower productivity. There may also be discrimination in the labor market by the native population (either as employers, co-workers, or consumers) against those who are less proficient in the dominant language or who speak it with an accent. Working within a linguistic enclave is a mechanism for sheltering oneself from, or mitigating the adverse labor market consequences of, limited destination language proficiency.

Living and working within a linguistic concentration area has feedback effects on destination language proficiency. The greater the extent to which an individual can avoid communicating in the destination language, the slower is likely to be the rate of acquisition of dominant language skills. Consider two individuals: one lives in a large linguistic concentration area where one can work, consume, socialize, and engage in other activities using the origin language and the other lives in a linguistically isolated area; communication can be done only in the dominant language. The latter may have a more difficult initial adjustment, but has a stronger incentive to acquire destination language skills and has greater exposure that facilitates learning the destination language.

Thus, what has emerged in many developed countries is the existence of distinct immigrant communities that differ in language, culture, and other characteristics from the host society. These immigrant/linguistic concentrations are expected to have an adverse effect on the immigrant's acquisition of dominant language skills. The immigrant's dominant language skills, as well as the size of the linguistic concentration area, will also affect the person's earnings, other things being the same. Greater proficiency would have a positive effect, and a larger concentration a negative effect on nominal earnings. These hypotheses are tested in the empirical analysis.

The data

Defining the population under study

The empirical analysis is performed using data from the 1990 Census of Population of the United States for adult male immigrants.[11] The U.S. Census provides a very large sample, a rich array of variables, and immigrants from diverse origins arriving at various periods of time. The analysis at this stage is limited to adult (nonaged) males as the analysis for females or aged males becomes more complex because of the need to model labor supply decisions, in addition to the language and earnings equations. Moreover, the formation of enclaves or concentrations is taken as exogenous for the individual in the empirical analysis, although there was a discussion in Section II as to why such concentrations are formed.

The data for the statistical analysis are from the 5 percent Public Use Microdata Sample from the 1990 Census. The sample is limited to males aged 25 to 64 years who were foreign born, but not from an English-speaking developed country. Thus, the native born, those born in a U.S. territory (e.g., Puerto Rico), born at sea, or born abroad of American parents are excluded, as are those born in the United Kingdom, Ireland, Canada, Australia, and New Zealand.

Defining the variables

The English language proficiency variable comes from question number 15 in the census long form. Respondents were asked if there was a language other than English spoken at home (other than just a few words), and if so the identity of that language and how well they spoke English, where the response categories were Very Well, Well, Not Well, and Not at All. For the purpose of this analysis, the foreign born who spoke only English or who spoke another language but reported that they spoke English "very well" or "well" were considered fluent: those who spoke English "not well" and "not at all" were considered not fluent.

The other dependent variable is earnings, which is the sum of wage, salary, and self-employment income in 1989. Those with earnings of less than $100, including those with negative earnings, were assigned a value of $100. Those who worked 0 weeks in 1989 were deleted from the sample for the analysis of earnings.

The enclave variable is a minority language concentration measure (CON).[12] The 24 languages other than English most frequently spoken in the United States were identified. The speakers of these top 24 languages constitute around 94 percent of those reporting a foreign language spoken at home. For each of these 24 languages, for the 50 states and the District of Columbia, the percent of the states' population aged 18 to 64 years (whether native or foreign born) speaking that language, was computed. The concentration measure for each respondent is the percentage speaking the person's origin language in the state of current residence. For other languages, since the number of speakers is too low, the percentage was assumed to be zero. Those who reported speaking only English were assigned the mean value of the concentration ratio for foreign language speakers in their birthplace group.

Within states, the density of population is less in rural areas than in urban areas. A variable for residence in a rural area (RURAL) is included because of a smaller concentration of origin language speakers in rural than in urban centers. The other explanatory variables are straightforward and are discussed in Appendix 17.B and as the variables are introduced in the text.

The statistical techniques

The main statistical methodology that is employed is ordinary least squares (OLS) with standard errors corrected for heteroskedasticity, and where indicated below instrumental variables (IV) analysis.

Analysis of language

The language model

This section presents the development of the model for dominant language proficiency. While largely based on previous work, in particular Chiswick and Miller (1995, 1998), the model is expanded to include new variables (refugees, persons from former colonies of English-speaking countries, and persons who lived abroad 5 years earlier). Particular attention is given to the variable measuring the degree of minority language concentration (CON).

The language proficiency model adopts a human capital approach in which destination language proficiency (LANG) is a function of three fundamental determinants, namely, "exposure," "efficiency," and "economic incentives." Since the application is to the English language for immigrants in the United States in the 1990 Census, the discussion of these variables will be in this context. The principles apply to any destination language, country, and data set.

Exposure variables

"Exposure" refers to exposure to the destination language either pre- or post-immigration. The Census identifies country of birth, but provides no other information on preimmigration experiences relevant for acquiring English language proficiency. A set of country-of-origin dichotomous variables is included in the analysis to control for country of origin fixed effects. Western Europe (other than the U.K. and Ireland) is the benchmark. Moreover, a dichotomous variable is created for whether the origin was a colony (COLONY) of an English-speaking country, that is, of either the United States or the United Kingdom. Recall that respondents born in current U.S. territories are excluded from the analysis.

Postimmigration exposure to English can be measured in time units and in intensity per unit of time. Time in the destination is measured as the number of years since migration and its square (YSM, YSMSQ). It is expressed as a quadratic variable to allow for the effect of an extra year in the United States to be larger in the early years than in subsequent years.

The duration variable refers to when the immigrant first came to the United States to stay. Exposure to English in the United States may have been interrupted by sojourns outside the country after the initial migration. For immigrants in the United States for more than 5 years the variable "lived abroad five years ago" (ABROAD5) is unity if this was the situation, otherwise it is zero. It is expected that, other variables being the same, having lived outside the United States would be associated with lesser proficiency in English compared to otherwise similar immigrants who did not live elsewhere in 1985.

Intensity of exposure per unit of time in the United States can be measured by several variables. Of particular interest is the minority language concentration measure (CON), which is computed on a state level, as was discussed above. Within states the density of population is less in rural areas than in urban areas. A variable for residence in a rural area (RURAL) is included because of a smaller concentration of origin language speakers in rural than in urban areas within states.

For immigrants from Mexico the analysis also includes an index for Spanish language media, namely, a variable for the number of radio stations in Spanish normalized for the size of the state in square miles and population (RADIO) (Chiswick and Miller, 1998). Because of possible endogeneity in this variable, a predicted value (IV technique) rather than an observed value for radio is used.

A marital status variable (MARR is unity if married, spouse present) is also included here. It is not possible in the 1990 Census to distinguish between pre- and post-migration marriages, but it was possible to do in the 1980 Census (Chiswick and Miller, 1992).

Efficiency variables

"Efficiency" refers to the ability to convert exposure into language skills. Greater efficiency means more language skills are acquired for the same level of exposure. The efficiency variables include age at migration (age with years since migration held constant), years of schooling, whether the respondent may have been a refugee, and a measure of the "distance" between the origin language and English.

Older immigrants (AGE) at arrival have greater difficulty learning a new language. Age is entered as a quadratic variable (age and its square) as it is expected that an extra year of age at migration would have a larger adverse effect among younger than among older immigrants.

Those with more schooling (EDUC) are assumed to be more able and to have more knowledge of the structure of languages, and hence are likely to be more efficient in learning new languages, including the destination language. It may also be that those with more schooling in the origin were exposed to English at higher grades prior to immigration, or that schooling in the United States enchanced proficiency.[13]

The refugee variable (REFUGEE) is included because refugees tend to be less favorably selected for a successful adjustment in the destination than are economic migrants. The migration decision of refugees is influenced to an important extent by factors other than the expectation of a successful economic adjustment. The refugee variable is based on country of birth and period of immigration.

Another efficiency variable is "linguistic distance" (DISTANCE), that is, a measure of how difficult it is for non-English speakers to learn English (see Chiswick and Miller, 1998). For example, Korean would be more "distant" from English than would be French. The more "distant" is the origin language from English, the lower the efficiency in learning English and hence the lower the expected proficiency in English.

Economic variables

"Economic incentives" is the most difficult conceptual variable to model. In principle, one would like to add an explanatory variable that measures the expected increment in earnings for a unit increase in proficiency for each respondent. Given currently available data it is not possible to do this. It has been found that immigrants with higher levels of schooling have a greater economic return from becoming proficient (Chiswick and Miller, 2003). That is, that there is a complementarity between language skills and education in generating labor market earnings. This effect on incentives to invest in destination language skills would be captured by the education variable (EDUC).

The economic benefits in the labor market and in other activities from increased proficiency in English would be greater the longer the expected duration in the United States. Immigrants from countries with a high propensity for return migration would expect a shorter period in the U.S. Data on emigration (EMIG) by country of origin are used for this purpose, but not for Mexico (Ahmed and Robinson, 1994). The methodology for developing the country-specific emigration rates in Ahmed and Robinson (1994) is not applicable to Mexico because of the 1986 amnesty and the very large proportion of illegal aliens among the foreign born from Mexico.

Immigrants from countries farther from the United States are more likely to be favorably self-selected as they have higher costs of migration (Chiswick, 1999). This implies a higher level of efficiency in learning English. They also have a lower return migration rate, again because of the higher migration costs. Those from origins at a greater distance from the United States are, therefore, expected to be more fluent (Chiswick and Miller, 1998). This is measured as the number of miles (XMILES) from the major city in the origin to New York, Miami, or Los Angeles, whichever is the shortest. It is entered as a quadratic variable.

While state-specific (fixed) effects are not held constant because the concentration ratio is based on state data, a control variable is entered for Southern states (SOUTH).

Statistical analysis

The means and standard deviations of the language variable (LANG) and the explanatory variables, overall and separately for Mexican and non-Mexican immigrants, are reported in Table 17.A1. The regression equations for English language proficiency are reported in Table 17.1 for all immigrants, non-Mexican immigrants, and Mexican immigrants.

The data are found to be consistent with the hypotheses developed above.

Table 17.1 Regression estimates of language equation, adult foreign-born men by origin, 1990

Variable	*Total Sample*[a]	*Excludes Immigrants from Mexico*[a]	*Immigrants from Mexico Only*[b]
Constant	0.409	0.478	0.440
	(26.28)	(26.76)	(10.64)
Age	–0.010	–0.011	–0.006
	(15.69)	(16.59)	(3.96)
Age squared/100	0.003	0.006	–0.003
	(4.96)	(7.55)	(1.33)
Years of education	0.029	0.030	0.028
	(141.10)	(119.57)	(64.67)
Years since migration (YSM)	0.021	0.018	0.027
	(100.30)	(83.11)	(48.35)
YSM squared/100	–0.025	–0.023	–0.027
	(56.69)	(50.82)	(22.23)
Married	0.033	0.020	0.053
	(19.29)	(11.02)	(12.06)
Rural	0.010	0.021	0.002
	(3.00)	(6.93)	(0.26)
South	0.013	0.010	0.028
	(7.16)	(5.41)	(5.75)
S. Europe	–0.028	–0.033	[c]
	(9.47)	(10.88)	
E. Europe	–0.047	–0.063	[c]
	(12.19)	(15.41)	
USSR	–0.030	–0.039	[c]
	(4.75)	(6.19)	
IndoChina	–0.090	–0.093	[c]
	(9.19)	(9.22)	
Philippines	0.024	0.014	[c]
	(3.43)	(1.78)	
China	–0.123	–0.128	[c]
	(17.17)	(16.15)	
S. Asia	–0.011	–0.020	[c]
	(1.38)	(2.26)	
Other Asia	–0.036	–0.044	[c]
	(3.27)	(3.91)	
Korea	–0.202	–0.207	[c]
	(21.62)	(18.04)	
Japan	–0.108	–0.116	[c]
	(10.74)	(9.84)	
Middle East	0.010	0.009	[c]
	(2.52)	(1.10)	

Sub-Saharan Africa	0.032	0.028	c
	(6.28)	(2.79)	
Mexico	–0.067	c	c
	(12.10)		
Cuba	0.044	0.040	c
	(5.17)	(3.97)	
C. and S. America (Spanish)	–0.019	–0.042	c
	(4.10)	(8.25)	
C. & S. America (non-Spanish)	0.219	0.208	c
	(32.97)	(30.76)	
Minority language concentration	–0.004	–0.003	–0.010
	(17.81)	(9.73)	(7.84)
Linguistic distance	–0.005	–0.006	c
	(0.36)	(0.44)	
Miles from origin/1,000	0.050	0.054	c
	(14.57)	(15.19)	
Square of miles from origin/1,000	–0.034	–0.038	c
	(9.16)	(10.19)	
Refugee	–0.123	–0.138	c
	(32.12)	(35.19)	
Colony	0.012	0.013	c
	(3.53)	(3.90)	
Resident overseas 5 years ago	–0.069	–0.046	–0.073
	(11.34)	(6.76)	(3.13)
Emigration rate	c	–0.010	c
		(2.37)	
Spanish radio	c	c	–21.98
			(4.11)
$\bar{R}^2$	0.3244	0.3176	d
Sample size	237,766	169,253	68,512

Notes: *t*-statistics have been computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator.

a Equation estimated using ordinary least squares; b equation estimated using instrumental variables (IV) estimator; c variable not applicable; d R^2 not defined for the IV Model. IV estimator used for Spanish Radio variable.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent sample.

In particular, English language proficiency is greater the higher the level of schooling, the longer the duration of residence (quadratic effect), the younger the age at immigration (negative effect of age), among those from a former British or American colony, and from countries more distant from the United States. It is less among refugees, among transients (i.e., immigrants who first came to the United States more than 5 years earlier—prior to 1985—but who were outside the United States in 1985), and where the expected duration in the United States (emigration rate variable) is shorter. The linguistic distance variable is not statistically significant when country of origin fixed effects are included in the analysis, as is the case in Table 17.1, but it is significant with the expected negative effect when the country dichotomous variables are excluded from the equation. This arises from the close relation between country of origin and language of origin.

The minority language concentration variable (CON) is highly statistically significant as is the rural variable (RURAL), which is a proxy for the concentration of foreign language speakers within areas in states. According to the regression for all immigrants, going from a minority language concentration of zero to the mean value of 7.8 percent lowers the probability of being fluent in English by 3.1 percentage points, which is 4.2 percent of the mean proficiency of 0.73 or 73 percent. Rural residence (5.5 percent of the foreign born) raises proficiency by 1.0 percentage point overall and by 2 percentage points among non-Mexican immigrants.

Among Mexican immigrants three variables reflect the effect of the linguistic concentration of Spanish speakers. One is the direct minority language concentration measure, the second is the rural variable, while the third is the (predicted) Spanish language radio station variable. The minority concentration measure and the radio station variable, but not the rural variable, are highly statistically significant with the expected negative signs.

Thus, the analysis of English language proficiency among immigrants from non-English origins in the United States indicates that the data are consistent with the model based on exposure, efficiency, and economic variables. Moreover, it is found that linguistic concentrations or enclaves are associated with a lesser proficiency in English among all, Mexican and non-Mexican immigrants.

Analysis of earnings

The earnings model

The econometric analysis of earnings is based on the human capital earnings function, modified for immigrant adjustment (Chiswick, 1978). In this specification, the natural logaritham of annual earnings (LNEARN) is regressed on years of schooling (EDUC), years of potential labor market experience and its square (EXP, EXPSQ), duration in the United States and its square (YSM, YSMSQ), the natural logarithm of weeks worked (LNWW), marital status (MARR), and place of residence (RURAL, SOUTH). Three dichotomous variables are added to the equation which take the value of unity for immigrants whose race is Black, are Veterans of the U.S. Armed Forces, and who are Citizens of the United States. Two other variables are also added to this equation, the respondent's proficiency in English (LANG), which is unity for those fluent in English, as defined above, and zero otherwise, and the minority language concentration measure (CON).

Statistical analysis

The earnings equation is estimated separately for all immigrants, Mexican immigrants, and non-Mexican immigrants. The means and standard deviations of the variables are reported in Table 17.A2, while Tables 17.A3–17.A5 report the regression equations for each group. A basic earnings function in these tables is reported in column (*i*) without the language and concentration variables, column (*ii*) adds the English language proficiency variable (LANG), column (*iii*) adds the concentration variable (CON) to the basic equation, column (*iv*) adds both variables, while column (*v*) substitutes a predicted English language proficiency variable obtained through the IV technique. (The auxiliary equation is reported in Table 17.A6.). A summary of the language and concentration variable results is presented in Table 17.2.

Ordinary least squares analysis

As has been found elsewhere, the basic determinants of earnings among immigrants are also found to be important here (see Tables 17.A3–17.A5). For immigrants from non-English speaking countries, earnings increase with years of schooling (by about 5 percent per year of schooling), duration in the United States (at a decreasing rate), preimmigration labor market experience (total experience when duration is held constant), and weeks worked (with an elasticity of annual earnings with respect to weeks worked close to unity), and are higher for married men (by about 20 percent) and citizens (9 percent). Earnings are lower for immigrants who are veterans of the U.S. Armed Forces (8 percent), and among those living in rural areas (4 percent) and in the south (11 percent).

Similar patterns are found when the analysis is done separately for

Table 17.2 Partial effects on earnings of the language and concentration variables, adult foreign-born men from non-English speaking countries, 1990

Variables	*Total Sample*		*Excludes Immigrants from Mexico*		*Immigrants from Mexico Only*	
	OLS	*IV*	*OLS*	*IV*	*OLS*	*IV*
Proficient in English	0.148	0.592	0.151	0.678	0.146	[a]
	(31.60)	(16.53)	(22.40)	(16.40)	(23.52)	
Minority language concentration	−0.0056	−0.0039	−0.0070	−0.0050	−0.0033	[a]
	(15.25)	(9.62)	(11.77)	(7.85)	(7.13)	

Notes: Full regression equations are reported in column (*iv*) and column (*v*) of Tables 17.A3–17.A5. OLS refers to ordinary least squares. IV refers to instrumental variables technique using predicted value of respondent's proficiency in English.

[a] IV equation is not computed for Mexico; *t*-ratios are in parentheses.

Source: Tables 17.A3–17.A5.

non-Mexican and for Mexican origin immigrants (Tables 17.A4 and 17.A5).[14] Note that the effects of several variables reflecting human capital are smaller for Mexican immigrants than for other immigrants. These include schooling, experience, and weeks worked, but not duration in the United States.

The OLS analysis in Tables 17.A3–17.A5 indicates that earnings are about 15 percent higher for all immigrants, Mexican immigrants and non-Mexican immigrants who are proficient in English, compared to those lacking proficiency. The difference is statistically significant and the magnitude of the effect and level of significance do not vary with whether the concentration measure is included in the analysis.

Assuming a long working life, the real rate of return on the investment in language proficiency can be estimated (approximately) as $r = b/k$, where r is the real rate of return, b is the regression coefficient of the language proficiency variable, and k is the number of full-year equivalents of lost earnings (including out-of-pocket expenditures and foregone earnings), to go from not proficient ("not well," "not at all") to proficient (English only, "very well" or "well"). Then, if the coefficient of the language variable is $b = 0.15$ and if the full cost is the equivalent of a full year's potential earnings ($k = 1$), the rate of return is about 15 percent. If the cost were the equivalent of 2 years of full-time equivalent earnings ($k = 2.0$), the rate of return on the investment would be about 7.5 percent. If proficiency required the equivalent of only 6-months foregone earnings ($k = 0.5$), the estimated rate of return would be about 30 percent. The rate of return would be even higher if the positive effects of proficiency on weeks worked in the year were included in the calculation and if the consumption benefits from English language proficiency could be estimated. Thus, investments in English language skills appear to be profitable for immigrants from non-English speaking countries.

The concentration measure is also statistically significant in all three analyses. Earnings are lower where the concentration index is higher.[15] The coefficient and level of significance are also largely invariant with respect to the inclusion in the analysis of the respondent's fluency in English. Among all immigrants, going from a zero concentration area to the mean level (7.8 percent) lowers earnings by about 4.4 percent (i.e., 7.8 times 0.0056 from Table 17.A3, column (*iv*)). For non-Mexican immigrants (mean concentration 3.9 percent) it lowers earnings by about 2.7 percent. Among Mexican immigrants, the mean of the concentration ratio is much higher (18.1 percent), but the coefficient of the concentration ratio is lower (–0.0033 compared to –0.0070 for other countries). For Mexican immigrants, the effect of going from a zero concentration to the mean concentration ratio is to lower earnings by about 6.0 percent. Thus, other variables the same, including the respondent's own proficiency in English, living in a linguistic/ethnic concentration area lowers the earnings of immigrants.[16] Moreover, the estimated magnitude of the effect is in a reasonable range.

The effect of the concentration ratio on earnings varies systematically with the level of education. If an education-concentration ratio interaction term is added to the regression in Table 17.A3, column (*iv*), it has a negative and highly significant effect.[17] That is, the adverse effect on earnings from living in a high-concentration area is greater the higher the level of schooling.[18] There is no effect for those with only 5 years of schooling, but the negative effect of living in a high-concentration area grows larger at higher levels of schooling. Alternatively, this can be expressed as the effect of education on earnings is smaller in the high-concentration (enclave) area than in an area where fewer other individuals speak the same origin language.

Instrumental variables analysis

There are several potential econometric problems with the OLS analysis using the respondent's reported level of English language proficiency. One problem is that language skills may be endogenous to, that is, determined by, earnings. Those who anticipate higher earnings if they were to become proficient will make greater investments to acquire proficiency (Chiswick and Miller, 1995).

A second problem is that there may be substantial measurement error in reported language skill. Purely random measurement error would bias the coefficient toward zero, but the measurement error need not be purely random (Kruger and Dunning, 1999). For example, those who are more successful in the labor market for unmeasured reasons may be more likely to overestimate their English language skills. A positive correlation in the measurement error terms could bias the coefficient upward.

A third problem is that there may be dimensions of ability that are not in the equation, but which enhance both English language proficiency and earnings. Those with greater innate ability among the foreign born may have superior English language skills and earn more, even though the higher earnings may be unrelated to their English proficiency. Yet there are no independent measures of ability in these data. This form of omitted variables bias would tend to overstate the true effect of language skills on earnings in an OLS equation.

Instrumental variables is a statistical technique that can, in principle, correct for these potential problems by using a predicted rather than the observed value of language proficiency. An auxiliary regression is computed (Table 17.A6), which includes at least some variables that are not in the earnings function and which has a more complex functional form (various quadratic and interaction terms) to permit statistical identification. This auxiliary regression is used to obtain predicted values of the language variable, and it is these values, rather than the reported or observed values, that are used in the earnings equation. Because the statistical identification is so dependent on variables that vary across countries of origin, a reliable IV model cannot be estimated using these data for immigrants from only one country, Mexico.

The results for the IV earnings function are reported in column (*v*) in Tables 17.A3 and 17.A4 and are summarized in Table 17.2 for all and non-Mexican immigrants. The IV technique results in a very large coefficient for the language proficiency variable. It implies about 80 percent higher earnings for those proficient in English in the all immigrant analysis.[19] Yet, similar very large coefficients on destination language skills have been found elsewhere and for other countries using this technique.[20] Perhaps the unbiased effect of English language fluency on earnings among immigrants is somewhere between the OLS and the IV estimates. Yet, even the OLS estimate of about 15 percent implies a large pay-off from obtaining English language skills.

Summary and conclusion

Summary

This chapter has been concerned with whether immigrant linguistic concentrations or enclaves affect immigrant adjustment in terms of destination language proficiency and earnings.

The reasons for the development of these concentrations are discussed. New immigrants tend to settle near ports of entry, where previous immigrants from their origin (friends and family) have settled and where their employment opportunities are best. The "friends and family" or chain migration effect is a consequence of economies in communication, information, consumption, and the labor market.

"Ethnic goods" are market and nonmarket goods and services consumed by members of an immigrant/ethnic group that are not consumed by others. Ethnic-specific goods are an important factor in location choice. Because of economies of scale in the production of ethnic goods, the full cost of ethnic goods is lower the larger the size of the immigrant/ethnic group. Then an immigrant would be indifferent between working in two alternative areas (equal real wages) only if the area with the high cost ethnic goods (lower concentration ratio) provided a higher nominal wage.

Several hypotheses emerge from the analysis. Linguistic concentrations are expected to have an adverse effect on the destination language proficiency of immigrants. Greater proficiency is expected to result in higher earnings and a larger ethnic/immigrant concentration is expected to have a negative effect on nominal earnings.

The modeling of the language equation is based on three fundamental variables, exposure (pre- and post-immigration) to the destination language, efficiency in destination language acquisition, and economic incentives for destination language acquisition. Variables are developed to measure the effects of these concepts. The concentration ratio and the rural variable measure, in part, postimmigration exposure to the destination language.

The earnings equation is based on the standard human capital earnings function augmented for immigrant adjustment. Two additional variables are the immigrant's proficiency in the destination language and the minority language concentration ratio.

The empirical testing is done using adult (nonaged) male immigrants in the United States from non-English speaking countries as reported in the 1990 Census 5 percent microdata sample. Immigrant language skills are found to vary positively with exposure to the destination language, efficiency in language acquisition, and economic incentives. In particular, English language proficiency is greater the higher the level of schooling, the longer the duration of residence, the younger the age at immigration, the further the origin from the United States, if the origin was a colony of the United States or the United Kingdom, if the immigrant was not a refugee, has a lower probability of return migration, and among immigrants who do not go back and forth between their origin countries and the United States. A smaller minority language concentration ratio and living in a rural area, and hence living among a lower density of origin language speakers, are both associated with greater proficiency in English. Among immigrants from Mexico, greater access to Spanish language radio stations are associated with poorer English language skills.

Annual earnings are found to increase with skill level (schooling, experience, duration in the United States), and weeks worked, and are higher among married men, those living in urban areas outside the south, those who are citizens, and those who are not black. Veteran status is associated with higher earnings among Mexican immigrants, but lower earnings among other immigrants. In the OLS analysis earnings are higher by about 15 percent for those proficient in English, compared to those lacking fluency, and are lower for those living in an area with a higher minority language concentration ratio. The earnings advantage from proficiency is even greater when the respondent's English language proficiency is estimated using the IV technique.

Policy implications

The answer to the question in the title is "yes." Enclaves matter for immigrant adjustment. Immigrant linguistic concentrations are associated with a lower level of proficiency in the destination language (English). Poorer English language skills result in lower nominal and real earnings. Even after controlling for one's own language skills, living within an immigrant/ethnic concentration area also results in lower nominal earnings, presumably because of the ethnic goods effect. Thus, linguistic concentrations have both an indirect effect (via destination language skills) and a direct effect on lowering the observed earnings of immigrants. The direct effect of concentration on earnings may be an equilibrium situation, where earnings differences are compensating differentials reflecting geographic differences in the cost of ethnic goods.

Immigrant/linguistic concentrations serve a useful role. They provide information networks and channels of communication in consumption and in the labor market for those without, or with only limited, destination-specific information and language proficiency, and they lower the cost of ethnic goods. On the other hand, they tend to retard the acquisition of, or investment in, destination-specific skills (e.g., language proficiency) and to lower nominal earnings. The assimilation or adjustment of immigrants is enhanced the smaller the extent of the immigrant/ethnic concentration.

It would be difficult to implement incentives for immigrants to settle outside of concentrated areas for their group. Focusing immigration on countries of origin "culturally similar" to the United States would be an unwarranted return to the pernicious national origins quota system in place from 1921 to 1965. A reduced emphasis on family ties in issuing immigration visas, and placing a greater emphasis on the applicant's own skills is likely to increase the diversity of origins and reduce the extent of immigrant–linguistic concentrations.

Yet, in the highly mobile United States these concentrations tend to be first generation, and at most also second-generation, phenomena. Reliance on self-correcting mechanisms is likely to be the most effective public policy, such as the acquisition of English language skills and the decline in the importance in the market basket of ethnic goods with a longer duration of residence.

Appendix 17A

Table 17.A1 Means and standard deviations of variables, sample used for language model

Variable	*Total Sample*	*Excludes Immigrants from Mexico*	*Immigrants from Mexico Only*
English proficiency	0.730	0.808	0.524
	(0.44)	(0.39)	(0.50)
Age	37.79	40.92	36.83
	(10.63)	(10.75)	(9.69)
Years of education	11.63	13.09	7.80
	(4.99)	(4.27)	(4.69)
Years since migration	15.21	15.43	14.64
	(11.08)	(11.51)	(9.84)
Married	0.655	0.673	0.610
	(0.48)	(0.47)	(0.49)
Rural	0.055	0.042	0.089
	(0.23)	(0.20)	(0.28)
South	0.237	0.234	0.244
	(0.43)	(0.42)	(0.43)
S. Europe	0.078	0.107	[a]
	(0.27)	(0.31)	
E. Europe	0.036	0.049	[a]
	(0.19)	(0.22)	
USSR	0.016	0.022	[a]
	(0.12)	(0.15)	
IndoChina	0.048	0.066	[a]
	(0.21)	(0.25)	

Philippines	0.051 (0.22)	0.070 (0.26)	[a]
China	0.062 (0.24)	0.085 (0.28)	[a]
S. Asia	0.049 (0.21)	0.067 (0.25)	[a]
Other Asia	0.012 (0.11)	0.016 (0.13)	[a]
Korea	0.031 (0.17)	0.044 (0.20)	[a]
Japan	0.015 (0.12)	0.020 (0.14)	[a]
Middle East	0.060 (0.24)	0.083 (0.28)	[a]
Sub-Saharan Africa	0.024 (0.15)	0.034 (0.18)	[a]
Mexico	(0.276) (0.45)	0.00 (0.00)	1.00 (0.00)
Cuba	0.051 (0.22)	0.070 (0.26)	[a]
C. and S. America (Spanish)	0.125 (0.33)	0.173 (0.38)	[a]
C. and S. America (non-Spanish)	0.009 (0.09)	0.012 (0.11)	[a]
Minority language concentration	7.784 (8.87)	3.816 (6.19)	18.178 (5.95)
Linguistic distance	0.515 (0.15)	0.542 (0.17)	[a]
Miles from origin	3841.1 (2574.9)	4756.6 (2475.4)	[a]
Refugee	0.096 (0.29)	0.133 (0.34)	[a]
Colony	0.147 (0.35)	0.203 (0.40)	[a]
Resident overseas 5 years ago	0.019 (0.14)	0.017 (0.13)	0.025 (0.16)
Emigration rate	[a]	2.049 (0.76)	[a]
Spanish radio	[a]	[a]	0.002 (0.01)
Sample size	237,766	169,253	68,512

Notes: Standard errors are in parentheses.
a Variable not applicable.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent sample.

Table 17.A2 Means and standard deviations of variables, sample used for earnings model

Variable	*Total Sample*	*Excludes Immigrants from Mexico*	*Immigrants from Mexico Only*
Natural log of earnings	9.787	9.942	9.387
	(1.03)	(1.04)	(0.90)
English proficiency	0.747	0.830	0.535
	(0.43)	(0.38)	(0.50)
Labor market experience	22.76	22.41	23.63
	(11.46)	(11.53)	(11.25)
Years of education	11.79	13.30	7.90
	(4.92)	(4.11)	(4.68)
Years since migration	15.43	15.75	14.60
	(10.85)	(11.30)	(9.52)
Married	0.673	0.691	0.627
	(0.47)	(0.46)	(0.48)
Rural	0.057	0.044	0.091
	(0.23)	(0.20)	(0.29)
South	0.240	0.238	0.244
	(0.43)	(0.43)	(0.43)
Race (Black)	0.033	0.044	0.004
	(0.18)	(0.21)	(0.06)
Citizen	0.417	0.484	0.247
	(0.49)	(0.50)	(0.43)
Veteran	0.068	0.083	0.030
	(0.25)	(0.28)	(0.17)
Log weeks worked	3.752	3.774	3.693
	(0.47)	(0.46)	(0.51)
S. Europe	0.078	0.108	[a]
	(0.27)	(0.31)	
E. Europe	0.036	0.050	[a]
	(0.19)	(0.22)	
USSR	0.013	0.019	[a]
	(0.12)	(0.14)	
IndoChina	0.041	0.057	[a]
	(0.20)	(0.23)	
Philippines	0.053	0.073	[a]
	(0.22)	(0.26)	
China	0.061	0.085	[a]
	(0.24)	(0.28)	
S. Asia	0.051	0.071	[a]
	(0.22)	(0.26)	
Other Asia	0.011	0.016	[a]
	(0.11)	(0.12)	
Korea	0.031	0.043	[a]
	(0.17)	(0.20)	
Japan	0.015	0.021	[a]
	(0.12)	(0.14)	

Middle East	0.059 (0.24)	0.082 (0.27)	[a]
Sub-Saharan Africa	0.024 (0.15)	0.034 (0.18)	[a]
Mexico	0.279 (0.45)	0.00 (0.00)	1.00 (0.00)
Cuba	0.051 (0.22)	0.069 (0.25)	[a]
C. and S. America (Spanish)	0.127 (0.33)	0.176 (0.38)	[a]
C. and S. America (non-Spanish)	0.009 (0.09)	0.012 (0.11)	[a]
Minority language concentration	7.834 (8.88)	3.850 (6.21)	18.129 (6.00)
Sample size	212,381	150,680	61,700

Notes: Standard errors are in parentheses.
a Variable not applicable.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent sample.

Table 17.A3 Regression estimates of earnings equation, adult foreign-born men from non-English speaking countries, 1990

Variable	*OLS*				*IV*
	(i)	*(ii)*	*(iii)*	*(iv)*	*(v)*
Constant	5.063 (173.18)	5.006 (171.67)	5.074 (173.47)	5.017 (171.96)	4.845 (150.58)
Years of education	0.049 (91.10)	0.045 (83.48)	0.048 (90.24)	0.045 (82.85)	0.035 (35.19)
Experience	0.023 (35.72)	0.025 (38.04)	0.023 (35.77)	0.025 (38.04)	0.029 (38.55)
Experience squared/100	–0.037 (31.33)	–0.038 (32.20)	–0.038 (31.47)	–0.038 (32.30)	–0.041 (33.27)
Years since migration (YSM)	0.028 (49.29)	0.025 (43.73)	0.028 (49.76)	0.025 (44.24)	0.017 (18.36)
YSM squared/100	–0.039 (30.24)	–0.035 (27.56)	–0.039 (30.54)	–0.036 (27.88)	–0.026 (16.85)
Log of weeks worked	0.970 (135.52)	0.964 (134.94)	0.967 (135.43)	0.963 (134.88)	0.952 (131.21)
Married	0.213 (55.22)	0.208 (54.02)	0.214 (55.43)	0.209 (54.23)	0.195 (47.54)
Rural	–0.037 (4.67)	–0.038 (4.89)	–0.043 (5.43)	–0.044 (5.58)	–0.047 (5.89)
South	–0.112 (26.11)	–0.113 (26.36)	–0.109 (25.40)	–0.110 (25.71)	–0.113 (25.90)
Race (Black)	–0.182 (12.36)	–0.190 (12.95)	–0.187 (12.68)	–0.195 (13.22)	–0.218 (14.48)

(*Continued Overleaf*)

Table 17.A3 Continued

Variable	*OLS* (i)	(ii)	(iii)	(iv)	*IV* (v)
Veteran	–0.078 (10.25)	–0.080 (10.48)	–0.079 (10.39)	–0.081 (10.61)	–0.085 (11.12)
Citizen	0.090 (21.36)	0.082 (19.35)	0.088 (20.87)	0.080 (18.94)	0.056 (11.83)
S. Europe	–0.063 (6.23)	–0.060 (5.98)	–0.058 (5.70)	–0.056 (5.51)	–0.049 (4.85)
E. Europe	–0.077 (6.40)	–0.073 (6.09)	–0.077 (6.44)	–0.074 (6.13)	–0.062 (5.14)
USSR	–0.133 (7.37)	–0.125 (6.95)	–0.134 (7.43)	–0.127 (7.02)	–0.103 (5.65)
IndoChina	–0.282 (23.21)	–0.270 (22.31)	–0.283 (23.31)	–0.271 (22.42)	–0.236 (19.02)
Philippines	–0.224 (21.11)	–0.234 (22.07)	–0.217 (20.39)	–0.227 (21.39)	–0.259 (23.42)
China	–0.274 (24.10)	–0.254 (22.41)	–0.270 (23.73)	–0.251 (22.11)	–0.193 (15.84)
S. Asia	–0.021 (1.83)	–0.028 (2.41)	–0.023 (2.00)	–0.029 (2.55)	–0.049 (4.13)
Other Asia	–0.201 (10.45)	–0.203 (10.54)	–0.202 (10.49)	–0.203 (10.57)	–0.208 (10.70)
Korea	–0.233 (14.95)	–0.209 (13.41)	–0.233 (14.94)	–0.209 (13.43)	–0.137 (8.25)
Japan	0.347 (18.75)	0.357 (19.45)	0.347 (18.76)	0.357 (19.44)	0.389 (20.97)
Middle East	–0.098 (8.26)	–0.104 (8.77)	–0.099 (8.36)	–0.105 (8.85)	–0.122 (10.18)
Sub-Saharan Africa	–0.064 (3.38)	–0.070 (3.71)	–0.062 (3.29)	–0.068 (3.62)	–0.087 (4.54)
Mexico	–0.341 (37.39)	–0.313 (34.39)	–0.235 (21.28)	–0.218 (19.80)	–0.167 (14.23)
Cuba	–0.242 (21.54)	–0.216 (19.22)	–0.172 (14.35)	–0.153 (12.77)	–0.095 (7.44)
C. and S. America (Spanish)	–0.244 (25.62)	–0.227 (23.89)	–0.168 (15.93)	–0.158 (15.05)	–0.129 (11.96)
C. and S. America (non-Spanish)	–0.081 (3.61)	–0.100 (4.45)	–0.073 (3.27)	–0.092 (4.12)	–0.150 (6.47)
Proficient in English	[a]	0.151 (32.26)	[a]	0.148 (31.60)	0.592 (16.53)
Minority language concentration	[a]	[a]	–0.0062 (16.75)	–0.0056 (15.25)	–0.0039 (9.62)
$\bar{R}^2$	0.4157	0.4185	0.4164	0.4190	[b]
Sample size	212,381	212,381	212,381	212,381	212,381

Notes: a Variable not entered; b R^2 not defined for the IV Model. IV estimator used for Proficient in English variable. *t*-statistics have been computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent sample.

Table 17.A4 Regression estimates of earnings equation, adult foreign-born men from non-English speaking countries other than Mexico, 1990

Variable	*OLS* (i)	(ii)	(iii)	(iv)	*IV* (v)
Constant	4.824	4.757	4.839	4.773	4.542
	(132.90)	(131.27)	(133.15)	(131.52)	(111.69)
Years of education	0.058	0.055	0.058	0.054	0.041
	(85.04)	(77.84)	(83.96)	(76.98)	(33.96)
Experience	0.023	0.025	0.024	0.025	0.030
	(29.48)	(31.39)	(29.80)	(31.64)	(33.21)
Experience squared/100	–0.037	–0.038	–0.038	–0.039	–0.042
	(24.66)	(25.42)	(25.05)	(25.77)	(26.90)
Years since migration (YSM)	0.026	0.024	0.027	0.024	0.016
	(37.82)	(33.87)	(38.05)	(34.15)	(15.63)
YSM squared/100	–0.037	–0.034	–0.038	–0.035	–0.024
	(24.46)	(22.43)	(24.66)	(22.64)	(13.87)
Log of weeks worked	0.994	0.991	0.994	0.990	0.978
	(108.94)	(108.53)	(108.89)	(108.50)	(105.98)
Married	0.218	0.215	0.217	0.215	0.205
	(44.77)	(44.23)	(44.70)	(44.18)	(40.86)
Rural	–0.002	–0.005	–0.006	–0.009	–0.018
	(0.18)	(0.43)	(0.50)	(0.72)	(1.48)
South	–0.087	–0.088	–0.091	–0.091	–0.093
	(16.22)	(16.38)	(16.85)	(16.95)	(16.90)
Race (Black)	–0.189	–0.197	–0.196	–0.203	–0.228
	(12.25)	(12.77)	(12.65)	(13.13)	(14.40)
Veteran	–0.093	–0.095	–0.094	–0.095	–0.101
	(11.12)	(11.33)	(11.18)	(11.38)	(11.90)
Citizen	0.107	0.098	0.105	0.097	0.069
	(19.99)	(18.41)	(19.63)	(18.12)	(11.85)
S. Europe	–0.028	–0.025	–0.023	–0.020	–0.012
	(2.72)	(2.46)	(2.18)	(1.96)	(1.17)
E. Europe	–0.069	–0.063	–0.071	–0.065	–0.045
	(5.66)	(5.21)	(5.81)	(5.36)	(3.67)
USSR	–0.133	–0.123	–0.136	–0.125	–0.090
	(7.28)	(6.74)	(7.44)	(6.90)	(4.84)
IndoChina	–0.266	–0.251	–0.269	–0.253	–0.199
	(21.30)	(20.11)	(21.50)	(20.32)	(15.19)
Philippines	–0.225	–0.232	–0.217	–0.225	–0.253
	(20.62)	(21.33)	(19.91)	(20.66)	(22.53)
China	–0.274	–0.251	–0.270	–0.248	–0.169
	(23.67)	(21.67)	(23.33)	(21.39)	(12.94)
S. Asia	–0.041	–0.044	–0.043	–0.047	–0.058
	(3.45)	(3.75)	(3.69)	(3.96)	(4.85)

(*Continued Overleaf*)

Table 17.A4 Continued

Variable	*OLS*				*IV*
	(i)	*(ii)*	*(iii)*	*(iv)*	*(v)*
Other Asia	–0.207	–0.206	–0.209	–0.207	–0.203
	(10.68)	(10.63)	(10.77)	(10.71)	(10.36)
Korea	–0.237	–0.208	–0.238	–0.210	–0.112
	(14.94)	(13.17)	(15.01)	(13.26)	(6.40)
Japan	0.339	0.353	0.338	0.352	–0.402
	(18.12)	(18.99)	(18.07)	(18.93)	(21.06)
Middle East	–0.105	–0.108	–0.107	–0.110	–0.121
	(8.72)	(9.01)	(8.89)	(9.16)	(9.99)
Sub-Saharan Africa	–0.071	–0.074	–0.068	–0.071	–0.082
	(3.63)	(3.78)	(3.48)	(3.65)	(4.17)
Cuba	–0.230	–0.203	–0.141	–0.120	–0.051
	(19.99)	(17.56)	(10.42)	(8.93)	(3.48)
C. and S. America (Spanish)	–0.217	–0.197	–0.124	–0.112	–0.067
	(22.14)	(20.07)	(10.21)	(9.17)	(5.25)
C. and S. America (non-Spanish)	–0.060	–0.077	–0.052	–0.069	–0.129
	(2.62)	(3.36)	(2.25)	(3.00)	(5.44)
Proficient in English	[a]	0.154	[a]	0.151	0.678
		(22.82)		(22.40)	(16.40)
Minority language concentration	[a]	[a]	–0.0076	–0.0070	–0.0050
			(12.71)	(11.77)	(7.85)
$\bar{R}^2$	0.3770	0.3792	0.3776	0.3797	[b]
Sample size	150,680	150,680	150,680	150,680	150,680

Notes: a Variable not entered; b R^2 not defined for the IV Model. IV estimator for Proficient in English variable. *t*-statistics have been computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent sample.

Table 17.A5 Regression estimates of earnings equation, adult foreign-born men from Mexico, 1990

Variable	*OLS*			
	(i)	*(ii)*	*(iii)*	*(iv)*
Constant	5.208 (115.30)	5.194 (115.28)	5.279 (114.92)	5.254 (114.70)
Years of education	0.027 (29.94)	0.024 (26.67)	0.027 (29.77)	0.024 (26.58)
Experience	0.015 (12.23)	0.016 (13.35)	0.015 (12.11)	0.016 (13.23)
Experience squared/100	–0.026 (12.74)	–0.026 (12.90)	–0.025 (12.64)	–0.026 (12.81)
Years since migration (YSM)	0.029 (29.47)	0.025 (25.36)	0.029 (29.82)	0.025 (25.70)
YSM squared/100	–0.037 (15.40)	–0.033 (13.69)	–0.037 (15.57)	–0.033 (13.85)
Log of weeks worked	0.918 (82.30)	0.913 (81.83)	0.918 (82.25)	0.913 (81.79)
Married	0.207 (33.54)	0.199 (32.43)	0.208 (33.78)	0.200 (32.65)
Rural	–0.098 (10.58)	–0.099 (10.73)	–0.105 (11.27)	–0.105 (11.29)
South	–0.184 (26.73)	–0.184 (26.81)	–0.174 (24.85)	–0.175 (25.16)
Race (Black)	–0.039 (0.82)	–0.055 (1.16)	–0.038 (0.80)	–0.054 (1.14)
Veteran	0.087 (4.73)	0.078 (4.24)	0.085 (4.63)	0.076 (4.17)
Citizen	0.042 (6.17)	0.028 (4.04)	0.040 (5.86)	0.026 (3.82)
Proficient in English	[a]	0.149 (23.98)	[a]	0.146 (23.52)
Minority language concentration	[a]	[a]	–0.0039 (8.53)	–0.0033 (7.13)
$\bar{R}^2$	0.4080	0.4135	0.4086	0.4139
Sample size	61,700	61,700	61,700	61,700

Notes: a Variable not entered; *t*-statistics have been computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent sample.

Table 17.A6 Regression estimates of language equation used in IV estimation, adult foreign-born men by origin, 1990

Variable	*Total Sample*	*Excludes Immigrants from Mexico*
Constant	0.350	0.398
	(25.24)	(24.06)
Experience	–0.007	–0.007
	(23.10)	(21.09)
Experience squared/100	0.004	0.004
	(6.62)	(6.87)
Years of education	0.022	0.022
	(67.15)	(65.15)
Years since migration (YSM)	0.013	0.012
	(54.17)	(44.62)
YSM squared/100	–0.019	–0.016
	(41.42)	(33.24)
Married	0.028	0.016
	(15.33)	(8.35)
Rural	0.010	0.020
	(3.07)	(6.62)
South	0.005	0.004
	(2.53)	(2.15)
Citizen	0.064	0.055
	(33.65)	(27.59)
Race (Black)	0.068	0.065
	(12.22)	(11.61)
Veteran	0.019	0.012
	(8.46)	(5.17)
Natural logarithm of weeks worked	0.026	0.022
	(14.14)	(10.77)
S. Europe	–0.055	–0.056
	(18.47)	(18.16)
E. Europe	–0.072	–0.087
	(18.72)	(20.88)
USSR	–0.039	–0.053
	(6.22)	(8.22)
IndoChina	–0.156	–0.134
	(15.12)	(12.78)
Philippines	–0.065	–0.038
	(9.10)	(4.72)
China	–0.163	–0.144
	(21.67)	(17.56)
S. Asia	–0.102	–0.072
	(12.31)	(7.79)
Other Asia	–0.120	–0.100
	(10.44)	(8.15)

Korea	–0.242 (25.25)	–0.196 (16.82)
Japan	–0.137 (13.30)	–0.101 (8.42)
Middle East	–0.038 (9.17)	0.008 (0.99)
Sub-Saharan Africa	–0.081 (10.71)	–0.024 (2.04)
Mexico	–0.110 (18.91)	[a]
Cuba	–0.024 (2.66)	–0.008 (0.73)
C. and S. America (Spanish)	–0.057 (11.40)	–0.053 (9.84)
C. and S. America (non-Spanish)	0.106 (13.69)	0.133 (15.99)
Minority language concentration (CON)	0.028 (7.61)	0.022 (5.76)
Linguistic distance	–0.002 (0.16)	0.009 (0.59)
Miles from origin/1,000	0.035 (8.43)	0.047 (10.12)
Square of miles from origin/10 m.	–0.012 (2.85)	–0.023 (5.08)
Refugee	–0.116 (28.94)	–0.113 (27.36)
Colony	0.019 (5.70)	0.022 (6.57)
Resident overseas 5 years ago	–0.066 (10.37)	–0.045 (6.37)
Emigration rate	[a]	–0.033 (7.61)
CON × years of education/1,000	–0.024 (0.84)	0.103 (2.12)
CON × experience/1,000	–0.206 (17.89)	–0.270 (15.59)
CON × YSM/1,000	0.527 (50.63)	0.666 (37.23)
CON × linguistic distance	–0.078 (9.26)	–0.078 (9.11)
CON × miles from origin/1 m	0.181 (1.40)	–0.349 (2.41)
CON × emigration rate	[a]	0.003 (8.00)
$\bar{R}^2$	0.3345	0.3164
Sample Size	212,381	150,680

Notes: a Variable not entered; *t*-statistics have been computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent sample.

Appendix 17.B

Definitions of variables

The variables used in the statistical analyses are defined below. Mnemonic names are also listed where relevant. The means and standard deviations are reported in Appendix Tables 17.A1 and 17.A2 for the samples used in the analyses of language attainment and earnings, respectively.

Data source: 1990 Census of Population, Public Use Microdata Sample, 5 percent sample of the foreign born, except where noted otherwise.

Definition of population: The sample used in this study comprises foreign-born men aged 25 to 64, born in countries other than the English-speaking developed countries (UK, Ireland, Canada, Australia, New Zealand), territories of the United States, at sea, or born abroad of American parents. Those who worked 0 weeks in 1989 were deleted from the analysis of earnings, as they were not labor force participants.

Dependent variables

English language fluency (LANG): LANG is set equal to 1 for individuals who speak only English at home, or if a language other than English is spoken in the home, who speak English either "very well" or "well." The variable is set to 0 where a language other than English is spoken in the home and the respondent speaks English either "not well" or "not at all."

Earnings (LNEARN): The natural logarithm of the sum of wage or salary income and self-employment income (either nonfarm or farm) received in 1989. Individuals with earnings less than $100, including those with negative earnings, were assigned a value of $100.

Explanatory variables

Minority language concentration (CON): Each respondent is assigned a measure equal to the percentage of the population aged 18 to 64 in the state in which he lives, who reports the same non-English language as the respondent. In the construction of this variable, only the 24 largest language groups nationwide are considered. In descending order there are: Spanish; French; German; Italian; Chinese; Tagalog; Polish; Korean; Vietnamese; Japanese; Portuguese; Greek; Arabic; Hindi; Russian; Yiddish; Thai; Persian; French Creole; Armenian; Hebrew; Dutch; Hungarian; Mon-Khmer (Cambodian). These constitute 94 percent of all responses that a language other than English is used at home. Representation in the other language groups is so small numerically that the proportions are approximately zero, and this value is assigned. Those who reported speaking only English are assigned the mean value of the CON measure for other language speakers of their birthplace group.

Location: The two location variables record residence in a rural area (RURAL) or in the Southern States (SOUTH). The states included in the latter are: Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia.

Birthplace (BIRTH): A number of non-English speaking birthplace regions are considered in the analyses: Western Europe; Southern Europe; Eastern Europe; former Soviet Union; Indochina; South Asia (which comprises the regions of British influence, for example, India, Nepal, Pakistan); Other South-East Asia; Korea; Japan; Middle East and North Africa; SubSaharan Africa; Mexico; Cuba; Central and South America (Spanish influence); Central and South America (non-Spanish influence). The benchmark group (omitted category) in the regression analysis is Western Europe.

Colony (COLONY): Countries that are current or former colonies of English-speaking countries are coded one. All other countries are coded zero. Dependencies of the United Kingdom, United States, Australia, New Zealand, and South Africa are coded as colonies under this definition.

Years since migration (YSM). The categorical Census information on year of immigration is converted to a continuous measure using the following values: 1987 to 1990 (1.75 years); 1985 to 1986 (4.25 years); 1982 to 1984 (6.75 years); 1980 to 1981 (9.25 years); 1975 to 1979 (12.75 years); 1970 to 1974 (17.75 years); 1965 to 1969 (22.75 years); 1960 to 1965 (27.75 years); 1950 to 1959 (35.25 years); before 1950 (49.75 years).

Lived abroad 5 years ago (ABROAD5): This dichotomous variable is defined only for immigrants who have resided in the United States for more than 5 years. It is set equal to 1 if the individual lived abroad in 1985, otherwise it is set equal to 0 for immigrants in the United States 5 or fewer years and for longer duration immigrants living in the United States in 1985.

Radio (RADIO): The number of radio stations broadcasting entirely or nearly entirely in Spanish in the state was obtained from *Broadcasting and Cable Yearbook, 1994* (1994), R.R. Bowker, New Providence, NJ, pp. B566–B567. In 1994, there were 315 Spanish language radio stations broadcasting in 25 states. Chiswick and Miller (1998) present details. The number of Spanish-language radio stations in the state was normalized by the area of the state to give the number of radio stations per 1,000 square miles. Then this variable was normalized by the number of Spanish speakers in the state of residence to give the number of Spanish language radio stations per unit of area per 10,000 Spanish speakers. This variable provides an index of the intensity of the infrastructure supporting the Spanish language in the state of residence. There were too few radio stations broadcasting in languages other than Spanish to compute a meaningful index for other languages. Because of the possible endogeneity of this variable, an IV approach was used.

Marital status (MARR): This is a binary variable that distinguishes individuals who are married, spouse present (equal to 1) from all other marital states.

Years of education (EDUC): This variable records the total years of full-time education. It has been constructed from the Census data on educational attainment by assigning the following values to the Census categories: completed less than fifth grade (2.5 years); completed fifth through eighth grade (7 years); completed ninth grade (9); completed tenth grade (10); completed 11th grade (11); completed 12th grade or high school (12); attended or completed college (14); Bachelor's degree (16); Master's degree (17.5); Professional degree (18); Doctorate (20).

Refugee (REFUGEE): This variable is constructed to identify the major sources of post-WWII refugees to the United States. It is defined only for immigrants who migrated at age 25 and older. Individuals who migrated from Cambodia, Laos, or Vietnam in 1975 or later, Iran in 1980 or later, Cuba in 1960 or later, or the USSR and Baltic States are assigned a value of one for this variable. All other immigrants are assigned a value of zero.

Linguistic distance (DISTANCE): This is a measure of the difficulty of learning a foreign language for English-speaking Americans. It is based on a set of language scores (LS) measuring achievements in speaking proficiency in foreign languages by English-speaking Americans at the U.S. Department of State, School of Language Studies, reported by Hart-Gonzalez and Lindemann (1993). It is described in detail in Chiswick and Miller (1998, Appendix B). For the same number of weeks of instruction, a lower score (LS) represents less language facility, and, it is assumed, greater linguistic distance between English and the specific foreign language. For example, French is scored at 2.5 (in a range from 1 to 3), while Japanese is scored at 1.0. The language groups reported in the Hart-Gonzalez and Lindemann (1993) study are then matched to language codes in the 1990 Census using the Ethnologue Language Family Index published by Grimes and Grimes (1993). Adam Makkai, Professor of Linguistics, University of Illinois at Chicago, assisted in the matching of language codes, and in expanding the list of languages for which scores were assigned.

In the construction of this variable, foreign-born persons who speak only English at home and hence do not report speaking a non-English language are assigned the mean value of the linguistic score measure for individuals reporting a foreign language from their birthplace group.

The variable in the regression equations is linguistic distance, which is 1 divided by the linguistic score, DISTANCE = 1/LS.

Emigration rate (EMIG): Yearly emigration rates of the foreign born by country of birth and sex are computed by dividing the yearly emigration levels between 1980 and 1990 from Ahmed and Robinson (1994) by the number of immigrants of the specific birthplace-gender group in 1980 from the

1980 U.S. Census. Thirty-three countries are separately identified in the data, together with seven residual regions.

Direct-line distances (MILES): The miles between the major city in the immigrant's country of origin and the nearest large port of entry in the United States (New York, Miami, Los Angeles) are constructed from data in Fitzpatrick and Modlin's (1986) *Direct Line Distances, United States Edition*.

Years of experience (EXP): This is computed as age minus years of education minus 5 (that is, EXP = AGE – EDUC – 5). A quadratic specification is used.

Log of weeks worked (LNWW): The number of weeks worked in 1989 is used in natural logarithmic form.

Race: This is a dichotomous variable, set to 1 if the individual is Black, and set to 0 for all other racial groups (White, Asian, and Pacific Islander groups, American Indian, other groups).

Veteran status (VETSTAT): This is a dichotomous variable, set to 1 where the respondent is a veteran of the U.S. armed forces. In all other cases it is set to 0.

Citizen (CITIZEN): This is a dichotomous variable, set to 1 for individuals who are naturalized citizens.

Acknowledgements

Earlier versions of this chapter were presented at the Conference on Magnet Societies, Loccum, Germany, June 2000, the Annual Meeting of the European Society for Population Economics, June 2000, Bonn, Germany, the Joint Center for Poverty Research, Northwestern University, February 2001, the Education and Employment Economics Group (UK) Annual Meeting, Leicester, July 2001, the Annual Meeting of the American Economic Association, Atlanta, January 2002, the Center for Economic Policy Research Conference on Discrimination and Unequal Outcomes, Le Mans, France, January 2002, the Gideon Fishelson Lecture at Tel Aviv University, January 2003, the Institute of Government and Public Affairs, University of Illinois at Chicago, October 2003, conference Trading Borders: Migration, Ethnicity and Incorporation in the Age of Globalism, New York, October 2003, and Department of Economics seminars at George Washington University (November 2001) and Hebrew University, Jerusalem (December 2001). Comments received from participants at these presentations are appreciated although all errors of omission and commission are solely those of the authors.

Chiswick acknowledges the research support of the Institute of Government and Public Affairs, University of Illinois, and Miller acknowledges the support of an Australian Research Council grant. This chapter received the Milken Institute Award for Distinguished Research in Regional and

Demographic Studies, 2001, and an earlier version with the same title was published in Hebrew in *Revon L'Kalcalah* (Vol. 50, No. 2, June 2003).

Notes

1 Other work on the determinants of immigrant or ethnic concentrations include Bartel (1989), Brettell (2003), Cutler and Glaeser (1977), Lazear (1999), Bauer, Epstein, and Gang (2002), and Sierminska (2002). Lazear (1999, p. S99) describes concentrations as forming "in large part because doing so enhances trade" in market and non-market goods and services.

2 Epstein (2003) distinguishes theoretically between "herd behavior" and "network externalities" in the choice of destination among those from the same origin. Herd behavior refers to following those from the same origin, even if they are few in number, under the belief that they have better information, while "network externalities" implies a larger group and a lower cost of settlement in a specific destination because of linguistic and information networks.

3 Ross (2002) develops a model in which preferences for social interaction by the majority or a minority (whether negative as in prejudice or positive as in cultural affinity) result in social segregation of neighborhoods.

4 For research on network externalities see Economides (1996) and Katz and Shapiro (1985).

5 Distinctiveness is important as the ethnic goods of English immigrants to the United States would be much less distinctive than would those of, say, Chinese immigrants. To some extent the cost of ethnic goods can be reduced if the host society "adopts" the ethnic good, as, for example, often happens for certain foods, such as in Chinese restaurants. The "Americanized" version of the ethnic good may well differ from the version consumed in the origin or by members of the ethnic group in the destination.

6 For a study of consumer network markets and group size, see Etziony and Weiss (2001).

7 For a discussion of Chinese schools, see Zhou and Li (2003).

8 Workers of a given level of skill can be thought of as randomly drawing wage offers from a given distribution of wage offers available in the high-concentration and the low-concentration areas. If ethnic goods are an important part of their market basket, the ethnic immigrants will move to or stay in a low-concentration area only if their wage offer in this area exceeds by a sufficient margin the wage offer from the high-concentration area to compensate for the higher cost of living. Once settled in a specific area explicit and implicit location-specific investments in human capital, relevant for consumption and the labor market, tend to reduce subsequent migrations. Thus, those who leave a high-concentration enclave for a low- or zero-concentration area will tend to be those who receive a high wage offer in the latter location and those for whom ethnic goods (including ties to the ethnic community) are least important.

9 The "ethnic goods" concept and its implications for concentrations and wage differentials can be applied to other affinity groups, for example, the gay population.

10 In the data under study for earnings, Mexican immigrants are 29 percent of the sample and have a mean schooling level of 7.9 years, in contrast to 13.3 years for the other immigrants.

11 The definition of the population under study and the variables used in the analysis are described in more detail in Appendix 17.B.

12 Bertrand, Luttmer, and Mullainathan (2002) also use language as the basis for their "networks" (concentrations) in an analysis of welfare participation.

13 The finding that a higher level of secular schooling is associated with greater proficiency in Hebrew among immigrants in Israel suggests that exposure to English in school prior to immigration is not the primary mechanism for the positive effect of schooling on English language skills in the United States (Chiswick and Repetto, 2001).

14 On the other hand, the effect of having been in the U.S. Armed Forces differs sharply between these two groups. Veteran status is associated with about 8 percent higher earnings for Mexican immigrants but 10 percent lower earnings for non-Mexican immigrants.

15 Clark and Drinkwater (2002) find that unemployment rates for racial and ethnic minorities are higher among those living in ethnically concentrated areas of England and Wales.

16 The labor supply or "crowding" hypothesis would imply a larger coefficient on the concentration measure for Mexican immigrants than for the much more heterogeneous group of immigrants from other countries. That the opposite is found suggests that the negative relation between concentration and earnings is not a consequence of ethnic crowding in the labor market.

17 Partial effects of education and the concentration ratio on earnings:

	Table 17.A3 Column (iv)	*Table 17.A3, Column (iv), Plus Interaction*
Education	0.045 (82.9)	0.056 (78.3)
Minority Language Concentration	–0.0056 (15.3)	0.0062 (10.9)
Education-Concentration Interaction	—	–0.0012 (26.9)

18 For a similar finding for Sweden, see Edin, Fredriksson, and Aslund (2002).

19 The regression coefficient is $ln(1 + X) = 0.59$, where X is the percentage increase in earnings. X is then 0.80 or 80 percent. $ln(1 + X)$ is approximately equal to X when X is a small number. When $ln(1 + X) = 0.15$, X is approximately 16 percent.

20 See Chiswick and Miller (1995), and the references therein, for the United States, Canada, Australia, and Israel, and Dustmann and van Soest (2001) for Germany. The difference between the OLS and IV effects on earnings is much smaller in the United Kingdom (Dustmann and Fabbri, 2000).

References

Ahmed, B., and Robinson, J. G. 1994. *Estimates of Emigration of the Foreign-Born Population, 1980–1990*. Technical Working Paper No. 9. Washington, DC: U.S. Bureau of the Census.

Bartel, A. 1989. "Where Do the New Immigrants Live?" *Journal of Labor Economics* 7(4), 371–391.

Bauer, T., Epstein, G.S., and Gang, I. N. 2002. *Enclaves, Language and the Location Choice of Migrants*. Institute for the Study of Labor, Bonn, Xerox.

Bertrand, M., Luttmer, E. F. P., and Mullainathan, S. 2002. "Network Effects and Welfare Cultures," *Quarterly Journal of Economics* August, 1019–1055.

Brettell, C. 2003. "Meet Me at the Chat Corner: The Embeddedness of Immigrant Entrepreneurs," Department of Anthropology, Southern Methodist University, Dallas.

Broadcasting and Cable Yearbook, 1994. 1994. New Providence. NJ: R.R. Bowker Publishers.

Chiswick, B. R. 1978. "The Effect of Americanization on the Earnings of Foreign-Born Men," *Journal of Political Economy* 86(5), 897–922.

Chiswick, B. R. 1999. "Are Immigrants Favorably Self-Selected?" *American Economic Review*, 89(2), 181–185.

Chiswick, B. R., and Hatton, T. 2003. "International Migration and the Integration of Labor Markets," in M. D. Bordo, A. M. Taylor, and J. Williamson (eds.), *Globalization in Historical Perspective*, pp. 65–119. Cambridge: National Bureau of Economic Research.

Chiswick, B. R., and Miller, P. W. 1992. "Language in the Immigrant Labor Market" in B. R. Chiswick (ed.), *Immigration, Language and Ethnicity: Canada and the United States*, pp. 229–296. Washington: American Enterprise Institute.

Chiswick, B. R., and Miller, P. W. 1995. "The Endogeneity Between Language and Earnings: International Analyses," *Journal of Labor Economics* 13(2), 246–288.

Chiswick, B. R., and Miller, P. W. 1998. "English Language Fluency Among Immigrants in the United States," *Research in Labor Economics* 17, 151–200.

Chiswick, B. R., and Miller, P. W. 2003. "The Complementarity of Language and Other Human Capital: Immigrant Earnings in Canada," *Economics of Education Review* 22, 469–480.

Chiswick, B. R., and Repetto, G. 2001. "Immigrant Adjustment in Israel: Literacy and Fluency in Hebrew and Earnings," in S. Djajic (ed.), *International Migration: Trends, Policy and Economic Impact*, pp. 204–228. New York: Routledge.

Clark, K., and Drinkwater, S. 2002. "Enclaves, Neighbourhood Effects and Employment Outcomes: Ethnic Minorities in England and Wales," *Journal of Population Economics* 15, 5–29.

Cutler, D. M., and Glaeser, E. L. 1997. "Are Ghettos Good or Bad?" *Quarterly Journal of Economics* 112, 827–872.

Dustmann, C., and Fabbri, F. 2000. "Language Proficiency and Labor Market Performance in the UK," University College London, London, Xerox.

Dustmann, C., and van Soest, A. 2001. "Language Fluency and Earnings: Estimation with Misclassified Language Indicators," *Review of Economics and Statistics* 83(4), 663–674.

Economides, N. 1996. "The Economics of Networks," *International Journal of Industrial Organization* 14, 673–699.

Edin, P.-A., Fredriksson, P., and Aslund, O. 2002. "Ethnic Enclaves and the Economic Success of Immigrants—Evidence from a Natural Experiment," Department of Economics, Uppsala University, Sweden.

Epstein, G. S. 2003. "Informational Cascades and the Decision to Migrate," Department of Economics, Bar Ilan University.

Etziony, A., and Weiss, A. 2001. "Coordination and Critical Mass in a Network Market: An Experimental Evaluation," Department of Economics, Bar Ilan University, October, Xerox.

Fitzpatrick, G. L., and Modlin, M. J. 1986. *Direct-line Distances: United States Edition*. Metuchen, NJ: The Scarecrow Press, Inc.

Grimes, J. E., and Grimes, B. F. 1993. *Ethnologue: Languages of the World*, (13th ed.). Dallas: Summer Institute of Linguistics.

Hart-Gonzalez, L., and Lindemann, S. 1993. "Expected Achievement in Speaking Proficiency, 1993," School of Language Studies, Foreign Service Institute, U.S. Department of State, Mimeographed.

Katz, M., and Shapiro, C. 1985. "Network Externalities, Competition and Compatibility," *American Economic Review* 75, 424–440.

Kruger, J., and Dunning, D. 1999. "Unskilled and Unaware of it: How Difficulties in Recognizing One's Own Competence Lead to Inflated Self-Assessments," *Journal of Personality and Social Psychology* 77(6), 1121–1134.

Lazear, E. P. 1999. "Culture and Language," *Journal of Political Economy* 107 (6, part 2), pp. S95–S126.

Ross, S. 2002. "Segregation and Racial Preferences: An Analysis of Choice Based on Satisfaction and Outcome Measures," Paper presented at the conference on Discrimination and Unequal Outcomes, Le Mans, France.

Sierminska, E. 2002. "Immigrants and State Clustering: Effect of Welfare Benefits," Department of Economics, Johns Hopkins University, October, Xerox.

White, H. 1980. "A Heteroskedastic Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48, 817–838.

Zhou, M., and Li, Y.-Y. 2003. "Ethnic Language Schools and the Development of Supplementary Education in an Immigrant Chinese Community in the U.S.," Department of Sociology, UCLA.

18 Ethnic networks and language proficiency among immigrants

I. Introduction

Immigrants in the main immigrant receiving countries (such as Australia, Canada, Germany, Israel and the United States) acquire fluency in the dominant language of the destination with the passage of time, and the possession of these skills enhances prospects for economic success (Chiswick and Miller 1992, 1995). Consequently, knowledge of the process through which the language skills are acquired is important for understanding immigrant well-being. Much of the research into language proficiency has focussed on individual characteristics that affect language choice and proficiency, for example, educational attainment, duration of residence and exposure to the dominant language prior to migration. It has been shown, for example, that destination language fluency rates are higher among the better educated, among those in the country for a longer period of time, and among those exposed to the language prior to migration. There are also systematic effects by marital status, family composition, country of origin, and motive for migrating. Estimation of language choice and proficiency models using different data sets for the same country, and using similar specifications for different countries, reveals that these individual effects are remarkably robust (Chiswick and Miller 1992, 1994a, 1995).

At the same time, researchers have been aware that group effects may alter the costs of, and benefits to, language acquisition, and thus affect fluency rates. Evans (1986, p. 243), for example, outlines but does not test three hypotheses concerning group differences. She suggests that members of a large immigrant community, or of an immigrant group that has large enclave markets, or a more inward-looking perspective will have less economic incentive to acquire dominant language fluency. Similarly, Chiswick (1991, p. 156) suggests that for inhabitants of the Los Angeles area, the presence of a sizable, Spanish-speaking community could reduce the incentive for Hispanic immigrants to acquire English-language skills. Clyne (1991, p. 88) notes that in Australia ". . . the rate of language maintenance/shift varies in accordance with the relative size of the community in a particular state (or capital city)." Veltman (1983, p. 215), on the basis of an examination of regional differences in the rates of acquiring English fluency in the United States, claims that "The existence of this basic pool of people usually speaking the minority language may have a braking effect on the anglicization of immigrants." Dustmann (1994) attributes the lower fluency of Turkish immigrants in Germany to their being the most numerous non-Germanic immigrant group and their reduced exposure to German because of linguistic enclaves. In these studies, however, the differences in dominant language fluency rates reported by linguistic origin were not systematically related to any single group identifier, suggesting that they may reflect regional economic differences or other unmeasured variables rather than group linguistic effects per se.

A more promising line of inquiry into the identification of group linguistic effects is contained in Chiswick and Miller (1992, 1995). In this research into variations in dominant language fluency rates among immigrants, a variable for the fraction of the regional population that, whether native or foreign born, speaks the same origin-language as the respondent is included in the micro-level equation. Comparable statistical analyses undertaken for Australia, Canada, Israel and the United States reveal that the minority-language concentration measure is an important determinant of dominant language fluency. Immigrants living in regions that have a relatively high representation of their language group are less likely to be fluent in the dominant language, other things being the same. The statistically significant effect is strongest in Australia and about the same in Canada, Israel and the United States. The effect of the minority-language concentration measure persists even when analyses are done within countries of origin for each of these four major destinations.

The minority-language concentration measure may well be a proxy for an array of variables not available in the census data studied to date that are associated with greater group density. For example, to what extent is the concentration effect reflecting the intensity of interaction with friends or social/ethnic networks? Group differences in the effect on language fluency of simple measures of linguistic concentration may well reflect differences in these more fundamental variables.

The primary focus of this chapter is to explore this issue with a unique data set that can be used to study various dimensions of the linguistic environment in which immigrants live. A second purpose is to exploit more fully information on three dimensions of language proficiency, that is, reading, writing and speaking, that are available in these data. Most census and survey data mechanically limit themselves to one dimension, speaking fluency, yet this may not be the most salient characteristic. Indeed, some recent research on special samples of immigrants in the United States, Israel and Germany suggest the greater importance of literacy rather than fluency in the destination language in an analysis of earnings (see, Beenstock 1993; Chiswick 1991; Dustmann 1994).

Section II introduces the data from the survey *Issues in Multicultural Australia 1988*, while Section III briefly describes the estimating equation.

Section IV presents estimates of models of language skill that consider proficiency in speaking, reading and writing. The relationship between a group minority-language concentration variable and various ethnic variables as determinants of language skills is decomposed using a method outlined in Johnson and Solon (1986). A summary and conclusion are contained in Section V.

II. The data

The analyses presented below are based on the survey *Issues in Multicultural Australia 1988* undertaken on behalf of the Office of Multicultural Affairs in the Department of the Prime Minister and Cabinet in the Federal Government of Australia. This survey contains detailed information on language use and the interaction of the individual with the environment that appear to be unique among contemporary data sets. In particular, in addition to the demographic and skill-level data collected in most surveys (e.g., age, years since migration, and educational attainment), information was collected on English-language reading, writing and speaking skills, on the number of the respondent's relatives in Australia, on ethnic identification and on affiliation with ethnic/social clubs.

In the statistical analyses reported below respondents are coded as being proficient in spoken English (SPEAK) if English is the first language spoken at home or, if a language other than English is the first language spoken at home but the respondents' spoken English is "very good." The respondent is not fluent if English is not the first language spoken at home and if the person's speaking skills are only "good," "fair" or "poor". Proficiency in reading (READ) and writing (WRITE) are defined in a similar manner.

The *Issues in Multicultural Australia 1988* data set comprises four separate independent random samples: the Australian population aged 15 years and over (1,552 cases); second-generation Australians (823 cases); immigrants from non-English-speaking countries (986 cases); and immigrants who had arrived in Australia from July 1, 1981 until the date of the survey (1,141 cases). These four samples provide 4,502 observations, of whom 2,532 were born overseas. Given the focus on English-language fluency, this study is restricted to immigrants from non-English-speaking countries. The study covers the language proficiencies of adults age 15 to 64. This is a slightly wider age group than usual in studies using the large samples in Census data (Evans (1986): ages 20 to 64, Chiswick and Miller (1995): ages 25 to 64). The country of origin and age restrictions and the deletion of the few observation with missing data result in a sample with 2,032 observations.

The sampling frames of the four samples collected in the *Issues in Multicultural Australia 1988* survey differ, and the ideal approach is to conduct separate analyses for each sample (see, for example, Kee 1989). The small sample sizes for immigrants, however, suggests that a more practical alternative is to pool the data from the three relevant surveys (see, for example, Chapman and Iredale 1993). Models estimated from the pooled data were checked against similar models estimated using a random sample from the 1986 Australian Census of Population and Housing (see Appendix Table 18.A2). The qualitative conclusions that can be drawn from the parallel analyses are the same and the magnitudes of many of the coefficients are very similar, particularly with respect to the minority-language concentration variable (–0.060 in the census analysis and –0.058 in the 1988 survey analysis). For all practical purposes, therefore, the pooled survey data can be viewed as having been drawn from the same statistical population as the Census sample.

Following Chiswick and Miller (1994b), both males and females are included in the analysis. The means and standard deviations of the variables used in the analysis are reported in Appendix Table 18.A1.

III. A model of dominant language fluency

Language use and proficiency may be analyzed within a human capital framework (see, for example, Breton 1978a, b; Chiswick and Miller 1992, 1995). Within this framework, attention is focussed on the factors that affect the costs of, and returns from, the investment in human capital. Chiswick and Miller (1995) discuss these in terms of three fundamental variables: exposure to the language, efficiency in second language acquisition, and economic benefits from language fluency. Exposure to the language has three components: exposure prior to immigration, time units of exposure in the destination, and the intensity of exposure per unit of time in the destination. Both the characteristics of the individual and of the environment in which an individual lives may affect these variables.

The variables included in the estimating equation are discussed in depth in Chiswick and Miller (1995), and their empirical counterparts employed in the current analysis are defined in Appendix 18B. In particular, it has been found for Australia, as well as for the United States, Canada, Israel and Germany, that destination language fluency increases with duration of residence and level of education and decreases with age at immigration. There are systematic differences by country of origin, with refugees and sojourner migrants having the lowest fluency and those with greater pre-immigration exposure to the destination language having the highest level of fluency among the immigrants.

The minority group language concentration variable provides a measure of the intensity of exposure in the destination country to the language of the origin through the medium of linguistic enclaves (Chiswick and Miller 1992, 1995). Where there is a large concentration of minority-language speakers, linguistic enclaves may form in which the language of the country of origin is used, at least in part, as a language of the home, at work and in the community. This would lower the benefits and increase the costs of learning the language of the destination. For example, respondents dwelling in areas with high minority-language concentration ratios may have a relatively large number of family members near by, or may limit their interactions to ethnic clubs or organizations or ethnic media which require a critical mass to function.

Reflecting this situation, the language equation estimated in previous studies and outlined above is augmented in this study with variables for ethnic affiliation, social/ethnic club membership, ethnic media and relatives in the destination. The construction of these variables is described in Appendix 18B. Including variables for these linguistic-ethnic influences in the analysis will facilitate an assessment of the determinants of the minority-language enclave effect.

Three foreign media variables were constructed from independent sources for use in the analysis and add to the data file. The first is the number of Australian newspapers published in the language other than English reported by the respondent. There are no newspapers published in many community languages with small representation in Australia, whereas there are 15 published in Greek, 14 in Turkish and 10 in Italian. The second variable recorded the weekly hours (h) of broadcast time in each community language on the eleven radio stations with the most hours of non-English broadcasts. This variable ranges in value from 0 h for small language groups to 68 h for Greek, 52 h for Italian and 33 h for Spanish. The third variable is the annual number of hours of transmission in origin languages on the government-run Special Broadcasting Service multicultural TV channel. This variable ranged in value from zero for many languages to high values for Italian (256 h), German (197 h), and French (137 h).

In addition, a variable that records whether English was used in the home when the respondent was 14 years of age is included in the model. The primary role of this variable is to control for initial language skills. It is not possible to do this in most data sets, although a notable exception is Chiswick (1991) where statistical controls for English-language skills at the time of migration to the United States are entered in the language fluency model.

IV. Estimation and decomposition

Models of English speaking (SPEAK), reading (READ) and writing (WRITE) skills are presented in this section (Table 18.1). All estimates reported in Table 18.1 are obtained using OLS. There are two well-known problems with this method of estimation when the dependent variable is dichotomous: the residuals are heteroscedastic and the predicted values may lie outside the range of the unit interval. Re-estimation of the models using a logit estimation procedure does not lead to any changes in the substantive findings of the analysis (see Appendix Table 18.A3). Accordingly, because the OLS estimates are easier to interpret (and consistent standard errors may be computed following White 1980) and are amenable to analysis using the

Table 18.1 Regression estimates of English-language fluency among adult immigrants from non-English-speaking countries, Australia, 1988 (dependent variables: SPEAK, READ, WRITE)

Variable	*SPEAK*		*READ*		*WRITE*	
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Constant	0.168	0.325	0.117	0.280	0.097	0.266
	(2.90)	(5.24)	(2.00)	(4.49)	(1.66)	(4.25)
Age	–0.004	–0.004	–0.004	–0.004	–0.004	–0.003
	(4.55)	(4.08)	(4.40)	(4.00)	(3.88)	(3.49)
Education	0.017	0.016	0.027	0.026	0.024	0.023
	(7.05)	(6.84)	(10.88)	(10.74)	(9.87)	(9.82)
Years since	0.025	0.025	0.022	0.022	0.021	0.021
migration (YSM)	(9.72)	(9.85)	(8.42)	(8.62)	(8.30)	(8.52)
YSM squared/100	–0.024	–0.028	–0.018	–0.022	–0.017	–0.021
	(4.05)	(4.97)	(3.04)	(3.95)	(2.85)	(3.79)
Female	–0.026	–0.024	–0.041	–0.040	–0.017	–0.015
	(1.51)	(1.47)	(2.40)	(2.39)	(0.99)	(0.93)
Married	–0.075	0.060	–0.078	0.058	–0.073	0.065
	(3.69)	(2.57)	(3.86)	(2.57)	(3.53)	(2.77)
English at age 14	0.225	0.191	0.205	0.172	0.228	0.194
	(9.58)	(8.26)	(9.14)	(7.81)	(9.83)	(8.44)
Northern Europe	0.248	0.150	0.229	0.132	0.217	0.119
	(7.97)	(4.83)	(7.45)	(4.25)	(6.79)	(3.75)
Eastern Europe	–0.014	–0.023	–0.023	–0.034	–0.078	–0.084
	(0.33)	(0.55)	(0.51)	(0.80)	(1.79)	(2.01)
Arabian countries	–0.008	0.035	0.006	0.049	–0.014	0.034
	(0.20)	(0.88)	(0.15)	(1.21)	(0.34)	(0.84)
South Asia	0.213	0.070	0.207	0.067	0.206	0.062
	(5.73)	(1.84)	(5.52)	(1.74)	(5.49)	(1.61)
Philippines	0.360	0.241	0.406	0.291	0.371	0.248
	(6.54)	(4.61)	(8.69)	(6.36)	(7.25)	(5.07)
Vietnam	–0.190	–0.117	–0.204	–0.134	–0.197	–0.117
	(6.08)	(3.37)	(6.24)	(3.72)	(6.13)	(3.31)
Other Asia	–0.042	–0.141	–0.082	–0.179	–0.075	–0.176
	(1.18)	(3.90)	(2.25)	(4.88)	(2.11)	(4.90)
South and Central	–0.040	–0.026	–0.077	–0.066	–0.070	–0.049
America	(0.70)	(0.47)	(1.30)	(1.17)	(1.21)	(0.90)
Africa	0.243	0.047	0.240	0.044	0.239	0.039
	(4.02)	(0.76)	(3.91)	(0.73)	(3.95)	(0.64)
Other countries	0.323	0.143	0.317	0.141	0.316	0.134
	(6.71)	(2.94)	(6.79)	(3.01)	(6.68)	(2.82)
Minority-language	–0.050	0.003	–0.046	0.005	–0.052	0.002
concentration	(5.17)	(0.23)	(4.90)	(0.42)	(5.41)	(0.19)
Foreign marriage	[a]	–0.222	[a]	–0.223	[a]	–0.224
		(9.00)		(9.20)		(9.16)
Family	[a]	–0.046	[a]	–0.057	[a]	–0.058
		(2.32)		(2.86)		(2.94)

(*Continued Overleaf*)

Table 18.1 Continued

Variable	*SPEAK*		*READ*		*WRITE*	
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Ethnic press	(a)	–0.041 (4.72)	(a)	–0.039 (4.42)	(a)	–0.045 (5.19)
Ethnic press squared	(a)	0.002 (3.22)	(a)	0.002 (2.99)	(a)	0.002 (3.75)
$\bar{R}^2$	0.4268	0.4658	0.4297	0.4690	0.4327	0.4746
Sample size	*2032*	*2032*	*2032*	*2032*	*2032*	*2032*

Notes
a = variable not entered.
The benchmark country category is Southern Europe.
'*t*' statistics in parentheses computed using White's (1980) heteroscedasticity-consistent covariance matrix estimator.

Source: Issues in Multicultural Australia Survey 1988.

methodology of Johnson and Solon (1986), priority is given to the OLS results.

It should be noted that the survey design means that the analyses of the three dependent variables are not independent: all individuals who report English as the first language spoken at home are classified as fluent in each of these skills. As these account for between 80 and 85% of the fluent group, the analyses should reveal similar patterns.

The analyses are presented for a pooled sample of males and females. F-tests of structural differences between the determinants of language fluency for males and females were conducted for each of the six models listed in Table 18.1. There was evidence that the models for males and females were not the same in only two cases (equations (1) and (4)), and even then the test statistics were at the margin of significance. Separate analyses for males and females are therefore not reported. This contrasts with the statistically significant difference in the structure by gender reported in Chiswick and Miller (1994a) for Australia using the Census. The difference is most likely associated with the much larger sample sizes available in the Census.

Language speaking, reading and writing skills are markedly lower in regions where there is a relatively high concentration of minority-language speakers. The coefficient on the minority-language concentration variable in a simple (one variable) linear regression is –0.088 (t = 7.96) for SPEAK, –0.093 (t = 8.47) for READ, and –0.095 (t = 8.61) for WRITE. That is, in a region where an additional one percent of the population speaks the same minority language as the respondent, the fluency rate is expected to be around 9 percentage points lower than otherwise.

For the reasons noted previously, this group effect in a simple regression will undoubtedly capture some unmeasured individual effects (omitted variables bias). The nature of the relationship can be determined in a straightforward manner. Let β_s be the estimated coefficient on the minority-language concentration variable obtained from a simple regression with SPEAK, READ or WRITE as the dependent variable, and β_m be the estimated coefficient on this variable from a multiple regression (with the same dependent variable) that includes I control variables for individual characteristics.

Then:

$$\beta_m - \beta_s = -\sum_{i=1}^{I} \beta_i b_{ic},$$

where β_i is the estimated coefficient on the ith control variable, and b_{ic} is the coefficient from a simple regression of the ith control variable on the minority-language concentration measure (see Johnson and Solon 1986).

As a first step in implementing this approach, multiple regression equations are estimated that include both the group and individual effects. Two specifications of the estimating equation are presented in Table 18.1 for each of the fluency measures. The first equation includes variables that are thought less likely to be closely related to the minority-language concentration variable. The second specification includes "foreign marriage," "family" and "ethnic media" variables that are alternative dimensions of linguistic enclaves.

The three media variables for access to foreign language newspapers and radio and TV broadcasts are highly positively inter-correlated. When all three are entered in the estimating equation at the same time, the ethnic press variable is statistically significant, whereas the ethnic radio and ethnic TV variables are not. Only the ethnic press variable is therefore included in the specifications reported below. To allow the impact of the ethnic press variable to vary with the number of newspapers available, a quadratic form is used.

Replacing the ethnic press variable by the ethnic radio variable results in no material change to the results. However, the ethnic TV variable is not statistically significant in an equation that includes the minority-language concentration variable. Broadcast time on the Special Broadcasting Service TV channel is allocated on the basis of the number of speakers of various languages *and* the availability of suitable films (Clyne 1991, p. 149). Presumably it is the second of these conditions that accounts for the different performance of the ethnic TV variable.

A variable recording membership in a social/sports club that was identified with a particular ethnic group was insignificant in each equation, as was a variable that recorded whether the respondent identified with a non-Australian ethnic or cultural group. The latter variable was marginally significant when the ethnic media variable was omitted. These variables were not included in the specifications reported in this paper.

Focussing first on the equation for speaking skills, the variable age (or age at migration given that years since migration are held constant) is negatively related to speaking fluency (Table 18.1, column 1). Each additional year of age leads to 0.4 of one percentage point lower fluency. This effect is about one-half of that reported from analyses of 1981 and 1986 Australian Census data (Chiswick and Miller 1995), but the discrepancy appears to be associated with the inclusion in the current analysis of a variable for language used at age 14 (see, for example, the results in Appendix Table 18.A2).

Language fluency rises with educational attainment, the estimated effects indicating 1.7 percentage points higher fluency per additional year of education. When education is divided into pre- and post-migration schooling the coefficients are the same. Hence, the strong finding of complementarity of the various types of human capital skills among immigrants reported in Chiswick and Miller (1995) carries over to the current analysis.

There is a curvi-linear relationship between English-language proficiency and years since migration, with language skills increasing at a decreasing rate with years of residence in Australia. Evaluated at a duration of 10 years, the partial effect of English-language proficiency with respect to duration of residence is 2.0 percentage points. This declines to 1.6 percentage points after 20 years of residence in Australia, but remains positive throughout the range of the data.

Current language skills depend on the use of the English language (either as a first or second language) in the home when the respondent was a youth (age 14). The partial effect of this variable (23 percentage points) is quite large. The most important role this variable plays in the analysis is standardizing for English-language fluency as a youth.

Language proficiency rates vary across birthplace groups, even after excluding from the data immigrants from English-speaking countries and controlling statistically for speaking English in the home at age 14. English-language fluency is greater among immigrants from countries in which English is a common second language and is lower among refugees. Compared to the Southern European control group, English-language proficiency rates are 25 percentage points higher for immigrants from Northern Europe, 36 percentage points higher for those from the Philippines, and 21 percentage points higher for immigrants from South Asia. Rates of English-language proficiency are 19 percentage points lower for immigrants from Vietnam, of whom nearly all are refugees. Language fluency rates do not differ between the Southern European benchmark group and immigrants from Arab countries, Other Asia, Eastern Europe, and South and Central America.

Finally, the estimated coefficient on the minority-language concentration variable is negative and highly significant. The magnitude of –0.050 indicates that a one percentage point increase in the representation of persons in the region of residence speaking the same minority language as the respondent is associated with a decline in the English proficiency rate of 5 percentage points. This partial effect is much lower than the 9 percentage point reduction determined from a simple (one variable) regression in which the concentration measure is the only variable. This suggests that some of the individual variables incorporated into the Table 18.1 column (1) specification must be correlated with the propensity to form language enclaves. Yet, there are no substantive changes in the coefficients of the other variables when the concentration measure is deleted from the equation in column (1).

In the Table 18.1, column (2) equation, three variables that are most obviously related to minority-language enclaves—foreign marriage, family in Australia and ethnic press—are included in the equation. The impacts of these variables are negative, and highly significant. Where the respondent is married and the spouse speaks the same minority language, English-language speaking skills are 22 percentage points lower than if the spouse was not of the same language group. Language skills are lower (by 5 percentage points) when the respondent has at least one family member (other than a spouse and children) present in Australia. Tests reveal that it is the presence of at least one family member, rather than the number of such members, that is important. Language skills are reduced, at a decreasing rate, the greater the number of foreign language newspapers, that is, by 2.5 percentage points per additional newspaper when evaluated at the mean number (four) of foreign language newspapers.

The ethnic network variables are also associated with a reduction to statistical and practical insignificance of the estimated effect of the minority-language concentration variable (from –0.050, $t = -5.2$ to 0.003, $t = -0.2$). When the foreign marriage variable is added to the Table 18.1 column (1) specification by itself, the coefficient on the minority-language concentration variable is –0.036 ($t = -3.89$), and when the foreign marriage and family variables are added to the Table 18.1 column (1) specification (but not the ethnic press variables), the coefficient on the minority-language concentration variable is –0.035 ($t = -3.75$). These findings indicate that in the restricted specification in Table 18.1, column (1), the minority-language concentration variable was a proxy for ethnic interaction factors captured in column (2) by the foreign marriage, family in Australia and ethnic press variables. This is consistent with the intention behind the inclusion of the minority-language concentration variable in previous studies where alternative measures were not available.

Columns (3) to (6) in Table 18.1 present estimates for models of English reading and writing skills. The pattern of effects for the models of speaking, reading and writing skills are generally the same. There are, however, some important differences. Education has a stronger effect on English reading and writing skills than it does on speaking skills. In other words, the degree of complementarity between education and English reading and writing skills is greater than that between education and English-speaking skills. As it seems reasonable to view reading and writing as more advanced skills, this finding is intuitively appealing. Moreover, while there is no gender difference in speaking and writing skills, female immigrants have poorer English reading skills than males, other things being the same.

The impact of the inclusion of the ethnic control variables in the equations for English reading and writing skills on the coefficients of the minority-language concentration variable is similar to what was found in the English-speaking equation. The estimated coefficient declines to statistical and practical insignificance. The *t*-ratios below unity imply that the adjusted coefficient of determination would increase if the minority-language concentration variable were deleted from the equation.

The inclusion of the three ethnic variables in the even numbered equations in Table 18.1 also raises the adjusted coefficient of determination of the equations. The increase of 3 to 4 percentage points in the (unadjusted) coefficient of determination is statistically significant. The inclusion of the three ethnic variables has no material influence on the coefficients of the other variables in the analysis, with the exception of the variable for being married. Those who are married to a spouse not of the same language origin have significantly higher levels of English-language skills than those who are single, but those who are married to a spouse from the same linguistic-country group have even lower fluency than those who are single.

Table 18.2 presents a decomposition of the influence of the individual characteristics on the coefficient of the minority-language concentration variable. Birthplace, age, education, English usage at age 14, foreign marriage and the ethnic press variables make positive contributions to the difference between the estimated impacts of the minority-language concentration variable in multiple and simple regressions. Years since migration, marital status and the gender variables make negative contributions.

Consider the positive contribution for education of 0.0139 in the equation

Table 18.2 Decomposition of the influence of regressors on the estimation of the minority-language concentration effect

	SPEAK	*READ*	*WRITE*
Difference between impacts estimated in multiple and simple regressions	0.0903	0.0978	0.0971
Derives from:			
Education	0.0139	0.0217	0.0198
Age	0.0079	0.0078	0.0068
Birthplace	0.0160	0.0155	0.0127
Years since migration[(a)]	–0.0348	–0.0322	–0.0319
Language at age 14	0.0133	0.0120	0.0136
Marital status	–0.0018	–0.0018	–0.0020
Female	–0.0001	–0.0001	0.0000
Foreign marriage	0.0249	0.0250	0.0251
Family	0.0021	0.0025	0.0026
Ethnic press[(a)]	0.0488	0.0472	0.0504

Note: a Includes effect of squared term.

Source: Table 18.1 and auxiliary regressions computed from the *Issues in Multicultural Australia Survey 1988*.

for speaking skills (Table 18.2). Education is associated with greater dominant language fluency (Table 18.1), but there is a negative correlation between education and the minority-language concentration variable, that is, better educated immigrants are less likely to live in language enclaves. Consequently,

$$-\beta_{\text{EDUC}}\, b_{\text{EDUC, CONC}}$$

is positive. As a result, controlling for education will reduce the impact of minority-language concentration on language fluency.

Years since migration has a positive effect on dominant language fluency. There is a positive correlation in these data between years since migration and the minority-language concentration variable, perhaps reflecting the different waves of migration to Australia, with earlier groups coming mainly from Europe while the more recent groups have a much wider representation across Europe and Asia, and hence a lower concentration measure.

Consequently,

$$-\beta_{\text{YSM}}\, b_{\text{YSM, CONC}}$$

is negative.

It is apparent from Table 18.2 that the reduction in the impact on dominant language fluency of the language enclave variable when the analysis is switched from a simple regression to a multiple regression framework is mainly due to the ethnic press (4.88 percentage points) and foreign marriage variables (2.49 percentage points). In other words, much of the substantial 8 to 9 percentage point effect of the minority-language concentration variable in a simple regression is due to effects associated with the more limited interactions an immigrant will have with the English language when the immigrant has access to ethnic-language media or when their spouse was born in the same linguistic-country group as the immigrant.

V. Conclusion

Dominant language fluency among immigrants has been shown in previous research to be adversely affected by residence in a region with a relatively large number of individuals that speak the same origin-language as the immigrant. Comparable statistical analyses conducted for Australia, Canada, Israel and the United States reveal that this relationship is remarkably robust.

The origins of the effect of the minority-language concentration measure are explored in this paper using a unique sample for Australia. It is demonstrated that the minority-language concentration measure reflects interactions in the marriage market, with family (other than a spouse and children) in Australia, and with ethnic media, and hence with formal ethnic networks. The addition of variables reflecting these concepts reduces to statistical and practical insignificance the estimated impact of the minority-language concentration variable, but results in only minor changes in other estimated effects. Thus, previous research which used the simple minority-language concentration index because other measures of language interaction were not available could not reveal the mechanism through which this process operated. When these variables are available they should be included in the analysis. Yet, the analysis also reveals that the simple minority-language concentration index can serve as a reasonable proxy for these other dimensions when the data are not available.

The analysis also shows that the model that was successful for analyzing the determinants of English speaking skills among immigrants in Australia is also successful for analyzing English reading and writing skills. The most noteworthy difference is that education is more important for explaining reading and writing than it is for explaining English-speaking proficiency.

Appendix 18A

Appendix Table 18.A1 presents the means and standard deviations of the variables used in the statistical analysis. Appendix Table 18.A2 presents estimates of a model of English-speaking skills among adult immigrants based on the 1988 *Issues in Multicultural Australia* (IMA) survey together with results from a comparable model based on the 1986 Australian Census of Population and Housing, one percent random sample of immigrants. The estimations are self-weighting; hence new arrivals are given relatively greater weight in the analyses based on the IMA. The pattern of the coefficients in the equation using the IMA survey is very similar to that obtained from the analysis of the large, random sample from the 1986 Census.

Appendix Table 18.A3 presents logit equations parallel to the OLS equations in Table 18.1. The substantive interpretations are the same.

Appendix 18.B

List of variables

English-language proficiency (SPEAK, READ, WRITE): The survey asks "Is English the first language spoken at home?" Respondents answering "Yes" were not asked additional questions on their English-language skills. Respondents answering "No" were then asked how well they speak, read and write English, with five response categories: Very Good, Good, Fair, Poor and Very Poor. SPEAK is set to one for individuals who report that English is the first language spoken at home, or if a language other than English is the first spoken in the home, speak English "very good". The SPEAK variable is set to zero where a language other than English is reported as the first spoken in the home and the respondent speaks English either "good", "fair", "poor" or "very poor." READ (English-reading skills) and WRITE (English-writing skills) are defined in a similar manner.

Table 18.A1 Means and standard deviations of variables

Variable	*Mean*	*Standard deviation*
Age	37.40	11.45
Education	12.74	3.69
Years since migration	12.74	12.29
Years since migration squared	313.25	480.98
Female	0.424	0.494
Married	0.729	0.444
English at age 14	0.283	0.451
Northern Europe	0.106	0.308
Eastern Europe	0.064	0.244
Southern Europe	0.214	0.410
Arabian countries	0.079	0.269
Southern Asia	0.130	0.337
Philippines	0.044	0.205
Vietnam	0.134	0.341
Other Asia	0.126	0.332
South and Central America	0.030	0.169
Africa	0.030	0.171
Other countries	0.043	0.204
Minority-language concentration	0.751	1.096
Foreign marriage	0.429	0.495
Family	0.700	0.458
Ethnic press	3.910	4.480
Ethnic press squared	35.347	57.215
SPEAK	0.496	0.500
READ	0.527	0.499
WRITE	0.501	0.500
Sample size	2032	

Source: Issues in Multicultural Australia Survey 1988 and for the ethnic press variable, Michael Clyne, *Community Languages: The Australian Experience*, Cambridge: Cambridge University Press, 1991, Tables 20, 21 and 22.

Age: Age in years for those age 15 to 64.

Years of education: This variable records the total years of full-time education. It has been created from the "Age Left School" and "Qualifications" variables. Years of education is calculated as age left school minus 5. Individuals who possess a trade certificate or diploma have been assigned an additional 2 years of education, individuals who possess a bachelors degree an additional 3.5 years of education, and individuals who have a higher degree have been assigned an additional 6 years of education.

Years since migration: The years since migration variable is computed as current age minus age on arrival.

Female: A binary variable, equal to one for females and zero for males.

Marital status: Married is a binary variable, defined to equal one for individuals who are married.

Table 18.A2 Regression estimates of English-language fluency among adult immigrants from non-English-speaking countries, Australia, 1986 and 1988 (dependent variable: SPEAK)

Variable	*1986 Census*	*1988 Survey*
Constant	0.267	0.288
	(9.11)	(5.07)
Age	–0.009	–0.007
	(22.23)	(7.30)
Education	0.034	0.019
	(25.14)	(7.53)
Years since migration	0.016	0.020
	(34.30)	(20.94)
Female	–0.044	–0.022
	(5.40)	(1.24)
Married	–0.038	–0.075
	(3.63)	(3.61)
Small urban location	–0.008	–0.039
	(0.52)	(0.88)
Rural location	0.032	0.005
	(2.07)	(0.13)
Northern Europe	0.267	0.258
	(19.83)	(8.32)
Eastern Europe	0.064	–0.065
	(3.94)	(1.52)
Arabian countries	–0.025	–0.030
	(1.27)	(0.73)
South Asia[b]	0.201	0.303
	(13.16)	(8.16)
Philippines	[a]	0.441
		(8.33)
Vietnam	–0.228	–0.212
	(11.21)	(6.76)
Other Asia[b]	[a]	–0.063
		(1.75)
South and Central America	–0.082	–0.044
	(2.68)	(0.76)
Africa	0.316	0.293
	(15.90)	(4.79)
Other countries	0.422	0.426
	(25.03)	(9.00)
Minority-language concentration	–0.060	–0.058
	(16.74)	(5.81)
Sample size	10157	2032
$\bar{R}^2$	0.3422	0.3964

Notes
(a) Variable not entered.
(b) Due to limited country-of-origin categories in the 1986 Census, the South Asia group includes a small number of cases from unidentified other parts of Asia.
The benchmark country category is Southern Europe.
'*t*' statistics in parentheses computed using White's (1980) heteroscedasticity-consistent covariance matrix estimator.

Source: Column (1): Australian Census of Population and Housing, 1986. Column (2): *Issues in Multicultural Australia Survey 1988*.

Table 18.A3 Logit estimates of models of English-language fluency among adult immigrants from non-English-speaking countries, Australia, 1988 (dependent variables: SPEAK, READ, WRITE)

Variable	*SPEAK*		*READ*		*WRITE*	
Constant	–1.881	–0.891	–2.463	–1.526	–2.529	–1.562
	(4.42)	(1.91)	(5.67)	(3.23)	(5.85)	(3.30)
Age	–0.040	–0.039	–0.037	–0.308	–0.034	–0.033
	(5.46)	(5.09)	(5.08)	(4.87)	(4.66)	(4.34)
Education	0.117	0.119	0.190	0.199	0.173	0.183
	(5.96)	(5.73)	(9.15)	(8.98)	(8.44)	(8.33)
Years since migration	0.172	0.178	0.154	0.161	0.149	0.156
(YSM)	(7.80)	(7.60)	(6.85)	(6.73)	(6.78)	(6.66)
$YSM^2/100$	–0.145	–0.179	–0.109	–0.140	–0.104	–0.139
	(2.66)	(3.08)	(1.92)	(2.33)	(1.90)	(2.37)
Female	–0.209	–0.219	–0.323	–0.331	–0.138	–0.131
	(1.67)	(1.67)	(2.54)	(2.51)	(1.09)	(0.99)
Married	–0.475	0.508	–0.511	0.530	–0.467	0.525
	(3.19)	(2.71)	(3.40)	(2.73)	(3.13)	(2.77)
English at age 14	1.558	1.515	1.565	1.491	1.588	1.531
	(9.30)	(8.51)	(8.80)	(7.93)	(9.29)	(8.43)
Northern Europe	2.366	1.697	2.167	1.484	1.880	1.198
	(7.25)	(4.96)	(6.60)	(4.31)	(6.25)	(3.75)
Eastern Europe	0.074	–0.090	–0.036	–0.205	–0.376	–0.578
	(0.27)	(0.30)	(0.13)	(0.70)	(1.38)	(1.94)
Arabian countries	0.205	0.402	0.248	0.461	0.114	0.335
	(0.83)	(1.49)	(1.10)	(1.74)	(0.46)	(1.25)
South Asia	1.299	0.162	1.204	0.110	1.188	0.063
	(5.17)	(0.55)	(4.76)	(0.37)	(4.76)	(0.21)
Philippines	2.161	1.163	2.785	1.780	2.316	1.277
	(6.10)	(3.03)	(6.55)	(3.94)	(6.08)	(3.12)

(Continued Overleaf)

Table 18.A3 Continued

Variable	*SPEAK*		*READ*		*WRITE*	
Vietnam	–1.127	–0.710	–1.097	–0.700	–1.161	–0.706
	(4.08)	(2.38)	(4.24)	(2.49)	(4.35)	(2.44)
Other Asia	0.064	–0.783	–0.217	–1.074	–0.215	–1.108
	(0.27)	(2.81)	(0.91)	(3.89)	(0.90)	(3.98)
South and Central America	0.088	0.158	–0.167	–0.144	–0.162	–0.085
	(0.24)	(0.39)	(0.45)	(0.37)	(0.44)	(0.21)
Africa	1.550	0.020	1.502	–0.060	1.458	–0.102
	(4.16)	(0.05)	(3.91)	(0.15)	(3.88)	(0.25)
Other countries	2.116	0.545	2.200	0.645	2.104	0.524
	(5.48)	(1.30)	(5.30)	(1.44)	(5.25)	(1.21)
Minority-language	–0.304	0.033	–0.291	0.034	–0.318	0.029
concentration	(5.22)	(0.43)	(4.92)	(0.45)	(5.40)	(0.38)
Foreign marriage	(a)	–1.481	(a)	–1.491	(a)	–1.480
		(8.47)		(8.33)		(8.44)
Family	(a)	–0.249	(a)	–0.311	(a)	–0.339
		(1.74)		(2.16)		(2.35)
Ethnic press	(a)	–0.326	(a)	–0.312	(a)	–0.356
		(5.31)		(5.13)		(5.82)
Ethnic press squared	(a)	0.016	(a)	0.015	(a)	0.018
		(4.02)		(3.85)		(4.61)
Chi-squared	1110.1	12367.0	1130.2	1253.3	1119.7	1254.2
Sample size	*2032*	*2032*	*2032*	*2032*	*2032*	*2032*

Notes
(a) Variable not entered.
Asymptotic '*t*' statistics in parentheses. The benchmark country category is Southern Europe.

Source: Issues in Multicultural Australia Survey 1988.

English at age 14: English was spoken in the home when the respondent was age 14.

Birthplace: The following birthplace regions were formed from the country codes available in the original data: Britain, Northern Europe, Southern Europe, Eastern Europe, Arabian countries, Philippines, Vietnam, South Asia (which primarily comprises the regions of British influence), Other Asia, Canada, United States, British West Indies, South and Central America, Africa, New Zealand, Other countries (including country not specified). Immigrants from Britain, Canada, United States, British West Indies and New Zealand are viewed as being from English-speaking countries and are deleted from the analysis. Immigrants from Southern Europe are used as the control group.

Ethnic affiliation: This is a binary variable coded to one where the respondent identified with a non-Australian ethnic or cultural group which was regarded as either very important or fairly important to him.

Family present: The FAMILY variable is a binary variable set equal to one if the respondent had either a mother, father, sibling or other relative (grandparents, grandchildren, aunts, uncles, parents-in-law, and brothers-and sisters-in-law) in Australia.

Social membership: The SOCIAL variable is a binary variable that is set equal to one where the respondent belonged to a social/sports club that was identified with a particular ethnic or cultural group.

Foreign marriage: Binary variable is set equal to one where the respondent's spouse was born in the same linguistic-country group as the respondent.

Location: Three binary variables were formed, the first for the benchmark group of individuals living in the major urban areas, the second for individuals living in "other urban areas" and the third for individuals living in "rural" areas. The "rural" and "urban" variables were insignificant in the analyses, see, for example, Appendix Table 18.A2 and therefore are not included in the analyses reported in the text.

Minority-language concentration: This variable is defined as the percentage of the population aged 15–64 in the region (defined broadly using information on location) in which the respondent lives that reports the same minority language as the respondent. The 12 minority languages coded in the Household Sample File of the 1986 Census from which the language data are derived are used in the construction of the variable. These are: Arabic/Lebanese, Chinese, Dutch, French, German, Greek, Italian, Maltese, Polish, Serbian and Croatian, Spanish, Vietnamese. It is set equal to zero for all other languages.

Ethnic press: This variable is defined as the number of Australian newspapers in 1986 in the language other than English reported by the respondent. 30 languages are identified.

Ethnic radio: This variable is defined as the weekly hours of broadcasts in 1986 in the non-English language of the respondent on the eleven radio stations in Australia with the most non-English broadcasts. Transmission times for over 60 languages are available.

Ethnic TV: This variable is a defined as the annual hours of community language broadcasts on the multicultural, Special Broadcasting Service, TV station in 1986/87. This station is government run and can be received in all of Australia. Broadcast times for over 30 languages are available. According to Clyne (1991, p. 149), "Time allocation is made on an annual basis, taking into consideration number of speakers and the availability of suitable films."

Note: All variables are dichotomous except education, age, duration in the destination, the minority concentration measure and the ethnic media variables. All of the variables are from the *Issues in Multicultural Australia Survey 1988*, except for the three ethnic media variables from Clyne (1991), Tables 20, 21 and 22.

Acknowledgements

The survey *Issues in Multicultural Australia 1988* analyzed in this chapter was made available by the Social Science Data Archives at the Australian National University. This chapter was presented at the European Science Foundation Research Conference on Migration and Development, Crete, Greece, October 1994. Comments on this chapter given by William Bridges, Evelyn Lehrer, the conference participants and the anonymous referees are appreciated.

References

Beenstock M (1993) Learning Hebrew and finding a job. Econometric analysis of immigrant absorption in Israel. Paper presented at the Conference on the Economics of International Migration: Econometric Evidence, February, Konstanz, Germany

Breton A (1978a) Billingualism. An economic approach. CD Howe Research Institute, Montreal

Breton A (1978b) Nationalism and language policies. Can J Econ 11:656–668

Chapman BJ, Iredale RR (1993) Immigrant qualifications. Recognition and relative wage outcomes. Int Migrat Rev 27:359–387

Chiswick BR (1991) Speaking, reading and earnings among low skilled immigrants. J Labor Econ 9:149–170

Chiswick BR, Miller PW (1992) Language in the labor market. The immigrant experience in Canada and the United States. In: Chiswick BR (ed) Immigration, language and ethnic issues: Canada and the United States. American Enterprise Institute, Washington DC, pp 229–296

Chiswick BR, Miller PW (1994a) Language choice among immigrants in a multilingual destination. J Popul Econ 7:119–131

Chiswick BR, Miller PW (1994b) Language and labor supply. The role of gender among immigrants in Australia. In: Slottje D (ed) Research on economic inequality, vol. 5. JAI Press, Greenwich, CT, pp 153–189

Chiswick BR, Miller PW (1995) The endogeneity between language and earnings. International analyses. J Labor Econ 13: No. 2, April

Clyne M (1991) Community languages. The Australian experience. Cambridge University Press, Cambridge

Dustmann C (1994) Speaking fluency, writing fluency and earnings of migrants in Germany. J Popul Econ 7:133–156

Evans MDR (1986) Sources of immigrants' language proficiency. Australian results with comparisons to the Federal Republic of Germany and the United States of America. Eur Sociol Rev 2:226–236

Johnson G, Solon G (1986) Estimates of the direct effects of comparable worth policy. Am Econ Rev 76:1117–1125

Kee P-K (1989) The language dimensions. Issues in multicultural Australia. Report to the Office of Multicultural Affairs

Veltman C (1983) Language shift in the United States. Mouton Publishers, Berlin

White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48:817–835

Part VI

Linguistic distance

19 The effect of linguistic distance and country of origin on immigrant language skills

Application to Israel[1]

With Michael Beenstock and Gaston L. Repetto

Introduction

Research on immigrant assimilation or absorption in the country of destination has increased in importance during the past two decades.[2] It has been concerned with implicit and explicit investments in human capital, including skills that are both internationally transferable and specific to the origin or the destination. An important aspect of this investment is the acquisition of language skills relevant for the destination labour market and consumption activities.[3] Immigrants may choose a destination, in part, because of the closeness of its dominant language to their own origin language, may make investments in the destination language prior to immigration, or do so after immigration, and, for some, all three factors may be relevant (Chiswick and Miller, 1994a).

Studies of immigrant proficiency in the destination language find that immigrants from some countries of origin are systematically less proficient in the destination language than immigrants from other countries, other variables being the same. What is not obvious, however, is whether these country-of-origin effects are due to the characteristics of the country *per se* or to characteristics of the origin language. If the origin language is more "distant" from the destination language, that is, if it is more difficult (more costly) for immigrants speaking the origin language to learn the destination language, the immigrants will be less proficient.[4]

As a result, a greater linguistic distance of the origin language from the destination language may reduce the extent of economic and social integration of immigrants in the host country and increase return migration. Alternatively, greater investments in destination language training would be required to achieve the same level of proficiency if the origin language is more distant from the destination language.

The purpose of this chapter is to separate statistically, for the first time, country of origin from language of origin effects in the immigrant's learning the destination language. This issue has not been addressed previously in the literature and would not be possible, statistically, if every country of origin had its own unique language, as there would be perfect collinearity among immigrants between country (C_i) and language (L_i). That this is not the case permits identification and statistical estimation of separate effects.

Moreover, the two data sets that are employed, one a census and the other a survey, have information on both fluency and literacy in Hebrew. Other modern censuses for immigrant receiving countries have data on fluency, but not on literacy in the destination language, and most surveys also focus on speaking skills. Thus, this chapter substantially expands the literature on determinants of destination language literacy.

Following a section that develops the methodology to be employed, data on immigrants in Israel from the two independent data sets employed for this analysis are discussed. Statistical results are then reported and the chapter closes with a summary and conclusion.

Methodology

The methodology employed in this chapter follows the methodology that has become standard in research on the economics of linguistic adjustment of immigrants (See Chiswick and Miller, 1992, 1995, 1998, 1999b). Conceptually, destination language proficiency is related to economic incentives, exposure to the destination language (prior to and after immigration), and efficiency in destination language acquisition. For empirical studies there are no direct measures of these concepts, and proxy variables are used. For example, it is hypothesized that destination language skills would be greater among those who immigrated at a younger age (efficiency), have been in the destination longer (exposure), and expect to remain longer (economic), and have more schooling (economic and efficiency), among other variables.

Country of origin is also frequently included in the analysis of destination language proficiency, and may play one of several roles. Country-of-origin dummy (dichotomous) variables may capture the propensity for the migration to be permanent, as temporary migrants have a lesser economic incentive to invest in skills specific to the destination.[5] Country of origin may also reflect pre-immigration exposure to the destination language if it is used as a lingua franca or as a primary or second language in the origin. The size of the group of immigrants from the same country (language) of origin may also matter. The larger the origin community in the destination, the easier it is for immigrants to avoid using the destination language in the labour market and in consumption activities, and hence the lower their proficiency.[6]

Alternatively, one could use dummy variables for language of origin ("mother tongue") rather than for country of origin in the statistical analysis. Language may matter because it is a proxy variable for the country—Korean-speaking immigrants come from Korea, Italian-speaking immigrants come from Italy, etc. Yet, language may be important independent of the country-of-origin effects. Some destination languages may be more difficult to learn by speakers of the origin language ("linguistic distance") (Chiswick and Miller, 1998, 1999b). Moreover, some languages may have an international role as a lingua franca, as does English in the early twenty-first century.

To identify country-of-origin as distinct from language-of-origin effects one needs to take advantage of a data source, or sources, that permit a statistical separation. This can occur if there are languages that are spoken in many countries or if there are countries of origin in which more than one language is spoken. In either case, there is a less than perfect one-to-one mapping between country of origin and language.

For the purpose of this study it is preferable to analyse a destination that has a unique language. The use of English as the destination language would create problems as English is spoken as a second language or a lingua franca in many countries. This rules out three countries with rich data sets on immigrants, the US, Canada and Australia. Another important immigrant-receiving country in the post-war period is Germany, but immigrants to Germany came primarily from a small set of countries, each with its own distinct language (e.g., Turkey and Greece).

Israel offers excellent prospects for the study of the separate effects of country of origin and language. The destination language, Hebrew, is not the dominant, second language or lingua franca in any of the origin countries, although immigrants may differ in the extent to which they learn Hebrew prior to immigration.[7] Unlike a destination country such as Germany, immigrants to Israel have come from a wide range of countries in Europe, North and South America, North Africa and the Middle East. The immigrants came from some countries with unique languages (e.g., Polish), and from some countries that share a common language (e.g., English, Arabic, French, Spanish). Moreover, as indicated in the next section, Israel has some data that, for other reasons, are superior to data from any other country for addressing the language issues.

The following terminology is used:

H is an index of Hebrew language skills,
C_i is a dichotomous variable that is unity for country i, and zero otherwise, $i = l, \ldots, N$,
L_j is a dichotomous variable that is unity for language j, and zero otherwise, $j = l, \ldots, M$,
YSM is years since migration,
AGEMIG denotes age at migration,
SCHOOL is years of schooling completed, and
X denotes a set of other demographic variables.

Then

$$H = f(C, L, YSM, AGEMIG, SCHOOL, X) \tag{1}$$

It is hypothesized that Hebrew language proficiency will increase with education and with duration in Israel, and decrease with age at migration. Regarding Israel, see Beenstock, 1996b; Beenstock and Ben Menahem, 1997; Chiswick, 1998; Chiswick and Repetto, 2001. The set of dichotomous variables C_i and L_j capture possible effects of country of origin and language of origin, respectively, on Hebrew language skills. If a language is spoken in only one country (for example, Norwegian in Norway), the dichotomous variables C_N for Norway and L_N for Norwegian are perfectly correlated and there is no point in including such observations in this analysis. If, however, the language is spoken in more than one country, or more than one language is spoken in a given country, then, in principle, the two effects can be identified statistically.

The effect of country of origin, as distinct from language, is measured by the coefficients of the country dichotomous variables. If there are no differences in country effects, the country-of-origin coefficients will not be statistically significant. The effect of language (linguistic distance and lingua franca effects) is measured by the coefficients of the language dichotomous variables. If there are no differences in language effects, the language coefficients will not be statistically significant.

The dependent variable in the statistical analysis is not a continuous variable, but rather is categorical. The statistical method employed is ordered logit analysis as there are three or more categories of the dependent variable.[8] The ordered logit model hypothesizes the existence of a latent variable, "Hebrew proficiency", which depends on the explanatory variables in equation (1).

The data

Data for Israel provide opportunity for estimating the separate effects of country and language. The two data sets employed are the 1972 Census of Israel, which is a 20 per cent simple random sample of the population, and the Immigrant Absorption Surveys (IAS) taken during the 1970s, both of which were conducted by the Central Bureau of Statistics.

The two data sets are complementary. The Census sheds light on language assimilation in the longer run and, unlike the IAS, includes data on immigrants from a wide range of countries who arrived during various periods. The mean duration of residence in the 1972 Census, for example, is 19 years. By contrast, the IAS has a comparative advantage in shedding light upon language assimilation shortly after arrival, that is, after one and three years. Linguistic distance (and other effects) may be empirically important in the longer term but not in the shorter term if it operates through the degree of linguistic assimilation. The sample size in the Census is much larger than in the IAS overall, but the IAS has a larger sample of those in Israel for 3 or less years. The IAS also has the advantage of permitting the estimation of cohort effects, and provides data for the respondents on Hebrew skills at immigration and participation in Hebrew language training programmes in Israel, variables that are not available in the Census.

Of the censuses of Israel, the 1972 Census, offers the richest data for this analysis. Unfortunately, the 1995 Census did not include any questions on Hebrew language skills. The 1983 Census included a question on languages spoken on a daily basis, but has no information on literacy. The 1972 Census, however, includes questions on speaking and writing, and appears to be the only modern census in a developed country that includes information on literacy. The 1972 Census included the following questions about Hebrew language skills:

(C1) "Do you know how to write at least a simple letter in Hebrew?"
(C2) "Which language or languages do you use in daily conversation? Note first, second and third".

Responses to the categorical Hebrew writing question can be viewed as ordered, where the responses are: H_w =0 if illiterate, H_w =1 if can write a simple letter in a language other than Hebrew, H_w =2 if can write in Hebrew and another language, and H_w =3 if can write only in Hebrew. The responses to the Hebrew speaking question can be thought of as ordered, where H_s =0 if Hebrew is not spoken at all, H_s = 1 if Hebrew is the secondary or tertiary language, H_s =2 if it is the primary but not only language, and H_s =3 if it is the only language spoken.

The Immigrant Absorption Survey (IAS) is a panel survey of immigrants who came to Israel at various years during the 1970s, and who were interviewed on arrival, after one and again after three years in Israel. In each wave of the survey they provided information on their self-assessed Hebrew language skills. The questions were:

(IAS 1) "Can you hold a simple conversation in Hebrew?"
(IAS 2) "Can you read a simple letter in Hebrew?"
(IAS 3) "Can you write a simple letter in Hebrew?"

The IAS asked only the speaking question on arrival but asked all three in years one and three.[9] Approximately 1,200 immigrants were interviewed in each immigrant cohort for the 1972/73, 1973/74, 1974/75 and 1978/79 immigration cohorts. See Beenstock (1996b) for the details regarding the IAS.

The Census and the IAS ask country of birth and country of emigration, where the two differ if there had been a prior international migration.[10] There are no data on the timing of international moves prior to immigration to Israel. Neither source asks language used in the origin. Origin language is inferred from country of emigration:

English: UK, Ireland, US, Canada, South Africa, Rhodesia, India, Australia, New Zealand.

French: France, Belgium, Switzerland, Luxembourg, Algeria, Tunisia, Morocco, Tangiers.

Arabic: Morocco, Tangiers, Libya, Algeria, Syria, Jordan, Egypt, Yemen, Iraq, Lebanon.

German: Germany, Switzerland, Austria.

Portuguese: Portugal, Brazil.

Spanish: Spain, Argentina, Chile, Uruguay, Mexico, and other Spanish-speaking Latin American countries.

A separate dichotomous variable is created for immigrants from countries in which two languages are important: Switzerland, Belgium, Morocco, Tangiers, Algeria. It is expected that those who grow up in a dual-language environment (BILCOUNTRY) will be more adept (efficient) at learning Hebrew. There is no mechanism for identifying those from Yiddish speaking origins.

Limiting analysis to the countries listed above sharply reduces sample size as many immigrants to Israel came from single language large countries (e.g., Russian in what was then the USSR). Nevertheless, there are still over 30,000 observations from the 1972 Census and over 800 observations from the IAS.

Statistical analysis

This section reports the statistical analysis of the determinants of proficiency in Hebrew to separate the effects of country of origin from language of origin. The analysis is performed for speaking and writing for adult males in the 1972 Census of Israel and for speaking, reading and writing for adult males and females in the Israel Immigrant Absorption Surveys of the 1970s.[11] As the dependent variable is a ranked categorical variable with more than two categories, ordered logit analysis is employed. The benchmark for the country groups is North Africa and for the language category Arabic.[12]

1972 Census

Table 19.1 reports the cross-tabulation of speaking and writing in Hebrew among foreign-born adult men from the set of countries under study, as reported in the 1972 Census. As would be expected, there is a positive relation between speaking and writing skills. Most of those who do not speak Hebrew on a daily basis are in the two lowest writing categories; most of those who speak Hebrew as their only language are in the two highest writing categories. There are, however, observations in which there is a divergence between reported speaking Hebrew on a daily basis and the ability to write a letter in Hebrew. For example, 6.5 per cent of the sample report that they do not speak Hebrew on a daily basis but they can write a letter in Hebrew.

The analysis of writing is performed without and with the statistical control for speaking Hebrew. This latter specification is implicitly based on a

Table 19.1 Cross tabulation of speaking Hebrew daily and Hebrew writing proficiency, adult male immigrants, selected sample, Israel, 1972 Census

Number of individuals in each speaking category who can write a letter in Hebrew

	Writing Hebrew				
Speaking Hebrew daily	*Illiterate*	*Other language*	*Hebrew and other*	*Hebrew only*	*Total*
None	920	984	716	183	2,803
Sec and Tir	989	1,116	2,629	648	5,382
Primary	1,328	1,069	14,160	5,441	21,998
Only	492	218	4,705	4,457	9,872
Total	3,729	3,387	22,210	10,729	40,055

Percentage of individuals in each speaking category who can write a letter in Hebrew

	Writing Hebrew				
Speaking Hebrew daily	*Illiterate*	*Other language*	*Hebrew and other*	*Hebrew only*	*Total*
None	32.8	35.1	25.5	6.5	100.0
Sec and Tir	18.4	20.7	48.8	12.0	100.0
Primary	6.5	4.9	64.4	24.7	100.0
Only	5.0	2.2	47.7	45.1	100.0
Total	9.3	8.5	55.5	26.8	100.0

Note: Detail may not add to total due to rounding.

Source: 1972 Census of Israel, 20 per cent sample, microdata file.

model in which writing is recursive to speaking, that is, speaking comes prior to writing.

Table 19.2 reports the means and standard deviations of the variables in the analysis of speaking Hebrew, as well as the mean and standard deviation of the Hebrew writing variable. All of explanatory variables are dichotomous, except for years of schooling, age at immigration, and years living in Israel. The dependent variables, speaking and writing, are each in four categories ranging from 0 to 3.

Table 19.3 reports the ordered logit regression results for the 1972 Census. The ordered logit model hypothesizes the existence of a latent variable which in this chapter is language skills. A positive coefficient in the ordered logit model implies that the expected value of the language skill is increasing with respect to the explanatory variable.

Other variables the same, Hebrew speaking (HEBPRIML) and writing (HEBWRIML) proficiency are higher for those with more schooling (SCHOOL), who have immigrated at a younger age (AGEMIG), and who have been in Israel longer (YSM) (Table 19.3). The effect of duration in Israel does not vary with age of migration; the coefficient on the interaction term (AGEMIGYSM) is not statistically significant. Those men who are married

Table 19.2 Means and standard deviations of relevant variables, adult male immigrants, selected sample, Israel, 1972 Census

Variable	*Mean*	*Standard deviation*
HEBPRIML	1.9711	0.8134
HEBWRIML	2.0029	0.8463
SCHOOL	8.2676	4.7329
AGEMIG	22.0229	11.9120
AGEMIGSQ	651.6729	648.7937
YSM	19.1700	8.3173
YSMSQ	436.6677	354.1900
AGEMIGYSM	386.0616	223.2200
MARR	0.8987	0.3017
MARROVER	0.3583	0.4795
CHILDREN	0.7972	0.4020
CHILDBOIS	0.7564	0.4292
GRNORAFR	0.4870	0.4998
GRASIA	0.3569	0.4791
GRWESEUR	0.1016	0.3022
GRNORAMER	0.0150	0.1217
GRLATAMER	0.0294	0.1689
GRSOUAFR	0.0048	0.0696
GROCEANIA	0.0009	0.0313
LRARABIC	0.4469	0.4972
LRENGLISH	0.0525	0.2231
LRFRENCH	0.3906	0.4879
LRSPANISH	0.0332	0.1792
LRPORTUGUES	0.0032	0.0573
LRGERMAN	0.0722	0.2588
BILCOUNTRY	0.3779	0.4848
HEBFLS	0.5496	0.4975
HEBSTLS	0.1329	0.3395
HEBNOLS	0.0688	0.2531

Note: All the means and standard deviations are calculated from the sample in the model reported in Table 19.3 for variable HEBPRIML, except for HEBWRIML, HEBFLS, HEBSTLS and HEBNOLS. Sample size 38,531.

Source: 1972 Census of Israel, 20 per cent sample, microdata file.

(MARR) have a higher level of speaking Hebrew than those not currently married if they married after immigration, but those who married prior to immigrating (MARROVER) have the same level as those not married. The only effect of marital status on writing skills is that those currently married but who married prior to immigrating have a higher level of proficiency in writing than those who married after migration. Having children (CHILDREN) significantly enhances speaking and writing proficiency of adult male immigrants, and for writing, especially if the children were born in Israel (CHILDBOIS). The analysis of writing in Table 19.3, column 3, shows that writing skills are lower the lower is the use of spoken Hebrew, but that controlling for the speaking variables does not materially alter the effect of the other variables.

Table 19.3 Ordered logit, analysis of Hebrew language proficiency, adult male immigrants, selected sample, Israel, 1972 Census

Variables	*HEBPRIML*		*HEBRWRIML*		*HEBWRIML*	
	Param.	*Std. Err.*	*Param.*	*Std. Err.*	*Param.*	*Std. Err.*
INTERCEPT1	–2.3888	0.1735	–0.2292	0.1753	0.6384	0.1780
INTERCEPT2	0.8117	0.1725	2.9308	0.1757	3.9231	0.1791
INTERCEPT3	2.4250	0.1736	3.8252	0.1762	4.8718	0.1797
SCHOOL	0.0608	0.0025	0.0836	0.0026	0.0680	0.0026
AGEMIG	–0.0635	0.0077	–0.1456	0.0079	–0.1389	0.0080
AGEMIGSQ	–0.0001	0.0001	0.0013	0.0001	0.0015	0.0001
YSM	0.1036	0.0098	0.0196	0.0099	–0.0061	0.0100
YSMSQ	–0.0016	0.0001	–0.0004	0.0001	0.0001	0.0001
AGEMIGYSM	0.0002	0.0002	0.0002	0.0002	0.0000	0.0002
MARR	0.1820	0.0392	0.0645	0.0399	0.0416	0.0402
MARROVER	–0.1764	0.0310	0.1747	0.0313	0.2285	0.0317
CHILDREN	0.2458	0.0342	0.1740	0.0346	0.1121	0.0351
CHILDBOIS	0.0489	0.0327	0.0861	0.0332	0.0577	0.0336
BILCOUNTRY	0.9046	0.1495	0.9281	0.1516	0.7083	0.1528
GRASIA	0.4530	0.0350	0.3105	0.0355	0.2351	0.0358
GRWESEUR	0.4589	0.1198	0.2515	0.1219	0.1438	0.1228
GRNORAMER	0.5591	0.1091	0.3953	0.1122	0.2821	0.1131
GRLATAMER	0.5542	0.1808	0.3337	0.1830	0.1891	0.1846
GRSOUAFR	0.0368	0.1574	0.4415	0.1645	0.4419	0.1659
GROCEANIA	0.0337	0.3239	0.2819	0.3302	0.2836	0.3311
LRENGLISH	–1.2392	0.0692	–0.8096	0.0698	–0.5202	0.0707
LRFRENCH	–0.8075	0.1473	–1.0251	0.1494	–0.8376	0.1506
LRSPANISH	–0.7687	0.1676	–0.7174	0.1697	–0.5350	0.1712
LRPORTUGUES	–0.4003	0.2439	–0.7040	0.2472	–0.6164	0.2489
LRGERMAN	–0.9324	0.1243	–1.0373	0.1263	–0.8413	0.1272
HEBFLS					–0.4847	0.0263
HEBSTLS					–1.2869	0.0400
HEBNOLS					–2.0540	0.0508
Sample size	38,531		37,840		37,840	

Source: 1972 Census of Israel, 20 per cent sample, microdata file.

The remaining variables in Table 19.3 reflect the country of origin and language of origin effects.[13] With North Africa the benchmark country (region), all the country coefficients are positive and in columns (1) and (2) are statistically significant for Asia, Western Europe, North America and, for speaking in Latin America, and for writing in South Africa. With Arabic as the benchmark, the language variables all have negative coefficients for speaking and writing, and are statistically significant for English, French, Spanish, German, and for writing, also Portuguese.[14] Country and language variables not statistically significant all have a very small number of observations. That is, the analysis permits a measurement of the separate effects of a North African origin and an Arabic speaking background, where by country of origin those from North Africa have the lowest level of proficiency and by language of origin those of Arabic speaking origins have the highest proficiency in Hebrew.

The variable for coming from a dual language/bilingual country (BILCOUNTRY) has a positive and statistically significant effect for both speaking and writing. Bilingual immigrants may have acquired a greater aptitude for learning third languages; it is as if the necessity to acquire more than one mother tongue makes it easier for them to learn additional languages.

The statistically significant country and language coefficients imply that there are separately identifiable country of origin and language of origin effects on proficiency in Hebrew among immigrants in Israel.

Immigrant absorption surveys

The Immigrant Absorption Surveys (IAS) permit an analysis for immigrants who arrived in the 1970s of the effect on proficiency of arrival cohort, as well as whether the immigrant entered a Hebrew language training programme (LEARN), and whether, in addition, the person completed the programme (FINISH). To enhance sample sizes, males and females are combined, and a test is made for gender differences (FEMALE).[15] Yet, the sample sizes are much smaller than in the Census and the immigrants have been in Israel for only a short period.

Table 19.4 reports the statistical analysis for speaking Hebrew at arrival in Israel. Hebrew speaking skills are greater, the higher the level of schooling, the younger the age at migration, and among those who arrived married. These reflect pre-immigration investments in the Hebrew language. There is no significant effect of cohort of arrival, sex, or number of children. Although the country coefficients are positive and the language coefficients are negative, none shows a statistically significant difference from the benchmarks, North Africa and Arabic.

Tables 19.5a and 19.5b are the analyses of the speaking, reading and writing skills of immigrants after one year in Israel, with and without the statistical controls for studying Hebrew. More dramatic differences are starting to emerge. For speaking, the 1974/75 cohort has a positive coefficient at the margin of statistical significance from the benchmark, the 1972/73 cohort (Table 19.5). After one year, women are significantly less proficient in speaking, reading and writing than are men. Those who finished a Hebrew language training programme are significantly more proficient than those who did not take such a programme. Some significant country and language-of-origin effects are beginning to emerge, notably for Latin America/Spanish for speaking Hebrew.

Table 19.6a and 19.6b perform the IAS ordered logit analysis for speaking, reading and writing at three years after immigration, with and without statistical controls for studying Hebrew. Compared with the benchmark, the 1972/73 cohort, the 1978/79 cohort is less proficient in speaking, reading and writing Hebrew, but this effect is weaker when there is the statistical control

Table 19.4 Ordered logit analysis of speaking Hebrew at arrival, adult male and female immigrants, selected sample, Israel, Immigrant Absorption Surveys, 1970s

Variable	*SPEAK0*	
	Parameter	*Standard Err.*
INTERCEPT1	–1.2903	0.9933
INTERCEPT2	0.2364	0.9921
73/74	–0.0066	0.1893
74/75	0.2165	0.1944
78/79	0.1332	0.1981
SCHOOL	0.0950	0.0198
AGEMIG	–0.1284	0.0280
AGEMIGSQ	0.0013	0.0003
MARROVER	0.3485	0.1672
NUMCHILDR	–0.0081	0.0683
FEMALE	–0.1746	0.1339
BILCOUNTRY	0.9308	0.6188
GRASIA	–0.5433	0.7691
GRWESEUR	0.9522	0.6466
GRNORAMER	1.0242	0.6928
GRLATAMER	1.5356	0.9179
GRSOUAFR	0.6232	0.7258
GROCEANIA	0.3411	0.8260
LRENGLISH	–0.2035	0.6550
LRFRENCH	–0.6668	0.6925
LRSPANISH	–0.9140	0.8854
LRPORTUGUES	–0.5083	0.9446
LRGERMAN	–0.9594	0.9833
Sample size	1,032	

Source: Immigrant Absorption Surveys, Central Bureau of Statistics, Israel.

for the Hebrew language training programme. Proficiency is greater for those with more schooling and who immigrated at a younger age. Speaking, reading and writing skills are lower for women than for men, and literacy (reading and writing) skills are lower among those with children.[16]

Among immigrants in Israel for three years, country of origin coefficients are all positive and at or near statistical significance in several instances. The language of origin variables all have negative signs, are statistically significant for speaking English and are generally statistically significant for reading and writing. These effects hold even when the Hebrew language training variables are held constant. The variable for having taken a Hebrew language training programme (LEARN) is not statistically significant, but the variable for completing the programme (FINISH) is highly statistically significant.

Tables 19.7a and 19.7b repeat the statistical analysis of Hebrew language proficiency after three years in Israel with statistical controls for Hebrew speaking skills at arrival in Israel. Those who spoke no Hebrew (HEB-SPENONE0) and those who spoke Hebrew with difficulty (HEBSPEWDIFO)

Table 19.5a Ordered logit, analysis of Hebrew language proficiency, adult male immigrants, selected sample, Israel, Immigrant Absorption Surveys, 1970s

Variables	*SPEAK1*		*READ1*		*WRITE1*	
	Param.	*Std. Err.*	*Param.*	*Std. Err.*	*Param.*	*Std. Err.*
INTERCEPT1	0.9815	1.0344	0.3290	1.1866	0.4777	1.2085
INTERCEPT2	3.7456	1.0428	2.8699	1.1930	3.1175	1.2163
73/74	–0.1301	0.1525	0.2614	0.1459	–0.1225	0.1484
74/75	0.2977	0.1638	0.3429	0.2548	0.0530	0.2586
SCHOOL1	0.1876	0.0189	0.1731	0.0199	0.1965	0.0206
AGEMIG	–0.1413	0.0271	–0.1571	0.0285	–0.1519	0.0287
AGEMIGSQ	0.0009	0.0003	0.0014	0.0003	0.0012	0.0003
MARR1	–0.0612	0.0902	0.0353	0.0925	0.0784	0.0938
MARROVER	0.1025	0.2449	0.3292	0.2531	0.4567	0.2587
CHILDREN1	–0.0331	0.0610	–0.1191	0.0707	–0.1519	0.0719
FEMALE	–0.3030	0.1325	–0.4145	0.1392	–0.3364	0.1414
BILCOUNTRY	0.7569	0.7367	0.9350	0.8573	1.0118	0.8756
GRASIA	0.3009	0.7780	–0.0999	0.9244	–0.6515	0.9472
GRWESEUR	0.8717	0.7357	0.6168	0.8610	0.5445	0.8798
GRNORAMER	0.7151	0.7727	0.7134	0.9000	0.2052	0.9197
GRLATAMER	2.6721	1.0180	1.4116	1.1182	1.4780	1.1514
GRSOUAFR	0.9300	0.8011	0.6215	0.9254	0.2803	0.9449
GROCEANIA	0.6976	0.8818	1.3009	1.0522	0.3873	1.0669
LRENGLISH	–0.8849	0.4617	–0.9682	0.5234	–0.9680	0.5333
LRFRENCH	–0.4660	0.5269	–0.8622	0.5864	–1.2349	0.5964
LRSPANISH	–2.2211	0.8335	–1.3361	0.8648	–1.7023	0.8914
LRPORTUGUES	–1.2670	0.9318	–0.6831	0.9536	–0.5327	0.9924
LRGERMAN	–0.8779	0.8064	–1.0519	1.0302	–1.3109	1.0426
Sample size	1,071		871		870	

Source: Immigrant Absorption Surveys, Central Bureau of Statistics, Israel.

at arrival are less proficient in Hebrew at three years in Israel, even after the statistical control for whether a Hebrew language training programme was taken in Israel. Controlling for initial speaking skills, having taken but not completed a language training programme enhanced all three measures of proficiency, but completing the programme had an even more favourable effect.

Country coefficients are all positive but not statistically significant. Language of origin variables are negative and statistically significant for English (speaking, reading and writing) and for French (reading), even after controlling for initial speaking ability.

The 1978/79 cohort in Tables 19.7a and 19.7b still shows lesser proficiency in writing, but not in speaking and reading, and the other demographic effects are as before.

In summary, the analysis of the IAS indicates that proficiency in Hebrew is greater for those with more schooling, who immigrated at a younger age, and among those in Israel at least one year, for those who completed a Hebrew language training programme in Israel and had better Hebrew language skills

Table 19.5b Ordered logit, analysis of Hebrew language proficiency after one year in Israel with language school variables, adult male and female immigrants, selected sample, Israel, Immigrant Absorption Surveys, 1970s

Variables	*SPEAK1*		*READ1*		*WRITE1*	
	Param.	*Std. Err.*	*Param.*	*Std. Err.*	*Param.*	*Std. Err.*
INTERCEPT1	0.4119	1.2394	−1.9768	1.7120	−1.0677	1.7627
INTERCEPT2	3.1318	1.2465	0.4353	1.7121	1.2969	1.7652
73/74	−0.1288	0.1870	−0.0445	0.1783	−0.2089	0.1830
74/75	0.2855	0.1973	0.1457	0.2758	0.0701	0.2810
SCHOOL1	0.1739	0.0216	0.1721	0.0233	0.1882	0.0241
AGEMIG	−0.1451	0.0306	−0.1457	0.0337	−0.1688	0.0346
AGEMIGSQ	0.0010	0.0003	0.0013	0.0004	0.0014	0.0004
MARR 1	−0.0081	0.1051	0.0916	0.1109	0.1106	0.1126
MARROVER	0.1724	0.2808	0.4767	0.3012	0.5419	0.3098
CHILDREN1	−0.0099	0.0677	−0.0685	0.0817	−0.0649	0.0831
FEMALE	−0.4667	0.1560	−0.4344	0.1671	−0.3259	0.1707
BILCOUNTRY	0.8872	0.8393	2.1041	1.2802	2.2692	1.3310
GRASIA	0.3093	0.8985	1.4018	1.3700	0.6462	1.4248
GRWESEUR	1.1076	0.8362	2.0876	1.2835	1.9083	1.3335
GRNORAMER	0.8563	0.8856	2.0927	1.3297	1.6220	1.3796
GRLATAMER	2.9944	1.1778	2.7445	1.5528	2.6619	1.6258
GRSOUAFR	1.3032	0.9530	2.6592	1.3824	2.1298	1.4294
GROCEANIA	0.9460	1.0233	2.9099	1.5278	1.8148	1.5561
LRENGLISH	−0.6335	0.5520	−0.5616	0.6528	−0.7251	0.6775
LRFRENCH	−0.4199	0.6272	−0.4676	0.7303	−1.0016	0.7525
LRSPANISH	−2.3100	0.9716	−0.9534	1.0311	−1.2772	1.0754
LRPORTUGUES	−0.9543	1.0684	0.1703	1.1223	0.5621	1.2034
LRGERMAN	−0.7402	0.8729	−0.7748	1.1058	−1.0269	1.1310
FINISH1	1.0257	0.1925	0.3806	0.2035	0.4792	0.2064
LEARN1	−0.0423	0.1884	0.2584	0.2092	0.1221	0.2138
Sample size	824		624		623	

Source: Immigrant Absorption Surveys, Central Bureau of Statistics, Israel.

at arrival. The 1978/79 cohort shows poorer performance in writing Hebrew, but not in speaking or reading Hebrew, than the other cohorts. There is no sex difference in Hebrew speaking skills at arrival, but a gap increases in favour of men with duration in Israel, even after controlling for taking a Hebrew language training programme and speaking level at immigration. This may be a consequence of the greater anticipated and actual labour market experience in Israel for men in the sample.

There is evidence in the IAS that country and language of origin do not influence Hebrew speaking skills at arrival, but their effect increases with duration in Israel, especially for English language speakers. Even after controlling for initial speaking skills and for taking a Hebrew training programme, among other variables, immigrants from English language origins are significantly less proficient in speaking, reading and writing Hebrew.

Table 19.6a Ordered logit analysis of Hebrew language proficiency after three years in Israel, adult male and female immigrants, selected sample, Israel, Immigrant Absorption Surveys, 1970s

Variables	*SPEAK3*		*READ3*		*WRITE3*	
	Param.	*Std. Err.*	*Param.*	*Std. Err.*	*Param.*	*Std. Err.*
INTERCEPT1	2.7786	1.2443	0.1915	1.0595	–0.1982	1.0168
INTERCEPT2	3.0084	1.2451	0.4676	1.0597	0.1361	1.0168
INTERCEPT3	4.9450	1.2520	2.0596	1.0623	1.8115	1.0192
INTERCEPT4	5.3686	1.2534	2.5697	1.0637	2.0881	1.0199
73/74	–0.0431	0.2176	–0.1882	0.1923	–0.3176	0.1933
74/75	–0.1184	0.2203	0.0020	0.1942	–0.0976	0.1949
78/79	–0.1058	0.2394	–0.3312	0.2051	–0.8305	0.2064
SCHOOL3	0.1801	0.0207	0.2071	0.0193	0.2159	0.0195
AGEMIG	–0.1442	0.0282	–0.1296	0.0266	–0.1059	0.0253
AGEMIGSQ	0.0009	0.0003	0.0011	0.0003	0.0008	0.0003
MARR3	0.0199	0.1120	–0.0022	0.0840	0.1103	0.0828
MARROVER	–0.1062	0.2583	0.0855	0.2094	0.2309	0.2075
CHILDREN3	–0.0497	0.0633	–0.1191	0.0599	–0.1525	0.0608
FEMALE	–0.3585	0.1627	–0.4100	0.1415	–0.3705	0.1413
BILCOUNTRY	0.2453	0.9238	1.3800	0.7544	0.8788	0.6908
GRASIA	0.3115	0.9547	1.0049	0.8051	0.7618	0.7554
GRWESEUR	0.7987	0.9129	1.2278	0.7555	0.7551	0.6956
GRNORAMER	1.1737	0.9618	1.4670	0.8016	1.2811	0.7469
GRLATAMER	0.9665	1.4414	2.1989	1.0867	2.0032	1.0399
GRSOUAFR	0.9633	1.0046	1.3237	0.8409	0.9890	0.7900
GROCEANIA	1.2104	1.1247	1.3463	0.9353	0.8102	0.8940
LRENGLISH	–1.4362	0.5065	–1.5866	0.4737	–1.7754	0.4802
LRFRENCH	–0.4238	0.6023	–1.03432	0.5469	–1.1899	0.5521
LRSPANISH	–0.7240	1.2382	–1.7794	0.9302	–2.1054	0.9280
LRPORTUGUES	–0.2773	1.3673	–1.4741	1.0051	–1.7788	0.9960
LRGERMAN	–0.9970	0.8664	–1.6817	0.7882	–1.5559	0.7947
Sample size	852		854		855	

Source: Immigrant Absorption Surveys, Central Bureau of Statistics, Israel.

Table 19.6b Ordered logit analysis of Hebrew language proficiency after three years in Israel, with language school variables, adult male and female immigrants, selected sample, Israel, Immigrant Absorption Surveys, 1970s

Variables	*SPEAK3*		*READ3*		*WRITE3*	
	Param.	*Std. Err.*	*Param.*	*Std. Err.*	*Param.*	*Std. Err.*
INTERCEPT1	2.6026	1.2360	–0.0326	1.0651	–0.3617	1.0224
INTERCEPT2	2.8352	1.2368	0.2432	1.0651	–0.0271	1.0224
INTERCEPT3	4.7984	1.2434	1.8532	1.0672	1.6622	1.0243
INTERCEPT4	5.2280	1.2449	2.3754	1.0684	1.9430	1.0249
73/74	–0.0238	0.2191	–0.1836	0.1933	–0.3232	0.1941
74/75	–0.1233	0.2215	0.0189	0.1951	–0.0897	0.1956
78/79	0.1306	0.2529	–0.0216	0.2229	–0.5729	0.2243
SCHOOL3	0.1667	0.0211	0.1972	0.0196	0.2081	0.0198
AGEMIG	–0.1485	0.0287	–0.1325	0.0275	–0.1066	0.0259
AGEMIGSQ	0.0009	0.0003	0.0012	0.0003	0.0009	0.0003
MARR3	0.0306	0.1125	0.0013	0.0841	0.1125	0.0829
MARROVER	–0.1234	0.2602	0.0482	0.2107	0.1905	0.2087
CHILDREN3	–0.0295	0.0640	–0.0951	0.0608	–0.1341	0.0615
FEMALE	–0.3823	0.1636	–0.4511	0.1426	–0.4017	0.1423
BILCOUNTRY	0.3492	0.9079	1.6041	0.7556	1.0454	0.6940
GRASIA	0.2813	0.9384	1.0471	0.8050	0.8061	0.7569
GRWESEUR	0.8590	0.8964	1.3643	0.7557	0.8655	0.6978
GRNORAMER	1.2180	0.9457	1.6136	0.8016	1.3990	0.7486
GRLATAMER	0.8742	1.4245	2.2575	1.0862	2.0470	1.0426
GRSOUAFR	0.9135	0.9928	1.4004	0.8432	1.0624	0.7938
GROCEANIA	1.2395	1.1157	1.4083	0.9379	0.8892	0.8981
LRENGLISH	–1.3592	0.5116	–1.5822	0.4783	–1.7774	0.4843
LRFRENCH	–0.4591	0.6066	–1.4664	0.5511	–1.2811	0.5555
LRSPANISH	–0.6139	1.2320	–1.7910	0.9312	–2.1000	0.9308
LRPORTUGUES	–0.1998	1.3553	–1.5272	1.0026	–1.7953	0.9976
LRGERMAN	–0.9107	0.8790	–1.6913	0.7909	–1.5705	0.7969
FINISH3	0.6429	0.2302	0.6565	0.1949	0.5358	0.1933
LEARN3	0.0628	0.1952	–0.0580	0.1754	–0.1090	0.1757
Sample size	851		853		854	

Source: Immigrant Absorption Surveys, Central Bureau of Statistics, Israel.

Table 19.7a Ordered logit analysis of Hebrew language proficiency after three years in Israel, controlling for Hebrew proficiency at arrival, adult male and female immigrants, selected sample, Israel, Immigrant Absorption Surveys, 1970s

Variables	*SPEAK3*		*READ3*		*WRITE3*	
	Param.	*Std. Err.*	*Param.*	*Std. Err.*	*Param.*	*Std. Err.*
INTERCEPT1	5.0880	1.3817	1.5440	1.1507	0.5863	1.1387
INTERCEPT2	5.3739	1.3837	1.8877	1.1515	1.0086	1.1392
INTERCEPT3	7.4091	1.3979	3.6315	1.1561	2.8446	1.1425
INTERCEPT4	7.9033	1.4004	4.2388	1.1579	3.1715	1.1432
73/74	0.0433	0.2402	–0.1233	0.2122	–0.2099	0.2139
74/75	–0.1122	0.2454	–0.0338	0.2152	–0.1023	0.2172
78/79	–0.0171	0.2544	–0.3035	0.2186	–0.8703	0.2211
SCHOOL3	0.1547	0.0222	0.1849	0.0207	0.1910	0.0209
AGEMIG	–0.1277	0.0319	–0.1132	0.0303	–0.0727	0.0282
AGEMIGSQ	0.0007	0.0003	0.0010	0.0003	0.0005	0.0003
MARR3	0.0452	0.1209	0.0535	0.0911	0.1958	0.0907
MARROVER	–0.2483	0.2788	0.1381	0.2259	0.2613	0.2252
CHILDREN3	–0.0184	0.0700	–0.1078	0.0662	–0.1491	0.0674
FEMALE	–0.3103	0.1783	–0.4384	0.1531	–0.3510	0.1536
BILCOUNTRY	–0.1056	0.9680	1.1515	0.8036	1.0733	0.7991
HEBSPEWDIF0	–1.6189	0.5060	–1.0315	0.3075	–1.3585	0.3005
HEBSPENONE0	–3.0315	0.4807	–2.3346	0.2861	–2.5918	0.2845
GRASIA	0.6191	0.9894	1.3192	0.8512	1.5576	0.8539
GRWESEUR	0.6713	0.9481	1.1234	0.8011	1.0929	0.7985
GRNORAMER	0.7902	1.0028	1.1591	0.8515	1.4920	0.8504
GRLATAMER	1.1494	1.5260	1.0743	1.2309	1.6371	1.2105
GRSOUAFR	0.9581	1.0474	1.3939	0.8919	1.5374	0.8921
GROCEANIA	1.4611	1.1719	1.5022	0.9814	1.4919	0.9852
LRENGLISH	–1.2318	0.5199	–1.4374	0.4820	–1.6871	0.4927
LRFRENCH	0.1388	0.6369	–1.0573	0.5676	–0.8480	0.5771
LRSPANISH	–0.6634	1.3155	–0.5472	1.0625	–1.1659	1.0474
LRPORTUGUES	–0.6506	1.4557	–0.5344	1.1581	–0.9058	1.1414
LRGERMAN	–0.3214	0.9826	–1.8808	0.8823	–1.5975	0.8949
Sample size	761		763		764	

Source: Immigrant Absorption Surveys, Central Bureau of Statistics, Israel.

Table 19.7b Ordered logit analysis of Hebrew language proficiency after three years in Israel, adult male and female immigrants, controlling for Hebrew proficiency at arrival and language school participation, Israel, Immigrant Absorption Surveys, 1970s

Variables	*SPEAK3*		*READ3*		*WRITE3*	
	Param.	*Std. Err.*	*Param.*	*Std. Err.*	*Param.*	*Std. Err.*
INTERCEPT1	5.1397	1.3888	1.6470	1.1768	0.6865	1.1712
INTERCEPT2	5.4306	1.3908	1.9899	1.1777	1.1081	1.1718
INTERCEPT3	7.5263	1.4054	3.7746	1.1830	2.9875	1.1757
INTERCEPT4	8.0409	1.4082	4.4263	1.1851	3.3350	1.1765
73/74	0.0991	0.2435	–0.0973	0.2148	–0.1820	0.2165
74/75	–0.0786	0.2480	0.0144	0.2182	–0.0650	0.2200
78/79	0.2299	0.2714	0.0254	0.2399	–0.6073	0.2427
SCHOOL3	0.1288	0.0228	0.1635	0.0211	0.1703	0.0214
AGEMIG	–0.1357	0.0332	–0.1330	0.0321	–0.0896	0.0305
AGEMIGSQ	0.0008	0.0004	0.0013	0.0004	0.0007	0.0003
MARR3	0.0762	0.1222	0.0604	0.0920	0.2023	0.0916
MARROVER	–0.2231	0.2812	0.1434	0.2279	0.2579	0.2275
CHILDREN3	0.0106	0.0709	–0.0684	0.0673	–0.1139	0.0686
FEMALE	–0.3715	0.1800	–0.5245	0.1553	–0.4399	0.1558
BILCOUNTRY	0.0358	0.9623	1.4623	0.8151	1.3457	0.8107
HEBSPEWDIF0	–2.0297	0.5173	–1.4753	0.3248	–1.7781	0.3183
HEBSPENONE0	–3.4302	0.4947	–2.7889	0.3095	–3.0340	0.3083
GRASIA	0.4949	0.9844	1.3158	0.8652	1.5395	0.8707
GRWESEUR	0.6498	0.9410	1.1920	0.8126	1.1429	0.8111
GRNORAMER	0.7510	0.9973	1.2530	0.8631	1.5249	0.8633
GRLATAMER	1.1442	1.5271	1.2114	1.2148	1.7316	1.1982
GRSOUAFR	0.6697	1.0444	1.2586	0.9043	1.3536	0.9054
GROCEANIA	1.2558	1.1647	1.3876	0.9916	1.3767	0.9968
LRENGLISH	–1.0823	0.5301	–1.3071	0.4946	–1.5336	0.5078
LRFRENCH	0.1048	0.6474	–1.1239	0.5808	–0.8893	0.5932
LRSPANISH	–0.7503	1.3254	–0.6807	1.0401	–1.2520	1.0313
LRPORTUGUES	–0.7806	1.4635	–0.7682	1.1315	–1.0735	1.1205
LRGERMAN	–0.0313	1.0123	–1.6997	0.8955	–1.4058	0.9108
FINISH3	0.5869	0.2461	0.6556	0.2105	0.5109	0.2095
Llearn3	0.5543	0.2243	0.5644	0.2043	0.5983	0.2065
Sample size	760		762		763	

Source: Immigrant Absorption Surveys, Central Bureau of Statistics, Israel.

Summary and conclusions

This chapter uses two data sets to study for the first time the separate effects of country of origin and language of origin on the destination language proficiency of immigrants. The analysis is for the Hebrew language fluency and literacy of immigrants in Israel in the 1970s. Due to data limitations there has been very little research for any country on immigrant literacy in the destination language. Data are from the 1972 Census of Israel, which includes the full spectrum of durations of residence, and from the Immigrant Absorption Surveys (IAS) of the 1970s, which were panel surveys conducted at arrival and after one and three years in Israel. Israel is well suited for this study as the Hebrew language is not an international language (as is English), is not a mother tongue in any particular origin, and immigrants to Israel come from a wide range of countries and linguistic origins. Data on adults who immigrated from linguistic origins that encompass more than one country are used for this analysis.

The findings on adult males in the 1972 Census and on adult males and females in the IAS are consistent with each other and with other studies. Proficiency in speaking, reading and writing Hebrew is greater for those who have more years of formal schooling, who immigrated to Israel at a younger age, and who have lived in Israel a longer period. Marriage prior to immigration is associated with lower Hebrew language skills compared with those who married after immigration in the 1972 Census, but not in the IAS. In the IAS the immigrant cohort that arrived in 1978/79 has poorer Hebrew writing skills than earlier immigration cohorts in the 1970s, other variables being the same.

The IAS has unique data on destination language training, data generally not available in other immigration surveys. These data indicate that proficiency in Hebrew is greater among those who took a Hebrew language training programme after arrival in Israel, and even greater among those who completed the programme.

The 1972 Census analysis indicates that immigrants from dual-language countries are more proficient in Hebrew than are other immigrants, other variables being the same. This may arise because experience with a second language as youths enhances language learning abilities (lower costs). However, according to the IAS data the dual-language effect does not show up in the first three years.

Separate effects of country of origin and language of origin are identified in the 1972 Census. Immigrants from most countries of birth are more proficient in Hebrew than those from North Africa. Immigrants from most languages of origin are less proficient in Hebrew than those from Arabic-speaking countries. Arabic appears statistically as the language closest to Hebrew, which is consistent with their both being Semitic languages.

English-origin speakers in Israel in the 1972 Census are at the greatest linguistic disadvantage in speaking and writing Hebrew. This result is also confirmed by the IAS. In the IAS, English speakers show the least proficiency in speaking, reading and writing Hebrew; however, this effect is detectable only after three years. Indeed, there is no evidence of a linguistic disadvantage for English speakers in the IAS after one year for any of the three language skills. Thus, the gap between English speakers and others increases with duration in Israel because of the slower progress of the English speakers in acquiring Hebrew language skills.

The "linguistic distance" of the origin language from the destination language may help solve an econometric problem in estimating earnings functions for immigrants. Because of the possible endogeneity of destination language skills and earnings, the estimation of the earnings function with language skills on the right hand side would generate biased estimates of the effect of language on earnings. A measure of linguistic distance may serve as an identifying variable that enters a language equation but not an earnings equation (Chiswick and Miller, 1999a [2002]).

These findings suggest that there are linguistic distance effects on acquiring proficiency in the destination language and that these effects can be important and take time to emerge. A greater linguistic distance implies greater difficulty in learning the destination language, which implies both a less successful economic and social adjustment as well as a higher propensity for return migration. The pattern for English suggests that its role as an international language and a lingua franca in Israel reduces the incentive of native speakers of English to invest in acquiring proficiency in Hebrew. This makes English a special case in both the short term and the long term. It is likely that this unique role for English applies in other immigrant receiving countries in which English is not the dominant language.

This analysis provides further evidence on the importance of both country of origin characteristics and the "distance" of the origin language from the destination language for understanding immigrant linguistic adjustment. While the parameters will differ in destinations other than Israel, and for languages other than Hebrew, the methodology should have broad applicability.

Notes

1 We are indebted to the Social Science Data Archive, Hebrew University and the Central Bureau of Statistics, Israel, for making available data used in this study.
2 This literature began with Chiswick (1978).
3 Hebrew language proficiency (speaking, reading and writing) has been shown to be an important determinant of earnings among immigrants in Israel in a variety of census and survey data. See, for example, Beenstock, 1996a; Beenstock and Ben Menachem, 1997; Berman et al., 1999; Chiswick, 1998; Chiswick and Repetto (2001).
4 This approach to "linguistic distance" is different from the standard linguistic approach which is concerned with the roots or evolution of languages. In explaining the relationship among languages, the *Cambridge Encyclopedia of Language* indicates: "The main metaphor that is used to explain the historical relationships is that of the language *family* or *family tree*" (Crystal, 1987: 292, italics in original).

See also Grimes and Grimes, 1993. The *Encyclopedia* (Crystal 1987: 283–340) has an extensive discussion of families of languages, but only two brief paragraphs in a side bar box on "interlingual distance": "The structural closeness of languages to each other has often been thought to be an important factor in FLL (foreign language learning). However, it is not possible to correlate linguistic difference and learning difficulty in any straight forward way, and even the basic task of quantifying linguistic difference proves to be highly complex, because of the many variables involved" (Crystal, 1987: 371). Chiswick and Miller (1998), however, develop an index of the linguistic distance of myriad languages from English, based on the difficulty of Americans learning these languages. This measure in turn is found to be an important factor in explaining variation in proficiency in English in the US, Canada and Australia among immigrants from various non-English speaking origins (Chiswick and Miller, 1998, 1999b, 1999c). As far as we are aware, no comparable index has been developed for the linguistic distance of world languages from any other language, including Hebrew.

5 High-skilled North American immigrants in Israel and low-skilled Mexican immigrants in the US have a high propensity for return migration. For an analysis of return migration from Israel, see Beenstock, 1996b, and for an analysis of the effect of expectations of emigration on the language skills of immigrants in the US, see Chiswick and Miller, 1998.

6 For Israel, see Chiswick, 1998 and Chiswick and Repetto, 2001. For the US, Canada and Australia, see Chiswick and Miller, 1992, 1995, 1998, 1999b.

7 To various degrees, both across diaspora countries and across Jewish communities within diaspora countries, Hebrew language skills are developed for religious as well as Zionist reasons. Yet there would not be any immigrants to Israel that are proficient in Hebrew but who lack proficiency in another language used in the origin. It will be shown that in the Immigrant Absorption Survey, Hebrew speaking skills at arrival in Israel of immigrants during the 1970s did not vary systematically by country or language of origin.

8 Ordered logit analysis has been used in previous research on the determinants of destination language proficiency. See, for example, Beenstock (1996b) for an analysis for Israel, Espinoza and Massey (1997) for Mexican immigrants in the US and Espenshade and Fu (1997) for immigrants in general in the US.

9 The response categories for the IAS language questions at arrival and in year one are:

H = 1 if unable
H = 2 if with difficulty
H = 3 if fluently (easily) or almost fluently.

At the third year interview the response categories are:

H = 1 if unable
H = 2 if with major difficulty
H = 3 if with difficulty
H = 4 if with minor difficulty
H = 5 if fluently (easily) or almost fluently.

10 In the vast majority of cases there is no difference between country of birth and country of emigration, and in those cases in which there was a prior international migration there are no data on when this occurred. Country of emigration is used here.

11 The 1972 Census provides a sufficiently large sample and by limiting the analysis to males avoids potential confounding effects of pooling males and females. Males and females are pooled in the IAS because of the relatively smaller sample sizes. It

is found in the IAS that Hebrew language proficiency is significantly lower for females after one and three years in Israel, but not at the time of immigration.

12 It is assumed that there are no differences among countries within the country groupings used in this chapter.

13 The variable label GRXXX refers to country group XXX while LRYYY refers to Language YYY.

14 The "closeness" of Arabic to Hebrew compared with other languages is consistent with the language families developed by linguists (Crystal, 1987; Grimes and Grimes, 1993). Hebrew is classified in the Hamito-Semitic (Afro-Asiatic) family. The Hamito-Semitic family has six branches, of which one is Semitic or Afrosemitic. Semitic includes Hebrew, Arabic and Amharic, among other languages. The majority of immigrants to Israel are from the Indo-European language group. The major branches in this group include Germanic (German, Yiddish, English, Dutch, among others), Italic (which includes the Romance languages, such as French, Spanish, Italian and Romanian), and Balto-Slavic (which includes Russian) languages. The Indo-European language family also includes an Indo-Iranian branch (which includes Persian, Kurdish and Hindi/Urdu, among others). The Uralic languages, originally from the Ural mountains, form a separate family which includes Hungarian. Thus, according to the historical evoluation of languages as discerned by linguists, Hebrew is linguistically closest to Arabic and Amharic. There were very few Amharic speakers in Israel in the 1970s.

15 Statistical tests did not reveal significant interaction effects between sex and variables for marital status and children.

16 Lower rates of destination language proficiency are also found for women in the US and Australia both overall and when other measured variables are held constant. Greater proficiency is associated with higher earnings and enhanced female labour supply. Part of the sex difference in proficiency appears to be due to the lower labour market attachment of women. See, for example, Chiswick and Miller (1994b) and Stevens (1986).

References

Beenstock, M.

1996a "Failure to absorb: remigration by immigrants into Israel", *International Migration Review*, 30: 950–978.

1996b "The acquisition of language skills by immigrants: the case of Hebrew in Israel", *International Migration*, 34: 3–30.

Beenstock, M., and Y.B. Menahem

1997 "The labour market absorption of CIS immigrants to Israel: 1989–1994", *International Migration*, 35: 187–224.

Berman, E., K. Lang, and E. Siniver

1999 "Language skill complementary: returns to immigrant language acquisition", Falk Institute discussion paper, Falk Institute for Economic Research in Israel, Hebrew University, Jerusalem, Israel.

Central Bureau of Statistics

1970 Various years, *Immigrant Absorption Surveys*, Jerusalem, Israel.

1972 *1972 Census of Israel*, 20 per cent sample of the population, microdata file, Jerusalem, Israel.

Chiswick, B.R.

1978 "The effect of Americanization on the earnings of foreign-born men", *Journal of Political Economy*, 86(5): 897–921.

1998 "Hebrew language usage: determination and effects on earnings among immigrants in Israel", *Journal of Population Economics*, 11(2): 253–271.

Chiswick, B.R., and P.W. Miller

1992 "Language in the immigrant labour market", in B.R. Chiswick (Ed.), *Immigration, Language and Ethnicity: Canada and the United States*, American Enterprise Institute, 1992, Washington: 229–296.

1994a "Language choice among immigrants in a multi-lingual destination", *Journal of Population Economics*, 7(2): 119–131.

1994b "Language and labour supply: the role of gender among immigrants in Australia", *Research on Economic Inequality*, 5: 153–189.

1995 "The endogeneity between language and earnings: international analyses", *Journal of Labor Economics*, 13(2): 247–287.

1998 "English language fluency among immigrants in the United States", *Research in Labor Economics*, 17: 151–200.

1999a [2002] "Immigrant earnings: language skills, linguistic concentrations and the business cycle", Center for the Study of the Economy and the State, working paper no. 152, June, University of Chicago, Chicago, (Republished under the same title in *Journal of Population Economics*, 15(1): 31–57).

1999b "Language practice among immigrants in Canada", Center for the Study of the Economy and the State, working paper no. 149, February, University of Chicago, Chicago.

1999c "Immigration, language and multiculturalism in Australia", *Australian Economic Review*, 32(4): 369–385.

Chiswick, B.R., and G.L. Repetto

2001 "Immigrant adjustment in Israel: literacy and fluency in Hebrew and earnings", *Papers in Jewish Demography*, edited by S. DellaPergola and J. Evens, Jerusalem: Hebrew University, 351–378.

Crystal, D.

1987 *The Cambridge Encyclopedia of Language*, Cambridge University Press, Cambridge.

Espenshade, T.J., and H. Fu

1997 "An analysis of English language proficiency among U.S. immigrants", *American Sociological Review*, 62: 288–305.

Espinoza, K.E., and D.S. Massey

1997 "Determinants of English proficiency among Mexican migrants to the United States", *International Migration Review*, 31(1): 28–50.

Grimes, J.E., and B.F. Grimes

1993 *Ethnologue: Languages of the World*, Thirteenth Edition, Summer Institute of Linguistics, Dallas.

Stevens, G.

1986 "Sex differences in language shift in the United States", *Social Science Research*, 71(1): 31–36.

20 Linguistic distance

A quantitative measure of the distance between English and other languages

Introduction

This chapter is concerned with the issue of "linguistic distance", that is, the extent to which languages differ from each other. Although the concept is well known among linguists, the prevailing view is that it cannot be measured. That is, no scalar measure can be developed for linguistic distance.

The next section of this chapter discusses the concept of linguistic distance. The third section presents and discusses a scalar measure of the distance of other languages from English, based on the ease or difficulty Americans have in learning these other languages. The fourth section gives an application of this measure of linguistic distance to understanding the determinants of English language proficiency among adult immigrants from non-English speaking origins in the USA and Canada.[1] The final section gives a summary and conclusion.

Linguistic distance

Studies of immigrant adjustment to the language of the host or destination country indicate that this adjustment differs significantly and substantially by country of origin, even after controlling statistically for the immigrant's personal (socioeconomic and demographic) characteristics.[2] That is, immigrants from some countries of origin appear to be less proficient in the dominant language of the destination than do other immigrants, even when other measured variables are held constant. To some extent this may be due to different incentives for investing in destination language skills, such as the likelihood of temporary or permanent return migration, the availability of access to language training programmes in the destination, or access to the destination language in the origin prior to migration. It would be expected, for example, that destination language skills would be greater among the foreign born if they did not expect to return to their origin, if they had access to destination language training in the destination and if they were exposed to the destination language in schools, in the media or in the marketplace in the origin prior to migration.

Another reason why immigrant groups differ in their proficiency may be differences in the "distance" between the various immigrant languages and the destination language.[3] If English is linguistically "closer" to Western European languages (such as French and German) than it is to East Asian languages (such as Korean and Japanese), it would be expected that Western European immigrants in the USA, UK, Canada and Australia would attain a higher level of proficiency in English, and would attain any given level of proficiency sooner, than immigrants from East Asia (see, for example, Corder, 1981: 95–102).

Languages are complex. They differ in vocabulary, grammar, written form, syntax and myriad other characteristics. This makes for difficulty in the construction of measures of linguistic distance. Even if one intuitively "knows" that English is closer to French than it is to Chinese, by how much is it closer? If the difference is "large", how large is "large"? (McCloskey, 1998: 104–106). While it is easy to rank French as closer to English than Chinese is to English, other rankings of closeness to English may be more difficult, such as between Arabic and Russian or between Chinese and Japanese.

Linguists have developed models of the origins of languages and these models are expressed as "language trees". "The main metaphor that is used to explain the historical relationship is that of the language *family* or *family tree*." (Crystal, 1987: 292, italics in original). Through a language tree one may, in principle, trace the evolution of languages although linguists differ in their construction of language trees. Through a language tree it is possible to "trace" modern English back to its origins, but there is no measure of how different modern English is from its predecessor languages (Old English), other branches on the same tree (modern German), or even from languages on other trees (Chinese). While language trees are useful, they may be a poor guide to the qualitative distance across languages, and do not provide a quantitative measure.

A knowledge of linguistic distance may be invaluable for understanding differences across groups in the acquisition of destination language skills by adult and child immigrants, among participants in language training programmes (such as English as a second language or English for special purposes in the USA or abroad), or the linguistic issues facing indigenous linguistic minorities (e.g. indigenous language-speaking peoples in Africa or Latin America), and the complexity of adaptation in multilingual societies (e.g. India and New Guinea).

Crystal (1987: 371), in *The Cambridge Encyclopedia of Language*, writes regarding linguistic distance:

> The structural closeness of languages to each other has often been thought to be an important factor in FLL (foreign language learning). If the L2 [the foreign language] is structurally similar to the L1 [the original language], it is claimed, learning should be easier than in cases where the L2 is very different. However, it is not possible to correlate linguistic

> difference and learning difficulty in any straightforward way, and even the basic task of quantifying linguistic difference proves to be highly complex, because of the many variables involved.[4]

The many variables being the myriad characteristics that make up the structure of languages.

It has been shown that "linguistic distance" affects the choice of destination among immigrants, and the language they adopt in multilingual destinations. For example, Chiswick and Miller (1994) show that immigrants to Canada are more likely to settle in Quebec if they came from a Romance language country rather than from a country with another mother tongue. Moreover, among immigrants in Quebec those from Romance language countries are more likely to become French speakers while those from other (non-English) linguistic origins are more likely to become English language speakers.

Beenstock *et al.* (2001) show that among Jewish immigrants in Israel, those whose origin language was Arabic are the most proficient in Hebrew, other variables being the same. They suggest that this is due to the short linguistic distance between Hebrew and Arabic. Hebrew and Arabic, along with Amharic, are part of the Semitic branch of the Hamito-Semitic (Afro-Asiatic) family (Crystal, 1987; Grimes & Grimes, 1993). Among the languages included in the Israel analysis, Arabic is the closest to Hebrew. (The data were from the 1970s and there were negligible numbers of Ethiopian Jews in Israel at that time). A measure of linguistic distance from Hebrew comparable to the measure reported here for English has not yet been developed.

This chapter reports a scalar or quantitative measure of the distance between English and a set of other languages. The value of this scalar measure of 'linguistic distance' is demonstrated through an analysis of the determinants of English language proficiency among immigrants in two predominantly English-speaking immigrant receiving countries, the USA and Canada. The chapter ends with a summary and conclusion.

Measuring linguistic distance

The quest among linguists for a scalar measure of linguistic distance has been in vain. There is no yardstick for measuring distances between or among languages, as there is for the geographic distance between countries (e.g. miles). This arises because of the complexity of languages, which differ by vocabulary, grammar, syntax, written form, etc. The distance between two languages may also depend on whether it is in the written or spoken form. For example, the written form of Chinese does not vary among the regions of China, but the spoken languages differ sharply. Alternatively, two languages that may be close in the spoken form may differ more sharply in the written form (for example, if they use different alphabets, as in the case of German and Yiddish).

Perhaps the way to address the distance between languages is not through language trees which trace the evolution of languages, but by asking a simpler question: how difficult is it for individuals who know language A to learn languages B_1 through B_i, where there are *i* other languages. If it is more difficult to learn language B_1 than it is to learn language B_2 it can be said that language B_1 is more "distant" from A than language B_2.[5] Language B_3 may be as difficult to learn as is language B_1 for a language A speaker, but that does not mean that language B_3 is close to language B_1. Indeed, it may be further from B_1 than it is from A.

Alternatively, if the issue is the adjustment of immigrants speaking languages B_1 through B_i in the linguistic destination A, one would want to know how difficult it is for speakers of B_1 through B_i to learn language A.[6] The US Department of State, School of Language Studies teaches English-speaking Americans a variety of languages spoken in all corners of the globe, other than Native American (American Indian) languages. Achievements in speaking proficiency in these languages are then measured at regular intervals. For the same number of weeks of instruction a lower score represents less language facility, and it is assumed that this means a greater distance between the language and English. On the basis of the assumption of linguistic symmetry, this provides a measure of the linguistic distance between English and a variety of other languages.

The paper by Hart-Gonzalez and Lindemann (1993) reports language scores for 43 languages for English-speaking Americans of average ability after set periods (16 weeks and 24 weeks) of foreign language training. These languages are reported in the stub of Table 20.1, with their matching Census of Population Public Use Microdata Sample (PUMS) language codes for the 1990 and 2000 Censuses reported in the "direct code" column. Using the *Ethnologue Language Family Index* published by Grimes and Grimes (1993), the right-most column indicates the linguistic score for that language after 24 weeks of instruction based on the Hart-Gonzalez and Lindemann (1993) report. The range is from a low score (harder to learn) of 1.00 for Japanese to a high score (easier to learn) of 3.00 for Afrikaans, Norwegian and Swedish. The score for French is 2.50 and for Mandarin 1.50. These scores suggest a ranking of linguistic distance from English among these languages: Japanese being the most distant, followed by Mandarin, then French and then Afrikaans, and Norwegian and Swedish as the least distant.

The data on language scores is extended to a much longer list of languages in the column labelled "close codes" (Grimes & Grimes, 1993). To as great an extent as possible, languages (other than Native American languages) not on the original list were matched with the assistance of a linguist for linguistic "closeness" to languages on the list.[7] Thus, Frisian (census code 612) is matched to Dutch (census code 610) which has a linguistic score of 2.75, and Icelandic and Farolse (census codes 617 and 618, respectively) are matched to Norwegian with a linguistic score of 3.00.

Language scores are reported in Table 20.1 for a wide range of languages that are spoken by foreign-born and native-born segments of the population

Table 20.1 Index of difficulty of learning a foreign language (language scores) and codes for languages reported in the US Census

Language	*Direct codes 1990, 2000 Censuses*	*Close codes 1990 Census*	*Changes for 2000 Census*	*Language score*
Afrikaans	611			3.00
Danish	615			2.25
Dutch	610	612		2.75
French	620	621, 622, 623, 624		2.50
German	607	608, 609, 613		2.25
Italian	619			2.50
Norwegian	616	617, 618		3.00
Portuguese	629	630		2.50
Rumanian	631	632		3.00
Spanish	625	626, 627		2.25
Swedish	614			3.00
Indonesian	732	730–731, 733–737		2.00
Malay	739			2.75
Swahili	791	792		2.75
Amharic	780			2.00
Bengali	664			1.75
Bulgarian	647	648		2.00
Burmese	717			1.75
Czech	642			2.00
Dari	660			2.00
Farsi	656	657, 658, 659, 661		2.00
Finnish	679	680		2.00
Greek	637			1.75
Hebrew	778			2.00
Hindi	663	662, 665–669, 678	Add 671	1.75
Hungarian	682			2.00
Lao	720			1.50
Cambodian	726			2.00
Mongolian	694	695, 716		2.00
Nepali	674			1.75
Polish	645	644, 646		2.00
Russian	639	640, 641		2.25
Serbo-Croatian	649–651	652		2.00
Sinhala	677			1.75
Tagalog	742	740, 741, 743–749		2.00
Thai	720	717, 718, 719	Add 725	2.00
Turkish	691	689, 690, 692, 693		2.00
Vietnamese	728	729		1.50
Arabic	777	779		1.50
Mandarin	712	713, 714, 715		1.50
Japanese	723	725	Delete 725	1.00
Korean	724			1.00
Cantonese	708	709, 710, 711, 721, 722		1.25

Note: Language codes in this table are from the 1990 US Census of Population and Housing, Technical Documentation and from the 2000 US Census of Population and Housing, Technical Documentation. There are minor differences in the language codes in the 1990 and 2000 Censuses. These differences are indicated in column (3). Column (4) is the language score for the direct codes.

Source of matching codes: (a) Grimes and Grimes (1993), (b) Adam Makkai, Professor of Linguistics, Department of English, University of Illinois at Chicago.
Source of Language Score: Hart-Gonzalez and Lindemann (1993).

in the USA. These scores can then be used to do statistical analyses of language issues.

Application of the measure of linguistic distance

This section reports the application of the measure of linguistic score in Table 20.1 to the analysis of proficiency in English among immigrants in the USA and Canada.

Using ordinary least-squares regression analysis (OLS), Table 20.2 reports the partial effects of "linguistic distance" on the English language proficiency of foreign-born adult male and female immigrants in the USA from non-English speaking countries, using data from the 1990 Census of Population. The linguistic distance (LD) is measured in this analysis as the inverse of the linguistic score (LS) in Table 20.1, that is, LD = 1/LS. The "other variables held constant" include years of schooling, age and its square, duration in the USA and it square, marital status, a minority language concentration measure in the region of residence specific to the respondent's minority language, urban/rural residence and a south/nonsouth region variable. Other variables being the same, LD is a highly statistically significant variable for both men and women. Going from Swedish to Japanese (LD = 0.33 to LD = 1.0)

Table 20.2 Partial effect of linguistic distance on the English language proficiency of foreign-born adults from non-English speaking countries, 1990 US Census[a]

	Males	*Females*
Other variable held constant[b]	–0.256	–0.263
	(–44.91)	(–51.95)
Other variables and distance of foreign country from the USA in miles, and its square[b]	–0.319	–0.320
	(–53.34)	(–60.10)
Other variables, distance in miles and its square, and country fixed effects[b,c]	0.007	–0.063
	(0.56)	(–4.39)

Notes: Sample size 237,770 for males and 243,496 for females. *t*-ratios are in parentheses.

a The measure of Linguistic Distance (LD) is the inverse of the Linguistic Score (LS) in Table 20.1. That is LD = 1/LS from Table 20.1. The dependent variable is unity if the respondent speaks only English at home or, if another language is spoken, English is spoken "very well" or "well". It is zero for those who speak English "not well" or "not at all". The foreign-born excludes those born in the English-speaking developed countries (UK, Ireland, Canada, Australia and New Zealand). Adults are persons aged 25–64 in 1990. Where only English is spoken at home, and hence a non-English language is not reported, LS is the mean value of the linguistic score measure for individuals reporting a foreign language from their birthplace group.

b Other variables held constant include years of schooling, age (and its square), duration of residence in the USA (and its square), marital status, an index of the extent to which their origin language is spoken in their state of residence and variables for urban/rural and south/nonsouth residence.

c Country fixed effects represented by 16 country/region of birth dichotomous variables.

Source: Chiswick and Miller (1998, Tables 2 and 6).

reduces the probability of being proficient in English by about 17 percentage points ($0.26 \times 0.67 = 0.174$), or the equivalent effect of about 5.4 years of additional schooling. The effect is larger (0.214) when the geographic distance (measured in miles) from the origin to the USA is also held constant. The partial regression coefficient and the *t*-ratio for the linguistic distance effect diminishes sharply (and disappears for men, but not for women) when country of origin is held constant through a set of dichotomous variables. This arises in large part because of the close correspondence of language and country—Korean is spoken in Korea, Italian in Italy, etc.

The linguistic distance measure was also applied to an analysis of English or French language proficiency among adult male immigrants in Canada from non-English speaking countries (Table 20.3).[8] Other variables being the same, the greater the linguistic distance, the less likely is the immigrant to speak English, or if the immigrant speaks English, the less likely he is to speak English at home. At a duration in Canada of five years, only one quarter (25%) of immigrants with the greatest linguistic distance (LS = 1.0, Korean and Japanese) can carry on a conversation in English or French, in contrast to 5% for those with the smallest linguistic distance (LS = 3.0 Afrikaans, Swedish, Norwegian).

Even after 15 years in Canada, the ability to carry on a conversation in English or French varies by linguistic distance. Fully 10% of those with the greatest origin language linguistic distance cannot do so, compared to only 1% for those with the smallest distance. By 15 years in Canada only 5% of those with the greatest linguistic distance in their origin language usually speak English or French at home, in contrast to 58% for those with the

Table 20.3 Predicted distributions across language categories by linguistic score and duration of residence, foreign-born adult males from non-English speaking countries, 1991 Census of Canada[a]

Linguistic score[b]	*After 5 years in Canada*			*After 15 years in Canada*		
	E1	**E2**	**E3**	**E1**	**E2**	**E3**
1.0	24.54	73.88	1.58	10.05	85.35	4.60
2.0	8.00	73.75	18.25	2.32	60.18	37.51
3.0	4.57	61.18	34.24	1.09	41.05	57.86

Notes: Sample size 32,168.

a Predicted values from a multinomical logit model. Adults are aged 25–64 in 1990, foreign-born exclude those born in the USA, UK and Ireland.
E1 = Cannot carry on a conversation in English or French.
E2 = Can carry on a conversation in English or French, but usually speak another language at home.
E3 = Can carry on a conversation in English or French and usually speak one of these languages at home.

b Language scores range from 1.0 (Japanese and Korean) to 3.0 (Afrikaans, Norwegian and Swedish).

Source: Chiswick and Miller (2001, Table 5).

shortest origin language distance. Thus, the linguistic patterns of immigrants in Canada, even after living there for 15 years, are influenced strongly by the distance between their origin language and English.

Summary and conclusion

This chapter develops and tests a scalar or quantitative measure of "linguistic distance". Although linguists are familiar with the concept of the distance among the myriad characteristics of languages, the prevailing view is that it cannot be measured or quantified. This chapter develops and tests such a measure.

The measure developed here is based on the ability of Americans to learn a variety of languages in fixed periods of time. The lower the scores on a standardised proficiency test, the greater is the distance between these languages and English. With the aid of a linguist, scores are inferred for languages for which a direct measure does not exist.

The measure of linguistic distance was then used in analyses of the English language proficiency of adult immigrants in the USA and Canada from non-English language origins, using census microdata. It is found empirically that the greater the distance between an immigrant's origin language and English, the lower is the level of the immigrant's English language proficiency, when other relevant variables are the same.

The measure of linguistic distance developed here can be used for other purposes. It can, for example, be used for research, evaluation, planning and diagnostic analyses for understanding the determinants of English language proficiency, in general or for specific purposes, among non-English speaking individuals, whether they are immigrants, non-English speaking linguistic minorities or learning in their country of origin.

The measure may also be useful for explaining patterns of international migration (i.e. choice of destination among immigrants), language adopted in multilingual destinations and patterns of flows of tourists.[9] The measure can also be applied to other forms of analysis. Hutchinson (2002), for example, uses the linguistic distance measure developed for this chapter in an analysis of international trade. He finds that, holding other relevant variables constant, a greater linguistic distance between the USA and other countries reduces both imports from and exports to the USA.

The methodology used here can, in principle, be developed for languages other than English. Thus, it would be possible to develop scalar measures of linguistic distance for other languages. This can permit the development of a full range of measures of linguistic distance.

Acknowledgements

B.R. Chiswick acknowledges the research support of the Institute of Government and Public Affairs, University of Illinois. P.W. Miller acknowledges the financial support of the Australian Research Council.

Notes

1 For an analysis of the determinants of second language acquisition from an economist's perspective, see Chiswick and Miller (1998), and from a linguist's perspective, see Ellis (1994).

2 For example, these studies have been conducted for the USA (Chiswick & Miller, 1998), Australia (Chiswick & Miller, 1995, 1996), Canada (Chiswick & Miller, 2001; Grenier & Vaillancourt, 1983), Germany (Dustmann, 1997), Israel (Beenstock, 1996; Chiswick, 1998) and the UK (Dustmann & Fabbri, 2003; Shields & Wheatley Price, 2002).

3 The story in Genesis about the Tower of Babel emphasises the difficulty of working cooperatively when there is a lack of communication among individuals based on differences in languages.

4 In their study of the English language proficiency of immigrants in the UK, Shields and Wheatley Price (2002: 145) indicate that their theoretical model calls for a measure of linguistic distance of the immigrants' origin language from English, but they do not have a direct measure and they use country of birth dichotomous variables to reflect this and other origin-specific effects.

5 Think of each language as having n dimensions, where the n dimensions represent the various aspects of language (Crystal, 1987: 371). Then each language can be thought of as being represented by a point in n-dimensional space, and could be described by a vector $(a^1, a^2, \ldots, a^n)$, $(b^1{}_1, b^2{}_1, \ldots, b^n{}_1)$, etc, where a^n is the amount of the nth dimension that characterises language A, and $b^n{}_i$ is the amount of the n th dimension that characterises language B_i, etc. The distance between any two languages is given by the Euclidean distance function. The measure of linguistic distance proposed in this paper can be thought of as a proxy for the Euclidean distance between language A and the various languages B_1 through B_i.

6 While the linguistic difference between A and, say, B_i, is a given magnitude, the impact of that measure of distance on proficiency in language B_i, among language A speakers may differ from the impact of the distance on language B_i speakers learning language A.

7 We are indebted to Adam Makkai, Professor of Linguistics, Department of English, University of Illinois at Chicago for helping us with this coding.

8 Those born in France cannot be separately identified in the Canadian Census due to the small number of immigrants from France.

9 Other variables being the same, tourist flows would be expected to be greater the smaller the linguistic distance between the languages of the origin and tourist destination.

References

Beenstock, M. (1996) The acquisition of language skills by immigrants: The case of Hebrew in Israel. *International Migration* 3, 3–30.

Beenstock, M., Chiswick, B.R. and Repetto, G.L. (2001) The effect of linguistic distance and country of origin on immigrant language skills: Application to Israel. *International Migration* 39 (3), 33–60.

Chiswick, B.R. (1998) Hebrew language usage: Determinants and effects on earnings among immigrants in Israel. *Journal of Population Economics* 11 (2), 253–271.

Chiswick, B.R. and Miller, P.W. (1994) Language choice among immigrants in a multilingual destination. *Journal of Population Economics* 7, 119–131.

Chiswick, B.R. and Miller, P.W. (1995) The endogeneity between language and earnings: International analyses. *Journal of Labor Economics* 13, 246–288.

Chiswick, B.R. and Miller, P.W. (1996) Ethnic networks and language proficiency among immigrants. *Journal of Population Economics* 9, 19–35.

Chiswick, B.R. and Miller, P.W. (1998) English language fluency among immigrants in the United States. *Research in Labor Economics* 17, 151–200.

Chiswick, B.R. and Miller, P.W. (2001) A model of destination language acquisition: Application to male immigrants in Canada. *Demography* 38 (3), 391–409.

Corder, S.P. (1981) *Error Analysis and Interlanguage*. Oxford: Oxford University Press.

Crystal, D. (1987) *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.

Dustmann, C. (1997) The effects of education, parental background and ethnic concentration on language. *Quarterly Review of Economics and Finance* 37, 245–262.

Dustmann, C. and Fabbri, F. (2003) Language proficiency and labour market performance of immigrants in the UK. *Economic Journal* July, 695–717.

Ellis, R. (1994) *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Grenier, G. and Vaillancourt, F. (1983) An economic perspective on learning a second language. *Journal of Multilingual and Multicultural Development* 4, 471–483.

Grimes, J.E. and Grimes, B.F. (1993) *Ethnologue: Languages of the World* (13th edn). Dallas: Summer Institute of Linguistics, Inc.

Hart-Gonzalez, L. and Lindemann, S. (1993) Expected achievement in speaking proficiency, 1993. School of Language Studies, Foreign Services Institute, Department of State (mimeo).

Hutchinson, W.K. (2002) Linguistic distance as a determinant of bilateral trade. Department of Economics, Vanderbilt University, Xerox.

McCloskey, D. (1998) *The Rhetoric of Economics* (2nd edn). Madison: University of Wisconsin Press.

Shields, M.A. and Wheatley Price, S. (2002) The English language fluency and occupational success of ethnic minority immigrant men living in English metropolitan areas. *Journal of Population Economics* 15 (1), 137–160.

US Bureau of the Census (1993) 1990 United States Census of Population and Housing. Technical Documentation, Washington, D.C.

US Bureau of the Census (2003) 2000 United States Census of Population and Housing. Technical Documentation, Washington, D.C.

Bibliography of language research

Papers are authored by Barry R. Chiswick and Paul W. Miller, except where indicated otherwise.

"Language Skills and Immigrant Adjustment: What Immigration Policy Can Do", in Deborah Cobb-Clark and Siew-Ean Khoo (eds.), *Public Policy and Immigrant Settlement*, Cheltenham, UK: Edward Elgar Publishing, 2006, pp. 121–148.

"Parents and Children Talk: English Language Proficiency within Immigrant Families" (with Yew Liang Lee), *Review of Economics of the Household*, **3**(3), 2005, pp. 243–268.

"The Linguistic and Economic Adjustment of Soviet Jewish Immigrants in the United States, 1980–2000" (by Barry R. Chiswick and Michael Wenz), *Research in Labor Economics*, **24**, 2006, pp. 179–216. Reprinted in Yaacon Ro'i, *et al.* (eds.) *Soviet and Post-Soviet Jewry: Essays in Honor of the Retirement of Prof. M. Altsuler* (forthcoming).

"Immigrants' Language Skills and Visa Category" (with Yew Liang Lee), *International Migration Review*, **40**(2), Summer 2006, pp. 419–450.

"A Longitudinal Analysis of Immigrant Occupational Mobility: A Test of the Immigrant Assimilation Hypothesis" (with Yew Liang Lee), *International Migration Review*, **39**(2), Summer 2005, pp. 332–353.

"Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages", *Journal of Multilingual and Multicultural Development*, **26**(1), 2005, pp. 1–11.

"Family Matters: The Role of the Family in Immigrant's Destination Language Skills" (with Yew Liang Lee), *Journal of Population Economics*, **18**(1), 2005, pp. 631–647.

"Do Enclaves Matter in Immigrant Adjustment?", *City and Community*, **4**(1), March 2005, pp. 5–35. An earlier version with the same title was published in Hebrew in *Revon LeKalkalah (The Economic Quarterly)*, **50**(2), June 2003, pp. 285–325.

"Immigrants' Language Skills: The Immigrant Experience in a Longitudinal Survey" (with Yew Liang Lee), *Annales d'Economie et de Statistique*, Special Issue on Discrimination and Unequal Outcomes, **71–72**, July/December 2003, pp. 97–139.

"Schooling, Literacy, Numeracy and Labour Market Success" (with Yew Liang Lee), *Economic Record*, **79**(245), June 2003, pp. 165–181.

"The Complementarity of Language and Other Human Capital: Immigrant Earnings in Canada", *Economics of Education Review*, **22**(5), 2003, pp. 469–480.

"Immigrant Earnings: Language Skills, Linguistic Concentrations and the Business Cycle", *Journal of Population Economics*, **15**(1), January 2002, pp. 31–57. Reprinted in Klaus F. Zimmermann and Amelie Constant (eds.), *How Labor Migrants Fare*, Berlin: Springer-Verlag, 2004, pp. 223–250, and in Barry R. Chiswick (ed.), *The Economics of Immigration*, 2005, Cheltenham, UK: Edward Elgar Publishing, pp. 71–97.

"A Model of Destination-Language Acquisition: Application to Male Immigrants in Canada", *Demography*, **38**(3), August 2001, pp. 391–409.

"The Effect of Linguistic Distance and Country of Origin on Immigrant Language Skills: Application to Israel", (by Barry R. Chiswick, Michael Beenstock and Gaston L. Repetto), *International Migration*, **39**(3), September 2001, pp. 33–60.

"Immigrant Adjustment in Israel: The Determinants of Literacy and Fluency in Hebrew and the Effects on Earnings" (by Barry R. Chiswick and Gaston L. Repetto), in Slobodan Djajic (ed.), *International Migration: Trends, Policy and Economic Impact*, New York: Routledge, 2001, pp. 204–228, and in Sergio DellaPergola and Judith Even (eds.), *Papers in Jewish Demography*, Jerusalem: Hebrew University, 2001, pp. 351–378.

The Economic Cost to Native-Born Americans of Limited English Language Proficiency, Report prepared for the Center for Equal Opportunity, August 1998.

"Indigenous Language Skills and the Labor Market in a Developing Economy: Bolivia" (by Barry R. Chiswick, Harry A. Patrinos and Michael E. Hurst), *Economic Development and Cultural Change*, **48**(2), January 2000, pp. 349–367.

"Immigration, Language and Multiculturalism in Australia", *Australian Economic Review*, **32**(4), December 1999, pp. 369–385.

"Census Language Questions in North America", *Journal of Economic and Social Measurement*, **25**(2), 1998, pp. 73–95.

"Language Practice and the Economic Well-Being of Immigrants", *Policy Options/Options Politiques*, **20**(4), May/Mai 1999, pp. 45–50.

"Language Skills and Earnings Among Legalized Aliens", *Journal of Population Economics*, **12**(1), February 1999, pp. 63–89. Reprinted in Klaus F. Zimmermann and Amelie Constant (eds.), *How Labor Migrants Fare*, Berlin: Springer-Verlag, 2004, pp. 279–306.

"Language Skill Definition: A Study of Legalized Aliens", *International Migration Review*, **32**(4), Winter 1998, pp. 877–900.

"English Language Fluency Among Immigrants in the United States", *Research in Labor Economics*, **17**, 1998, pp. 151–200.

"Hebrew Language Usage: Determinants and Effects on Earnings Among Immigrants in Israel" (by Barry R. Chiswick), *Journal of Population Economics*, **11**(2), May 1998, pp. 253–271.

The Economics of Language: Application to Education (by Barry R. Chiswick and Harry A. Patrinos), Training Manual, Human Development Network, World Bank, July 1996, Revised January 1998.

"The Economics of Language: The Roles of Education and Labor Market Outcomes" (by Barry R. Chiswick), *Human Capital Development Working Papers* (HCDWP 70), World Bank, September 1996.

"Soviet Jews in the United States: Language and Labor Market Adjustments Revisited", (by Barry R. Chiswick), in Noah Lewin-Epstein, Yaacov Roi and Paul Ritterband (eds.), *Russian Jews on Three Continents: Migration and Resettlement*, London: Frank Cass Publishers, 1997, pp. 233–260. Reprinted in Elite Olshtain and Gabriel Horenczyk (eds.), *Language, Identity and Immigration*, Jerusalem: Magnes Press, 2000, pp. 275–300.

"Literacy, Numeracy and the Labour Market", in *Aspects of Literacy: Assessed Skill Levels*, Canberra: Australian Bureau of Statistics, Catalogue No. 4228.0, September 1997, pp. 73–79.

"The Languages of the United States: What is Spoken and What it Means", *READ Perspectives*, Fall 1996, pp. 5–41.

"Ethnic Networks and Language Proficiency Among Immigrants", *Journal of Population Economics*, **9**(1), March 1996, pp. 19–35. Reprinted in Klaus F. Zimmermann and Thomas K. Bauer (eds.), *The Economics of Migration*, Cheltenham, UK: Edward Elgar Publishing, September 2002, Volume II, pp. 518–534.

"The Endogeneity Between Language and Earnings: International Analyses", *Journal of Labor Economics*, **13**(2), April 1995, pp. 246–288. Reprinted in Klaus F. Zimmermann and Thomas K. Bauer (eds.), *The Economics of Migration*, Cheltenham, UK: Edward Elgar Publishing, September 2002. Volume II, pp. 475–517, and in Donald M. Lamberton (ed.), *The Economics of Language*, Cheltenham, UK: Edward Elgar Publishing, 2003, pp. 198–242.

"Language and Earnings Among Immigrants in Canada: A Survey", in Sally Zerker (ed.), *Change and Impact: Essays in Canadian Social Sciences*, Magnes Press, 1994, pp. 247–264. Reprinted in Harriet Duleep and Phanindra Wunnava (eds.), *Immigrants and Immigration Policy: Individual Skills, Family Ties and Group Identities*, Greenwich, CT: JAI Press, 1996, pp. 39–56.

"Language and Labor Supply: The Role of Gender Among Immigrants in Australia", *Research on Economic Inequality*, **5**, 1994, pp. 153–189.

"Language Choice Among Immigrants in a Multi-Lingual Destination", *Journal of Population Economics*, **7**(2), 1994, pp. 119–131.

"Soviet Jews in the United States: An Analysis of their Linguistic and Economic Adjustment" (by Barry R. Chiswick), *International Migration Review*, **27**(2), Summer 1993, pp. 260–285, and published in Hebrew in *Revon LeKalkalah* (*The Economic Quarterly*), July 1991, No 148, pp. 188–211.

"Language in the Immigrant Labor Market", in Barry R. Chiswick (ed.), *Immigration, Language and Ethnicity: Canada and the United States*, Washington: American Enterprise Institute, 1992, pp. 229–296, and 471–476.

Immigration, Language and Ethnicity: Canada and the United States (edited by Barry R. Chiswick), Washington: American Enterprise Institute, 1992.

"Speaking, Reading, and Earnings Among Low-Skilled Immigrants" (by Barry R. Chiswick), *Journal of Labor Economics*, **9**(2), April 1991, pp. 149–170. Reprinted in Klaus F. Zimmermann and Thomas K. Bauer (eds.), *The Economics of Migration*, Cheltenham, UK: Edward Elgar Publishing, September 2002, and in Barry R. Chiswick (ed.), *The Economics of Immigration*, 2005, Cheltenham, UK: Edward Elgar Publishing, pp. 296–317.

"Earnings in Canada: The Roles of Immigrant Generation, French Ethnicity, and Language", *Research in Population Economics*, **6**, 1988, pp. 183–228.

Index