Testing the Untestable in Language Education

NEW PERSPECTIVES ON LANGUAGE AND EDUCATION
Series Editor: Professor Viv Edwards, *University of Reading, Reading, Great Britain*
Series Advisor: Professor Allan Luke, *Queensland University of Technology, Brisbane, Australia*

Two decades of research and development in language and literacy education have yielded a broad, multidisciplinary focus. Yet education systems face constant economic and technological change, with attendant issues of identity and power, community and culture. This series will feature critical and interpretive, disciplinary and multidisciplinary perspectives on teaching and learning, language and literacy in new times.

Full details of all the books in this series and of all our other publications can be found on http://www.multilingual-matters.com, or by writing to Multilingual Matters, St Nicholas House, 31-34 High Street, Bristol BS1 2AW, UK.

Testing the Untestable in Language Education

Edited by
Amos Paran and Lies Sercu

Contents

Contributors

Phil Benson is a Professor in the English Department at the Hong Kong Institute of Education. He has published widely on the subject of autonomy, including the book *Teaching and Researching Autonomy in Language Learning* (Pearson, 2001) and a 'State-of-the-Art' paper in *Language Teaching* (2007). His current research interests include language learning histories, popular culture and language learning, and language teacher education.

Leni Dam took the first steps towards developing learner autonomy in language teaching and learning in her own classes at secondary school level in Denmark in the early seventies. Even then, a strong focus was students' evaluation of their own learning. She was later educational adviser at University College, Copenhagen, which provided excellent opportunities for implementing the principles of learner autonomy as well as evaluation into teacher education. From 1993 to 1999, she was co-convenor of the scientific commission of learner autonomy within AILA, and also carried out the research project, Language Acquisition in an Autonomous Learning Environment, together with Lienhard Legenhausen. She has published widely and given numerous lectures and workshops around the world. In 2004, she was awarded an honorary doctor's degree in pedagogy from Karlstad University, Sweden. In 2007, she stepped down from official duties and is now freelance.

Kaisu Korhonen holds a PhD degree in intercultural communication and a Master's degree in business English. She has a long experience in working as principal lecturer at a university of applied sciences in Finland. She has worked with engineering and business students as well as with professionals in various fields, providing intercultural communication and ESP training. She has published two study books for engineering students and a number of papers for conference and professional publications, discussing intercultural competence, its development with

the aid of intercultural communication training, and the effectiveness of such training. Her current interests include the Finnish style of communication.

Terry Lamb works in the School of Education, University of Sheffield. He has authored and edited numerous publications in the areas of learner autonomy, multilingualism and teacher development, and is editor of the *International Journal of Innovation in Language Learning and Teaching*. He has presented keynote papers in conferences around the world. Terry is former President of the Association for Language Learning and current President of FIPLV (Fédération Internationale des Professeurs de Langues Vivantes). In 2008, he was appointed Chair of the Development Partnership for the Diploma in Languages and International Communication by the Secretary of State for Education. He is also Chevalier des Palmes Académiques, an honour awarded by the French Government.

Lienhard Legenhausen received his PhD from Stuttgart University in 1974. He wrote his thesis on *Error Analysis and Error Evaluation* (Berlin: Cornelsen, 1975). A year later, he became a Professor of English and Applied Linguistics at Düsseldorf University. In 1992, he accepted an offer from Münster University where he was responsible for Second Language Acquisition Research and Language Pedagogy. His main research interests include technology-enhanced language learning and the analysis of learner languages. Together with Leni Dam, he started the research project, Language Acquisition in an Autonomous Learning Environment (LAALE). Since 2006, Lienhard Legenhausen has been an Emeritus Professor.

Constant Leung is Professor of Educational Linguistics at King's College London. He is Chair of the MA English Language Teaching and Applied Linguistics programme, and Director of MA Assessment in Education programme in the Department of Education and Professional Studies. His research interests include education in ethnically and linguistically diverse societies, language assessment, second/additional language curriculum development, language policy and teacher professional development. He has written and published widely on issues related to ethnic minority education, additional/second language curriculum, and language assessment nationally and internationally.

Anthony J. Liddicoat is Professor in Applied Linguistics at the Research Centre for Languages and Cultures in the School of Communication, International Studies and Languages at the University of South Australia. His research interests include: language and intercultural issues in education, conversation analysis, and language policy and planning. In recent years, his research has focussed on issues relating to the teaching and learning of culture through language study and his work has contributed to the development of Intercultural Language Teaching and Learning.

Hui-wei Lin received his PhD from University of Nottingham in 2005. He is now Director of the Centre for Learning and Teaching at Chia Nan University of Pharmacy and Science. A specific focus of his current research is how literature renders qualia through stylistic patterning. His research seeks to explore the possibilities for the assessment of learner achievement in literary education, with a concomitant purpose of initiating innovation of the traditional paradigm of literature teaching and literature testing in the ESL/EFL classroom.

Marylin Low is motivated by her concerns about the complexities of teaching and learning (in) multiple languages and literacies, especially in the diverse and rich cultural contexts of Pacific small island nations. Assessment-as-teaching opens to an abundance of knowing that teachers work hard to restrict in their attempt to 'get it right'. Her interest is in breaking open such restrictions and getting in on the real, living conversations that constitute teaching/learning/assessment in everyday classroom life. She is a Senior Specialist at Pacific Resources for Education and Learning, a non-profit organisation that serves the US affiliated states of the Pacific.

Bernard Mohan is Emeritus Professor, Language and Literacy Education Department, University of British Columbia, and Research Fellow, Department of Education and Professional Studies, King's College London. He has published widely on language as a medium of learning and the integration of language and content learning in second language learning, drawing on the theory and analysis of systemic functional linguistics. His longstanding research interest in linguistic aspects of assessment developed when he was a member of the Division of Measurement and Experimental Design, Department of Educational Psychology, University of Calgary. He is currently Chair, North American Systemic Functional Linguistics Association.

Angela Scarino is an Associate Professor of Applied Linguistics and the Director of the Research Centre for Languages and Cultures at the University of South Australia. She has had extensive experience in the conceptualisation, design, implementation and evaluation of curriculum and assessment frameworks for languages learning across all levels of education. She has conducted research, development and evaluation projects in the areas of languages curriculum, assessment and teacher development. Her experience includes working in various states and nationally in Australia, Hong Kong, Singapore, Malaysia and New Zealand.

Tammy Slater is an Assistant Professor of Applied Linguistics/TESL in the English Department at Iowa State University of Science and Technology. Her research, which draws upon analytic methods from systemic functional linguistics, seeks to understand the development of academic language through content-based and project-based teaching and learning, particularly as it informs the field of education for English language learners.

Jane Spiro is Principal Lecturer running International and TESOL MA programmes at Oxford Brookes University. She has previously worked at the College of St. Mark and St. John, Pécs University, Hungary, and Nottingham University, and has been consultant worldwide on projects in materials writing, test writing, curriculum development and creative writing. She was co-editor of the journal *Reading in a Foreign Language*, and judge for the first *Poetry in a Foreign Language* competition. She is the author of *Creative Poetry Writing* (Oxford University Press, 2004) and *Storybuilding* (Oxford University Press, 2007) and is a published poet and novelist with a research/teaching commitment to the connections between creative practitioner and educator. Her PhD on this subject was completed in 2008 and published in full online by the University of Bath.

Chapter 1

More than Language: The Additional Faces of Testing and Assessment in Language Learning and Teaching

AMOS PARAN

Testing and Assessment: A Ubiquitous Phenomenon

'Teaching involves assessment' (Rea-Dickins, 2004: 249). This simple, three-word sentence hides what is in fact a whole world, a world where 'young people in many countries... are now faced with an unprecedented number of exams and tests as they go through school and higher education' (Broadfoot, 2005: 125). It is a world which has been developed into a testing society (Broadfoot, 2005), where standardised testing is a major part of the assessment regime, which in some countries, e.g. the UK, starts as early as the age of seven (for an overview, see Leung & Scott, 2009).

Unsurprisingly, such a ubiquitous phenomenon as testing exerts an extremely powerful influence on its environment; it is now recognised that tests have powerful washback effects, what Cheng and Curtis (2004: 7) call, 'a set of relationships, planned and unplanned, positive and negative, between teaching and testing'. These effects extend throughout the educational system and, indeed, throughout society, becoming, as Shohamy (2007: 120) has argued, *de facto* instruments of language policy: 'since tests are often more powerful than any written policy document, they lead to the elimination and suppression of certain languages in societies... Tests can also be used as tools to privilege certain forms and levels of language knowledge... Thus, language tests, given their power and influence in societies, play a major role in the implementation and introduction of language policies'.

In language teaching, the field of testing and assessment has an additional effect: a number of important models of language competence,

such as Bachman's (1990) model or the Common European Framework of Reference (CEFR) (Council of Europe, 2001) originate in the need to specify language competences for testing and, in the case of the CEFR, finding a way 'to compare the objective and achievement standards of learners in different national (and local) contexts' (Morrow, 2004: 6; see also Alderson, 2004). Thus, our view of language learning and language competence is strongly influenced by our understanding of language testing and assessment.

An additional issue is reflected in the title of this book: the assumption that everything we teach in the language classroom can in fact be tested. Within language education, since often more is taught than only language (see below for a discussion of this point), the case can be made that not only language should be tested.

This volume brings together 12 chapters in which educators from around the globe grapple with issues that arise from these points. In this introductory chapter, I start by looking at some of the recent critiques of policy and practice in language testing, and at some of the responses to the current situation. I then present the four areas in language education on which the present volume focuses, and provide an overview of the different chapters. I end with a discussion of the themes emerging from the different chapters in the book.

Language Tests: A Narrowing Agenda?

In spite of the ubiquity of testing, there is nevertheless 'a widespread perception that the needs of teachers and learners are not currently well served by assessment practice and by assessment research' (McNamara, 2001: 340). The reason for this becomes clear when we consider the way language testing has developed over the last half-century. Spolsky (2008) charts the major trends in language testing, highlighting the dominance of the psychometric approach and the industrialisation of tests. McNamara and Roever (2006: 1) suggest that in language testing, 'psychometrics became the substrate discipline... and language was virtually poured into these preexisting psychometric forms'. Leung and Lewkowicz (2006: 212) voice a similar view, suggesting that 'the form that has been most prevalent in ELT all over the world in the past 50 or more years has been standardised, psychometrically oriented testing'. Leung and Lewkowicz (2006) attribute the tendency towards standardisation in a wish (or requirement) for fairness, but make it clear that a commitment to standardisation will come at the expense of acknowledging differences between test takers.

The drive for achieving standardisation, alignment and conformity has another important consequence: it will almost always come at the expense of broadness of vision. Wall (1997) provides a history of the worry about the narrowing of education as the result of tests, tracing it back to the beginning of the 19th century. In language testing, this trajectory in the history of tests is exemplified in Weir's (2003) fascinating account of the development of the Certificate of Proficiency in English (CPE) during the 20th century. What emerges from Weir's account is how the examination, through its numerous revisions, increasingly focused on language and on language only. With each revision and expansion of the construct of language proficiency, the examination shed aspects that did not reflect this construct, reflecting 'a gradual but critical change of the examination to one of language as against language, literature and culture' (Weir, 2003: 18). This process may well be underway in other countries as well – Eckes *et al*. (2005: 373) seem to imply a criticism of language tests that 'assess far more than language proficiency proper (e.g. they also tap knowledge of German literature, history, and civilisation)'; elsewhere they mention the importance of increasing reliability in the marking of essays on a literary theme in Slovenia. It seems logical to assume that this will be accompanied by a narrowing of the scope of the examination. Thus, 'the psychometric, the industrial and the scaling trends' (Spolsky, 2008: 450) continue to dominate, and they often entail a more focused – and hence narrower – approach to what is tested and how this is done. This development receives an added dimension in Broadfoot's (2005: 135) discussion of 'the inseparability of the affective and cognitive domains in learning', and the resulting implications for testing.

Together, the points that have been presented here – the ubiquity of tests, their powerful washback effect, which can ultimately dictate what is taught in the language classroom, and the narrowing vision of language tests – have meant that one consequence of large-scale language tests has been to circumscribe the content that is taught in the language classroom. The way in which language tests have developed has meant that language education, in many countries, is now concerned with teaching only language. True, developments in our view of language, of communication and of language learning mean that we have a more nuanced and more complex view of the language classroom than it would have been half a century ago. McNamara and Roever (2006), for example, provide an extensive discussion of assessing second language (L2) pragmatics, and Spolsky (2008: 450) does note 'the broadening of the content to include sociolinguistically influenced

aspects of language'. However, the focus is still on language, with an artificial separation between language and content.

At the same time, our understanding of language, language teaching and language learning has moved away from this separation, and includes resistance to a reductionist view of language education where all that is taught is what can be easily tested. This volume is therefore an attempt to bring together and record the endeavours of language educators in the global arena (the 15 contributors to this volume work in 10 different countries) to incorporate assessment and integrate it into four different areas of language and language development: intercultural competence; autonomy; literature; and content teaching.

Current Critiques of Language Testing

For some time now, views of language testing have recognised many of the problems that beset this field. There has been a focus on the social consequences of tests as well as on their use as instruments of power and control, as the passage from Shohamy (2007) quoted above suggests. The critique of language testing has focused on critical views of the uses of testing (Shohamy, 1996, 2005, 2007); there has been a move to use-oriented testing (Shohamy, 2005) and critiques of many of the uses of language tests (McNamara, 2005; McNamara & Roever, 2006). This is linked to ethical issues in testing (see, e.g. Lynch, 2001, for a discussion of ethical issues in different approaches to language testing).

One approach has been alternative assessment (Huerta-Macías, 2002; Fox, 2008), sometimes conceptualised as 'alternatives in assessment' (Brown & Hudson, 1998). Indeed, Shohamy (1996: 144) has suggested that we are now in an 'alternative era'. Birenbaum (1996) views alternative assessment as being grounded in an alternative approach to instruction, and as integrating assessment and instruction. She then lists a number of what she calls 'alternative assessment devices', such as 'authentic performance tasks, simulations, portfolios, journals' and many others (Birenbaum, 1996: 8). There is much talk of the 'ethical dimension in so far as (testing) affects people's lives' (Weir, 2005: 1).

However, the critique of testing and of language tests normally focuses on the teaching and learning activities that are being affected, rather than the content of what is being taught. Shohamy (1996: 150) points out that even Bachman's (1990) elaborate model of communicative competence 'does not account for the domain knowledge in performance testing'. In Weir's (2005: 212) discussion of washback, it is clear that what is at stake (and to some extent, rightly so) is success on the test, and the

washback effects considered are, for example, 'training teachers in the new content and methodology required for the test' and 'support in the forms of appropriate teaching materials'. Even where washback is conceived in a wider context than the traditional sense, what Weir (2005: 213) calls 'the wider impact of the test, its effects on other systems in the administrative and academic contexts of the tests, and on the attitudes and behaviour of the stakeholders in these', we are still talking about the *test* as the main factor.

More than Language

This volume is a response to what I have described above, and presents the reactions of educators to the challenges that testing raises, through a focus on additional aspects of language testing, either the content of what is being tested, or additional skills that are related to language learning. The educators in the chapters that follow all realise that in the language classroom, more can and should be taught than language. What is taught in addition to language can either be conceptualised as content (e.g. literature or science) or as transferable skill (e.g. intercultural competence, creative writing, literary competence, autonomy). What this volume implicitly does is to call for an end, at least in some contexts, to the attempt to isolate linguistic competence and test it without reference to other competencies and other areas of knowledge.

Cumming (2009: 91) discusses three approaches to the question of 'how integrally language tests, curricula and pedagogical practices should be aligned and what benefits or consequences may arise when they are'. One of the approaches he discusses is 'to adapt assessment policies and practices for particular populations, such as test accommodations for certain learner groups or setting performance standards for occupational purposes' (Cumming, 2009: 91). The chapters in this collection may be interpreted as taking this approach, in that each section looks at the way in which testing practices have been developed and adapted, though not for particular populations, but for particular modifications and viewpoints of the curriculum/construct.

However, it is not only in terms of format and procedures that alternative assessment is important. Fox (2008: 97) points out that 'alternative assessment represents a conception of language that is diametrically opposed to that of traditional tests'. In terms of this volume, this opposition is presented in terms of what is actually being tested, assessed or evaluated, in addition to language.

Mirroring its concern with the content being tested, this volume is structured around the four educational areas of interest: intercultural communication; autonomy; literature; and assessment of content and language. These areas are then examined for their interaction with issues of testing and assessment. The four areas were chosen because of their educational relevance to the majority of language learners, in the majority of contexts; indeed, the contributors deal with a variety of language learning contexts starting with kindergarten and primary education, through secondary schooling to university.

The four areas chosen are central to the educational endeavour. The intercultural dimension of language learning is increasingly coming to the fore in many situations, and has led to an expansion of communicative competence to include intercultural communicative competence (Byram, 1997). Autonomy is important in that it is linked to lifelong learning, which is increasingly an explicit goal of education (Broadfoot, 2005; Jones & Saville, 2009). The relevance of literature and its importance in education is acknowledged by most educational systems, where a knowledge of first language (L1) literature is a requisite; although the link between L2 learning and literature may not be as strong as it was in the past, the use of literature is more prevalent than is commonly thought, and in many cases it never left the language curriculum (see Gilroy & Parkinson, 1996; Maley, 2001; Paran, 2006). The final section deals with the linking of assessment of content and assessment of language in contexts where English as a second language (ESL) students are in mainstream classes, learning alongside English L1 students (see Mohan *et al.*, this volume; Leung & Lewkowicz, 2006).

Outline of the Four Sections

The four areas listed above provide the focus for each of the sections in this volume. The first section, Intercultural Competence, opens with a survey by Sercu (Chapter 2) that discusses the way in which the definition of the construct of intercultural competence has evolved, and looks at issues of reliability and validity in this context, including issues of defining levels of intercultural competence. Chapter 3, by Korhonen, then provides an example of an intercultural training programme and the testing it involved. Of particular interest is the way Korhonen used web-based learning with her students (with a connection here to issues of learner autonomy, the focus of Chapters 5–7). Liddicoat and Scarino (Chapter 4) then present two case studies in which they explore issues of data elicitation in two contexts (secondary school and university) and a number

of foreign languages (French, German and Indonesian). Liddicoat and Scarino examine not only the students' responses, but also the way in which teachers conceptualised the construct of intercultural competence. The next section, devoted to Autonomy, opens with Chapter 5 by Benson, in which he closely examines a number of studies that involved the assessment of autonomy, with a particular focus on the different measures used. Chapter 6, by Lamb, presents an overview of a large study that incorporated a large number of collection instruments and presents snapshots of different groups of students who differed in their views of autonomy and control. This section ends with Chapter 7, by Dam and Legenhausen, who examine longitudinal data from a number of classes to illustrate the importance of listening to the learners' voices. The third section in the book deals with Literature, and presents three very different aspects of the literature and language interface. Paran (Chapter 8) presents an overview of item types that can be used to test literature in language classrooms. Spiro (Chapter 9) looks at a different aspect of literature in language teaching, namely, creativity and creative writing. Lin (Chapter 10) presents another aspect of this interface by showing how teaching Shakespeare in the English as a foreign language (EFL) classroom can enhance language awareness. The last section of the book is devoted to a group of chapters dealing with Language and Content, all of which are based in systemic functional linguistics. Chapter 11, by Mohan, Leung and Slater, opens this section and provides a review of issues to do with L2 assessment, language and content, and a register approach to assessment. They illustrate their approach with an example from causal explanations. Low (Chapter 12) then looks at the way in which teachers make assessment decisions, and Slater and Mohan (Chapter 13) expand on the way that causal explanations can be assessed formatively.

Emergent Themes

Questioning testing and assessment

Unsurprisingly, and as reflected in the title of this volume, the most important point that emerges from the chapters taken as a whole is the critical attitude (and in some cases, scepticism) that the different contributors bring to the assessment process. The title of the chapter by Benson (Chapter 5) is a case in point: 'Measuring Autonomy: Should we put our Ability to the Test?' Benson poses a number of questions, and usefully contrasts what we *can* assess with what we *should* assess. Importantly, although Benson concludes that it is possible to measure certain aspects of autonomy, he continues to be sceptical about the

meaning of such measurement and of its educational value. This ties in with the suggestion by Paran (Chapter 8) that the objectives of teaching and the objectives of testing are in many ways incompatible. An example of the competing tensions in this relationship is Low's (Chapter 12) vivid illustration of the dilemmas that arise in assessment and the way in which teachers need to assess different components of learner production.

The continuing tension between assessment of learning and assessment for learning

This questioning of assessment procedures and assessment in general leads us to considering different types of assessment and their contribution to the learning process. As Broadfoot (2005: 133) says, 'assessment that is integral to teaching and designed to help learning raises standards more than any other potential intervention in the learning process', and all the contributors to this volume confront this issue, either implicitly or explicitly. Korhonen (Chapter 3) discusses formative versus summative evaluations. Lamb (Chapter 6) sets out what is now a commonplace distinction in some circles, assessment *of* learning versus assessment *for* learning (Black & Wiliam, 1998). He then takes this distinction one step further, and looks at assessment *of* autonomy versus assessment *for* autonomy. Spiro (Chapter 9), too, makes this an important part of the teaching and learning process, conceptualising assessment as a scaffold, or indeed a trigger, for creativity. Mohan, Leung and Slater (Chapter 11) provide examples of the way in which micro-cycles of formative assessment occur in the classroom within learner-teacher exchanges. Slater and Mohan (Chapter 13) suggest that learners need to express their knowledge in order to take it further, illustrating the interconnectedness of assessment and learning.

Eliciting samples of behaviour: Multiple sources of data

The different chapters all indicate the issues that are involved in eliciting samples of the constructs that are being assessed. In many cases this is not straightforward, and the researchers used multiple sources of data. Thus, Liddicoat and Scarino (Chapter 4) used two elicitation devices in the first study they describe: one for the teachers (video-recording in class followed by a retrospective account of the recording) and one for the learners (a combination of two writing tasks, followed by group interviews). Benson (Chapter 5) surveys a number of studies and, through a discussion of what can be assessed in autonomous language learning, focuses on methodological issues and the measurable outcomes

examined in the studies he examines. Sercu (Chapter 2) discusses the tensions between the search for a holistic measure for assessing intercultural competence, and the large number of different tests and test types that already exist in the field and that can be used to measure different dimensions of the construct. Portfolio assessment, discussed by Paran (Chapter 8), also relies on multiple sources of data, often produced or elicited at different times, but this time chosen by the student (i.e. the assessee) and commented on by them.

An important element in eliciting data is the view expressed by Brown and Hudson (1998: 670) that 'multiple sources of information are important to think about in selecting assessment strategies and in interpreting their results' (see also, e.g. Shohamy, 1996; Fox, 2008). Echoing this, a consideration of research methodology and samples of learner behaviour has resulted, in many of the chapters in this volume, in a decision to elicit samples of learner behaviour in more than one way. Paran (Chapter 8) suggests that all assessment of literature in language learning should be constructed of a number of different task types in order to provide a multifaceted view of the learner's skills. Many of the contributors suggest the use of portfolios, although Liddicoat and Scarino (Chapter 4) make the point that 'the portfolio itself doesn't elicit; it evidences' (p. 71). Lamb (Chapter 6) provides a long list of different types of activities used as elicitation instruments to provide a rich source of data, including drawings, self-rating scales, brainstorming, questions and projective techniques. Korhonen (Chapter 3), too, used both quantitative and qualitative data, the latter in the form of the rationales that the students wrote for their choices. Interestingly, Korhonen not only used different sources of data, but also looked at different types of teaching, and evaluated the added value that a web-based learning period provided for her learners.

The voice of the learners

This is the explicit title of one of the sections in Dam and Legenhausen (Chapter 7), but other chapters include these as well. This is, to a large extent, the result of research and elicitation methods that give priority to learners' voices in the form of interviews, group interviews, recordings of classroom data, observations and written production by the learners. The voice of the learners is most pronounced in the chapter by Spiro (Chapter 9), where what is being examined is learners' own literary creations.

Teachers learning from assessing their learners

As should be clear from the examples above, assessing different competencies results in finding different ways to elicit language data and to assess learner progress. This means that, within a classroom context, the teacher or the tutor is also developing their own understanding of assessment procedures. Thus, Spiro (Chapter 9) shows how an assessment exercise 'allowed an opportunity for tutors to hone their definition of a ''creative'' task' (p. 186) – thus tutors can expand their views of learning through an engagement with assessment tasks.

Integration and interdisciplinarity

What emerges from a consideration of the various chapters as a whole is a picture of interdisciplinarity and interrelatedness, very much in line with the broader vision of education discussed above. Lamb (Chapter 6) shows how autonomy cannot be discussed without recourse to exploring metacognitive knowledge. Lin (Chapter 10) shows how literary awareness cannot be achieved without language awareness as its backbone. The final section of the book is a prime example of the way in which different areas in language teaching and applied linguistics are brought together to provide an integrated view of assessment: working from a register approach to assessment, within a systemic functional view of language, Mohan, Leung and Slater (Chapter 11) demonstrate the interrelationship of meaning and language; Low (Chapter 12) provides a discussion of the way in which teachers consider language and content in their decision making; and Slater and Mohan (Chapter 13) go on to show how learners use their linguistic resources to construct and express causality.

The responsibility of the tester

One of the important points that emerge from a reading of Mohan *et al.* (Chapter 11) is the issue of responsible testing: the assessor has a responsibility not only to validity and reliability, but to the learners as well. In general, the four general topics dealt with in this volume are evidence of the wish of teachers and testers to assume responsibility for the learners and to assist the learners to develop more than simply linguistic competence. Mohan *et al*. bring out the fact that validity and reliability are not just abstract concepts; they are an important part of the responsibility of the assessor to provide information that is correct both to the person who is undergoing the assessment and also to any external bodies or stakeholders. In terms of responsibility to the learners, one of the ways in which this responsibility to the learners is expressed is

through the clarity and transparency of criteria, as a number of contributors stress (Paran, Chapter 8; Spiro, Chapter 9; Low, Chapter 12). This aspect of the different chapters in the book is thus in line with recent concerns about the use of tests (Shohamy, 1996; McNamara, 2005; McNamara & Roever, 2006).

A focus on the classroom

In each section, the opening chapter surveys the field. The two remaining chapters focus very much on the classroom, and on classroom processes. Some of the chapters focus on specific courses (e.g. Korhonen, Chapter 3; Lin, Chapter 10). Others focus on longitudinal data from a number of years and a number of classes (Dam and Legenhausen, Chapter 7; Spiro, Chapter 9). The chapters deal with small-scale assessment in classrooms, thus filling the gap noted (e.g. Leung & Lewkowicz, 2006).

Conclusion

Although the discussion above presents a number of themes that emerge from a reading of the different chapters, it is important to note that the contributors to this volume do not present 'a united front'. Rather, they present different responses, in different educational contexts, to the issues involved in testing and assessment where more than one competence is involved. The differences in approach to teaching and to assessment exhibited by the contributors are just as important as the similarities. It is also important to note that many of the contributors do not see themselves as primarily engaged in language testing: they are engaged in language assessment only to the extent that any teaching involves assessment (as in the quote by Rea-Dickins that opened this chapter).

One issue that arises is that the specificity of each context and the specificity of the constructs involved mean that large-scale validation of the instruments used is difficult. However, the chapters indicate the extent to which work is being done in this area in a large variety of international contexts. The large variety of educational contexts discussed in this volume may well serve as an indication to educators of what is possible to achieve in this area at classroom and institutional level.

References

Alderson, C. (2004) The shape of things to come: Will it be the normal distribution? In M. Milanovic and C. Weir (eds) *European Language Testing in*

a Global Context. Proceedings of the ALTE Barcelona Conference July 2001. Studies in Language Testing 18. Cambridge: Cambridge University Press.

Bachman, L.P. (1990) *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Birenbaum, M. (1996) Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum and F. Dochy (eds) *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge* (pp. 3–29). Boston, MA: Kluwer Academic.

Birenbaum, M. and Dochy, F. (eds) (1996) *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge*. Boston, MA: Kluwer Academic.

Black, P. and Wiliam, D. (1998) Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice* 5 (1), 7–74.

Broadfoot, P.M. (2005) Dark alleys and blind bends: Testing the language of learning. *Language Testing* 22 (2), 19.

Brown, J.D. and Hudson, T. (1998) The alternatives in language assessment. *TESOL Quarterly* 32 (4), 653–675.

Byram, M. (1997) *Teaching and Assessing Intercultural Communicative Competence*. Clevedon: Multilingual Matters.

Cheng, L. and Curtis, A. (2004) Washback or backwash? A review of the impact of testing on teaching and learning. In L. Cheng and Y. Watanabe with A. Curtis (eds) *Washback in Language Testing: Research Contexts and Methods* (pp. 3–17). Mahwah, NJ: Lawrence Erlbaum Associates.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Cumming, A. (2009) Language assessment in education: Tests, curricula, and teaching. *Annual Review of Applied Linguistics* 29, 90–100.

Eckes, T., Ellis, M., Kalnberzina, V., Pizorn, K., Springer, C., Szollás, K. and Tsagari, C. (2005) Progress and problems in reforming public language examinations in Europe: Cameos from the Baltic States, Greece, Hungary, Poland, Slovenia, France and Germany. *Language Testing* 22 (3), 355–376.

Fox, J. (2008) Alternative assessment. In E. Shohamy and N.H. Hornberger (eds) *Encyclopedia of Language and Education. Volume 7: Language Testing and Assessment* (pp. 97–109). New York: Springer.

Huerta-Macías, A. (2002) Alternative assessment: Responses to commonly asked questions. In J.C. Richards and W.A. Renandya (eds) *Methodology in Language Teaching* (pp. 338–343). Cambridge: Cambridge University Press.

Jones, N. and Saville, N. (2009) European language policy: Assessment, learning and the CEFR. *Annual Review of Applied Linguistics* 29, 51–63.

Leung, C. and Lewkowicz, J. (2006) Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly* 40 (1), 211–234.

Leung, C. and Scott, C. (2009) Formative assessment in language education policies: Emerging lessons from Wales and Scotland. *Annual Review of Applied Linguistics* 29, 64–79.

Lynch, B. (2001) Rethinking assessment from a critical perspective. *Language Testing* 18 (4), 351–372.

Maley, A. (2001) Literature in the language classroom. In R. Carter and D. Nunan (eds) *The Cambridge Guide to Teaching English to Speakers of other Languages* (pp. 180–185). Cambridge: Cambridge University Press.

McNamara, T. (2001) Language assessment as social practice: Challenges for research. *Language Testing* 18 (4), 333–349.

McNamara, T. (2005) 21st century Shibboleth: Language tests, identity and intergroup conflict. *Language Policy* 4 (4), 1–20.

McNamara, T. and Roever, C. (2006) Language testing: The social dimension. *Language Learning* 56, Supplement 2.

Morrow, K. (2004) Background to the CEF. In K. Morrow (ed.) *Insights from the Common European Framework* (pp. 3–11). Oxford: Oxford University Press.

Paran, A. (2006) The stories of literature and language learning. In A. Paran (ed.) *Literature in Language Teaching and Learning* (pp. 1–10). Alexandria, VA: TESOL.

Rea-Dickins, P. (2004) Understanding teachers as agents of assessment. *Language Teaching* 21 (3), 249–258.

Shohamy, E. (1996) Language testing: Matching assessment procedures with language knowledge. In M. Birenbaum and F. Dochy (eds) *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge* (pp. 143–160). Boston, MA: Kluwer Academic.

Shohamy, E. (2005) The power of tests over teachers: The power of teachers over tests. In D.J. Tedick (ed.) *Second Language Teacher Education: International Perspectives* (pp. 101–111). Mahwah, NJ and London: Lawrence Erlbaum Associates.

Shohamy, E. (2007) Language tests as language policy tools. *Assessment in Education* 14 (1), 117–130.

Shohamy, E., Donitsa-Schmidt, S. and Ferman, I. (1996) Test impact revisited: Washback effect over time. *Language Testing* 13, 298–317.

Shohamy, E. and Hornberger, N.H. (2008) *Encyclopedia of Language and Education. Volume 7: Language Testing and Assessment* (2nd edn). New York: Springer.

Spolsky, B. (2008) Language assessment in historical and future perspective. In E. Shohamy and N.H. Hornberger (eds) *Encyclopedia of Language and Education. Volume 7: Language Testing and Assessment* (pp. 445–454). New York: Springer.

Spolsky, B. (2009) Editor's introduction. *Annual Review of Applied Linguistics* 29, vii–xii.

Wall, D. (1997) Impact and washback in language testing. In C. Clapham and D. Corson (eds) *Encyclopedia of Language and Education*. Volume 7: *Language Testing and Assessment* (pp. 291–302). Dordrecht: Kluwer Academic.

Weir, C. (2003) A survey of the history of the Certificate of Proficiency in English (CPE) in the twentieth century. In C. Weir and M. Milanovic (eds) *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002. Studies in Language Testing 15*. Cambridge: Cambridge University Press.

Weir, C. (2005) *Language Testing and Validation: An Evidence-based Approach*. Basingstoke: Palgrave Macmillan.

Part 1

Intercultural Competence

Chapter 2

Assessing Intercultural Competence: More Questions than Answers

LIES SERCU

Introduction

The importance of intercultural competence is increasingly recognised by educators and employers alike. Employers want their employees to be skilful negotiators in increasingly intercultural work situations. Educators believe learners should be prepared for living in an intercultural world. They design foreign language-and-culture curricula that help learners develop intercultural competence,[1] which is conceived of as a general humanistic educational goal.

Both in the professional and educational domain, ways are sought to test or assess whether or not people have actually developed intercultural competence. Employers are looking for reliable tools that can predict whether a particular employee possesses the necessary knowledge, skills, attitudes and personality traits to be successful in intercultural business contacts. Educators, too, are looking for elicitation procedures and evaluative instruments that can provide them with feedback on whether their learners have benefited from their intercultural teaching and that can indicate to society that learners have attained the educational targets underlying school curricula.

In the professional domain, organisations offering intercultural training to companies are increasingly using self-awareness inventories as tests of intercultural competence, despite the fact that these inventories were not originally developed for external (high stakes) testing, but rather to serve as springboards for exploring thinking patterns and behavioural styles in a training context, and thus as aids for self-assessment and developing learning rather than as clustered test item batteries. Examples of self-awareness inventories are the intercultural development inventory (Hammer *et al*., 2003); the cross-cultural

17

adaptability inventory (Kelley & Meyers, 1999); the overseas assignment inventory (Tucker, 1999); or the four-value orientation exercise (Casse, 1999) (for a useful overview, see Fantini, 2006 or Humphrey, 2007).

In the domain of foreign language education, too, ways are sought to demonstrate the (partial) absence/presence of, or progress in intercultural competence. Attempts have been made to design frameworks for measuring and assessing intercultural competence (Byram, 1997; Sercu, 2004; Arasaratnam & Doerfel, 2005; Sinicrope *et al*., 2007; Alvarez, 2007), and some concrete suggestions have been put forward as to what test formats can be used (see e.g. Fowler & Mumford, 1995; Bartz & Vermette, 1996; Lange, 2003; Schulz, 2007). Many of the assessment techniques proposed are actually also teaching techniques aimed at enhancing learners' culture-specific and/or culture-general awareness: cultural minidramas, critical incidents, culture assimilators or simulation games, combined with oral or written reflective work and documentation in the foreign language. Thus, authentic foreign language documents originating from a foreign culture are used both as input for teaching (language and culture), and for assessment purposes, for example when assessees are asked to explain the underlined culturemes in the text against the background of the text and against their larger understanding of a particular culture or of culture-general insights.

The increasing weight given to the promotion in learners of essential skills for self-directed and autonomous life-long learning has resulted in a parallel emphasis in assessment on self-assessment and assessment of the quality of the interculture learning process itself, next to learning outcomes (Sercu, 2002). Within the same vein, assessment has come to be viewed as assessment for promoting learning, not for judgement. Indeed, any assessment, also in the area of foreign language and culture teaching, should promote the learners' 'capacity for self-assessment so that they can become reflective and self-managing' in their learning, providing learners with 'constructive guidance about how to improve' and a 'shared understanding of the criteria by which they are assessed' (Assessment Reform Group, 2002: 2), such as their ability to learn from others and participate in learning dialogues.

Given this important paradigmatic shift in education, intelligently described by Anna Sfard (1998) as a movement from an acquisition metaphor towards a participation metaphor of learning, without, however, discarding the first, it should not come as a surprise that the issue of how to elicit data on and determine growth in intercultural competence is anything but wound up, also and especially because the construct 'intercultural competence' itself has proved hard to define

systematically and in ways that make it observable and assessable. Additionally, intertwined with these debates is the as yet unresolved issue of how to assess *language* development and *intercultural* development simultaneously, for poor foreign language skills can prevent learners from demonstrating high intercultural competence.

Till now, many proposals, which have unfortunately been formulated largely independently of one another, have been thrown at teachers and researchers alike, who are now experiencing frustration and uncertainties about the expected standards, outcomes, assessment formats and rating scales to use (Sercu *et al.*, 2005).

Providing an introduction to the two case study chapters on the assessment of intercultural competence presented next in this book, this chapter first briefly addresses the question of why the assessment of intercultural competence is important. Secondly, basic dimensions and components of intercultural competence are outlined both from the perspective of training for intercultural business and from that of school education. Next, existing test formats are presented and some major caveats in assessing intercultural competence reliably and validly are discussed. After addressing some points that deserve special attention when attempting to define levels of intercultural competence, we conclude with an appeal to more systematically define which specific behavioural evidence teachers can draw on in order to assess progress in learners' intercultural competence, for such work is direly needed.

The Importance of Assessing Intercultural Competence in Foreign Language Education

The importance of assessment in foreign language and culture education cannot be overestimated. Assessment is important for all parties concerned. Learners want to know whether they are making adequate progress and in which areas improvement is needed. Teachers want to find out whether their learners are actually learning what they are teaching. They also expect to get feedback from assessment results regarding the way in which their teaching might be adjusted. In addition, they realise that all communication in a foreign language is intercultural, and therefore their teaching should promote the acquisition of intercultural competence. At the same time, they know that learners tend to overlook what is not assessed and therefore demand that good assessment tools be developed. Parents want to see their children's progress documented. Society holds teachers accountable and wants to see the effects of teaching efforts.

Assessment is also important because of the 'backwash effect' it has on teaching. Teachers tend to teach what will be tested (McNamara, 2000). If national examinations would assess 'communicative competence in a foreign language' via multiple-choice grammar questions, then teachers would probably focus their teaching on grammar and how to approach multiple choice tests, so as to enhance their learners' chances to do well on the test. Likewise, tests that assess intercultural competence via quizzes that aim to document the extent of learners' knowledge of the cultural elements touched upon in class, will probably lead to teaching that promotes the acquisition of cultural knowledge only, and neglects the other dimensions of intercultural competence discussed in the next section.

Intercultural Competence: Construct Definition

When discussing intercultural competence in foreign language education, it is important to underline that 'intercultural competence' always implies 'communicative competence', and therefore always also has a linguistic, sociolinguistic and discourse component. We will first briefly touch upon evolutions in the definition of 'communicative competence' and on the implications these have had for assessment, before commenting on changes over time in the concept of 'intercultural competence'. Table 2.1 provides an overview of the different dimensions of intercultural competence that should be addressed in today's education, and therefore also in assessment.

In the 1960s, language competence was defined in terms of 'command of the structures of the language' (Lado, 1961). Consequently, language tests assessed knowledge of vocabulary and especially grammar in an isolated and de-contextualised way. This tradition came to be known as 'discrete point testing'. It arose in the wake of psychometrics, the new scientific discipline that tried to measure cognitive skills in an objective way. This testing tradition was discarded in favour of integrative tests, which combined the assessment of systemic knowledge with the ability to apply that knowledge for communicative purposes. The cloze test (Oller, 1979) is a typical representative of this testing tradition. Integrative tests can be said to have been the predecessors of what are now known as communicative tests, which focus on the learner's ability to use the foreign language appropriately for a variety of communicative purposes and in a variety of contexts encountered (for a more elaborate discussion, see McNamara, 2000).

The construct of intercultural competence, too, has undergone changes, which have been reflected in the terminology used. The aim of culture teaching changed from 'promoting familiarity with the foreign culture' through 'assisting learners to gain cultural awareness' and 'fostering intercultural communicative competence in learners' to 'turning out intercultural beings' and 'life-long learners of interculture'. Whereas cultural awareness already refers to understanding the fact that culture always affects communication, and to such essential qualities of an interculturally competent individual, such as respect and tolerance, intercultural communicative competence, in addition, emphasises performance aspects relating to communication in a foreign language in intercultural situations. The actualisation of human beings who think democratically and are respectful of all human beings, who understand the complexity of the world, the subjectivity of human perspectives, the continuous development and diversity of societies and who can act in (one or more foreign) language(s) accordingly constitutes the focus of educational approaches aiming to turn out 'intercultural beings' (Byram, 1997; Alred *et al.*, 2002; Phipps & Gonzalez, 2004). More recently, education for interculture supports the demand of society to turn out young people with life-long learning skills, so that they can develop a global mind and intercultural skills in a self-managed and goal-directed process (Sercu, 2004; Broadfoot, 2005). Whereas previously, intercultural (language) education viewed learning foremost (and in many cases exclusively) in terms of individual enrichment and the acquisition of knowledge, this acquisition metaphor of learning has now come to exist next to a participation metaphor, which views learning as community building and as participating in, contributing to and interacting with one's (multicultural and multilingual) learning communities. In learning as acquisition, the learner is predominantly viewed as an individual outsider to any foreign culture, even if interacting with them, as acquiring knowledge conveyed by the teacher and as developing a profound interest in and critical understanding of foreignness. By contrast, in learning as participation, the learner is viewed as a participant of intercultural and multilingual communities, learning through interacting with their members and finding assistance in a teacher-coach who helps to clarify learning and stimulates practice of learning to achieve the attainment target set.

Turning to definitions of 'intercultural competence' in the professional domain now, four dimensions tend to be distinguished: knowledge, skills, attitudes and traits. Chen and Starosta (1998), for example, provide a model of the effective interculturalist, which focuses on affective

(intercultural sensitivity), cognitive (intercultural awareness) and behavioural (intercultural adroitness) components. Affectively speaking, the effective interculturalist is said to have a positive self-concept, to be open-minded, non-judgemental and relaxed in social interaction. Cognitively speaking, the effective interculturalist is culturally self-aware and knowledgeable and shows cultural awareness and familiarity with different areas in which culture affects language and language use in communicative interaction, which help to reduce the ambiguity and uncertainty that are inherent in intercultural interaction. Behaviourally speaking, the effective interculturalist possesses good message skills, the skill of appropriate self-disclosure and interaction management, behavioural flexibility, as well as social skills, both in the verbal and non-verbal domain. The personality traits that are conducive to intercultural competence are: empathy, respect, interest in cultures, flexibility, tolerance, open-mindedness, initiative, sociability and positive self-image (Kealey & Ruben, 1983). Reflecting current business concerns, Kim (1996, 2002) adds 'global intelligence' to the repertoire of terms concerning intercultural business competence, defining it as the capacity to recognise our own cultures and those of others (1) for increasing personal and professional effectiveness, (2) for creating efficiency and productivity in the global workplace, and (3) for promoting harmony and humanity in the world. According to Kim, global intelligence consists of seven pillars, namely, global literacy, mentality, identity, competency, technology, integrity and global humanity. Though a humanistic vein is present in this definition of global leadership, it primarily points towards instrumental knowledge and skills that will help a company's leader to survive and make profits in our increasingly multilingual, multicultural and globalising world. Statements such as 'be aggressively curious about other culture' (global mentality), 'shift their paradigms as necessary' (global mentality) or 'challenge the negative cultural influence on the status quo' (global competency) strikingly reveal the sometimes missionary and patronising principles underlying what seems a very laudable goal to pursue.

Whereas in business training, instrumental goals relating to enhancing the trainees' ability to master intercultural contacts with specific cultures clearly have been dominant and continue to be dominant, in educational approaches, pedagogical humanistic ones dominate. Table 2.1 presents an overview of the different learning goals of general education in regard to intercultural competence, reflecting current tendencies in the understanding of learning interculture, and distinguishing between three educational pillars, namely, '*Assist the learner in improving learning of*

Table 2.1 Dimensions of promoting intercultural competence in education

Assist the learner in improving of interculture	Assist the learner in perfecting intercultural communicating skills	Assist the learner in imagining and adopting an interculturalist identity
Learning goals are: – improve language competencies for intercultural communication; – acquire culture-specific (own and other) and culture-general knowledge; – become aware of cross-cultural relationships (similarities and differences; political, economic, global, cultural,…relationships); – gain insight into ways in which culture affects language and communication; – gain understanding of learning demands of the subject matter and the learning process; – gain insight into one's own interculture learning preferences (e.g. via direct or indirect contacts; in a holistic/analytic way); – gain insight into advantages and disadvantages of community (group) learning and individual learning of interculture; – know and understand the assessment criteria and learning goals for interculture.	*Learning goals are:* – acquire the skill to operate cultural and intercultural knowledge, attitudes and skills under the constraints of (real-time) communication and in intercultural dialogue (skills also include language skills); – acquire the skill to clarify one's own understanding of culture in relation to others' understanding of culture; – acquire the ability to communicate in and learn from intercultural dialogue; – acquire the ability to operate (metacognitive) strategies to direct one's own and others' learning of interculture; – acquire the skill to assess one's own learning outcomes and learning processes.	*Learning goals are:* – develop intercultural sensitivity, that is, mindfulness of and attention to intercultural issues at play in interaction; – develop an essentially positive disposition towards learning intercultural competence and acting interculturally; – develop the willingness to engage with otherness and learn from it; – develop the ability to assess oneself with respect to being or not being an intercultural being.

interculture', '*Assist the learner in perfecting intercultural communicating skills*' and '*assist the learner in imagining and adopting an interculturalist identity*'. Each of the specific goals listed can be used by teachers as starting points for designing teaching tasks and activities that will allow them to assess learning of interculture in their language learners, and set in motion a new cycle of learning, assessing for learning, learning and again assessing and showing possibilities for learning.

Implications for Assessment: More Questions than Answers

The question addressed next is how the construct of intercultural competence, as outlined in Table 2.1, can be assessed reliably and validly. Answering this question requires answering a number of subquestions: Does an instrument exist for testing intercultural competence holistically? What test formats can be used to elicit data on the cognitive, competence and affective dimensions of intercultural competence? Can stages or levels in intercultural competence be distinguished? What can be considered indicators of progress in learning? These subquestions are the bases for the sections that follow.

Does an instrument exist for testing intercultural competence holistically?

What is quite clear from Table 2.1 is that assessing intercultural competence in education will imply assessing several subcompetencies, such as foreign language skills, the ability to read, relate and explicate cultures, the ability to systematically organise cultural information and information about interculture into cross-cultural and interculture schemata, foreign language and interculture learning skills, social and interacting skills and critical thinking skills. If one wanted to assess intercultural competence holistically, one would also have to assess to what extent learners can be viewed as intercultural beings, assessing the presence or absence of intercultural values and attitudes (e.g. be intrinsically interested in understanding cultures from within) and recognising demonstrations of such attitudes or values in particular student attitudinal behaviours [e.g. students show through their behaviour that they pursue the goal of understanding cultures from within: they read foreign language texts written by authors from within the foreign culture to better understand that foreign culture. They also consistently question in speaking and writing whether their perception

of a cultural phenomenon (e.g. wearing a headscarf) as an outsider to the culture matches insiders' perception.]

As yet, such a holistic measure is not available. It would probably be composed of disaggregatable composite measures for which a summary total score could be calculated; this would then provide an index of intercultural competence. Such a 'questionnaire' could contain statements to be scored by learners, for example regarding how often they do something to learn about cultures (e.g. read articles on foreign cultures), how important they believe a particular skill is (e.g. the skill of opening a conversation in a culturally appropriate manner) or to what extent they assess themselves to be interculturally competent (e.g. to what extent can they be critical of their own culture, or show respect for other cultures?), to name just a few examples. Such an instrument could also be turned into a self-assessment instrument, for example using can-do-statements similar to those proposed by the Council of Europe (2001) for self-assessment in language learning. Example statements could be 'When I meet foreign youngsters, I can easily make contact with them'; 'I can participate in the foreign language in any conversation dealing with daily life topics being sensitive to what other participants consider as ''normal behaviour'' in their cultures'; 'When I need information on an aspect of a foreign culture, I know which websites to use because they provide reliable and thorough information'; 'I find it hard to explicate an element of my own culture, for example, why we celebrate 1 May and have a day off then at school'.

If such a tool existed, teachers would have something to hold on to when planning practice activities and designing learning paths fit to promote individuals' or groups' learning, sharing with learners what specific goals they should achieve and what criteria will be used to assess to what extent these goals have been met.

Eliciting and interpreting data on the cognitive, competence and affective dimensions of intercultural competence

Even if a validated test like the one described above does not exist yet, it is not difficult to imagine ways of assessing intercultural competence in its different dimensions. The knowledge dimension, to start with, can be assessed in a number of ways. Bartz and Vermette (1996) list 16 prototype test formats for the assessment of intercultural competence: multiple choice tests, culture-general assimilators (Fowler & Mumford, 1995); critical incidents (Fowler & Mumford, 1995); solving cross-cultural conflict situations; answering written or oral questions on appropriate

behaviour or about the significance of a fact in a particular culture; analysing visual examples of authentic and cultural situations; giving verbal descriptions of a typical or an unlikely situation; reports on reading; simulated interactions or situations; examining cultural significance of underlined words or phrases; identification of significant features in a literary passage; describing a photo or drawing of a culture-specific situation showing social behaviour; observing an audio or video document for sociolinguistic behaviour and explaining what is happening and what causes a misunderstanding to arise; demonstrating knowledge of sources of information; organising, making sense and explicating one's cultural observations.

Some of these prototypes can be used to assess interculture learning ability, for example when asking learners to organise their cultural observations, or self-report in writing on how they have approached a particular learning task, what they have learned from this approach and how they could possibly improve it. (For a more elaborate discussion of how to assess learning skill, the reader is referred to the section on the assessment of autonomous learning in this book, Chapters 5, 6 and 7.)

With respect to the assessment of attitudes, one may think of continuous attitude scales or questionnaires. Teachers can gauge where students are in their attitudes towards interculture or towards a particular idea concerning interculture by observing how far along the dimension they are, as indicated by their verbal and non-verbal behaviours. Questionnaires could ask learners to score statements, such as 'It is best to keep quiet in a group when you don't know the cultural backgrounds of its members' (dare to engage in otherness) or 'I find it difficult to accept that I have to cover my body completely when visiting country x' (acceptance of differences; ability to refrain from disapproval of what is normal within a particular cultural context). On the basis of their answers, students could again be situated on a continuous line with stages, showing where they are at with respect to the attitudes expected of them. With respect to the attitude 'tolerance of otherness', the different marking points on the continuum could be as shown in Figure 2.1.

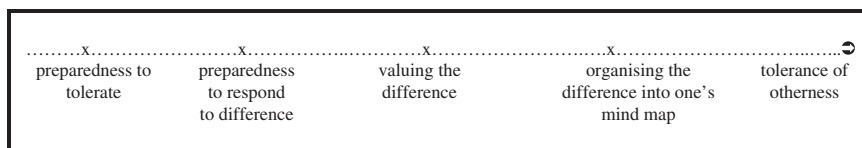| preparedness to tolerate | preparedness to respond to difference | valuing the difference | organising the difference into one's mind map | tolerance of otherness |

Figure 2.1 Continuous attitudinal scale for assessing 'tolerance of otherness'

Additional assessment issues

With respect to the assessment of each of the different dimensions of 'intercultural competence' individually, a number of questions suggest themselves and need our special attention. As regards the assessment of cultural knowledge, understanding and interpreting, the question arises whether it is possible to score learners' answers objectively. When assessing the acquisition of factual information dealt with in class, it may be possible to assign test takers an 'objective' score that reflects the degree of comprehensiveness and correctness of their knowledge. However, the actual choice of what culture-specific information to select to illustrate particular principles of interculture is very difficult and will depend on teachers' insights and preferences as well as on learners' interests and needs. (For a discussion of how to select cultural information for the foreign language curriculum, see Sercu, 2002 and Schulz, 2007). Some even argue that objectivity is not possible when dealing with culture because culture is always subjectively experienced and construed (Atkinson, 1999; Sercu, 2002). Therefore, assessors must also ask themselves what score learners will be assigned when they put forward a personal interpretation of a particular cultural phenomenon that differs from the interpretation put forward in class, but which is perfectly feasible when accepting other cultural points of view. Will they still pass or will they fail? Will they be assigned a point for their answer or not?

With respect to the assessment of learners' cognitive and metacognitive strategies, one question that suggests itself is whether the fact that a learner cannot *make explicit* what strategies she/he used also means that she/he is not an able problem-solver or intercultural learner. Other questions include: Are the problem-solving strategies demonstrated and taught in the classroom better than the ones developed by the learners themselves? If learners cannot solve a particular intercultural problem, is it because they are not skilful with respect to the learning dimension of intercultural competence? Or is it because of a difference in understanding the task due to differences in cultural backgrounds between the assessee and the assessor? Additionally, the question needs to be asked whether perhaps their level of mastery of the foreign language hampered adequate task completion, and whether allowing learners to react (for example, to a video showing culture-specific social behaviour) in their mother tongue might not have led to a different assessment of where they are, interculturally speaking.

With respect to the assessment of the affective dimension of intercultural competence, the question arises whether it is desirable that learners be assessed with respect to particular attitudes or personality traits. Does education want to be prescriptive about the intercultural attitudes learners should develop and can learners be punished for not having particular desired personality traits, such as 'interest in cultures' or 'positive self-image', which have been identified as characteristic of the effective intercultural person? What could possibly be the learning advantage of being labelled a person with 'high preference for certainty', if it is not explained simultaneously how this trait of character is advantageous to a person living in culture x, but not in culture y, where ambiguity can be left to exist for a longer period of time?

The above distinction between different test formats for assessing different dimensions of intercultural competence is not intended to promote the impression that these dimensions can indeed be assessed separately from one another. Test developers and assessors need to be aware and attentive to a degree of interdependence between the different dimensions of intercultural competence that have been distinguished. When designing tests for assessing 'interculture learning ability' for example, they should be aware that the extent of a learner's culture-specific or culture-general knowledge may affect the learner's learning skill. When designing ways to assess the acquisition of cognitive or metacognitive strategies, they should reflect on the culture-specific nature of particular task types, and on the particular attitudes that learners may bring to the assessment task. And, of course, at all times, assessors must realise that differences in communicative competence in the foreign language may cause assessees with the same level of intercultural competence to perform differently.

Levels of intercultural competence

The final question addressed in this chapter is whether learning interculture can be viewed as a quantifiable step-by-step process from one level to the next, much like communicative competence in a foreign language has been defined in terms of six levels, namely, A1, A2, B1, B2, C1, C2 (Council of Europe, 2001), where it is clear, however, that the different competencies making up language competence, namely, reading, speaking, listening and writing skills, can be developed towards different levels at the same point in time in a particular learner. Thus, even if levels have been circumscribed, it is recognised that language

competence, though perceived as a holistic competence, should actually be assessed as if it were a competence consisting of different competencies. Just like language learning, intercultural competence learning should be viewed as helix-shaped, as learning that comes back to what has already been learned, but tries to deepen that learning, lingering on spiral 3 for a long time, integrating learning from spirals 1 and 2 and automatising this learning to a point where it becomes impossible to explicate changes experienced towards a deepened intercultural sensitivity or the enhanced ability to participate in intercultural dialogue in a less prejudiced manner. In interculture learning, learners will be offered the same cultural text types on a repeated basis in order to assist them to deepen their understanding of a particular cultural issue. Learning is viewed as a deepening of insights and as indefinite in the sense that it is impossible to define where level 1 stops and level 2 begins, when learning is halfway or at 75%.

Especially in the domain of training for interculture in business contexts, a number of attempts have been made to rate trainees and measure progress in intercultural skill (for an overview of assessment tools, see Humphrey, 2007), describing progress in terms of levels. The work done within the framework of the INCA project (www.inca.org), for example, proves the possibility and at the same time the difficulty of identifying levels of competence sufficiently distinct from one another and sufficiently well-defined to ensure that assessors can relatively straightforwardly assign test takers to a particular level. This is evident in the following quote from the INCA Assessor Manual:

> The intercultural competence framework on the following page has been designed to describe three levels of performance. The framework will help the assessor to evaluate observations or answers given by the assessee.

As can be seen from the following extensive quote from the INCA website, growth is determined by means of words, such as *quicker* or *larger (repertoire)*, which could indeed be measured in terms of the number of seconds needed or the number of elements in one's repertoire, but also by words such as *easier, more systematically* or *confident enough* (full competence), for which it is far more difficult to imagine what could count as evidence of a trainee having reached full competence. Though a description of what would characterise a person with basic, intermediate and full competence may be helpful for trainees to assess themselves mentally, it is hard to see how a description like the one quoted below could 'help the assessor to evaluate observations or answers given by the

assessee'. Another problem with this type of description is the suggestion that training will culminate in full competence, which once acquired will remain acquired. There is no hint at the possibility that a person might become less willing to empathise with or act tactfully towards members of particular cultures after many frustrating and disappointing experiences in communicating with them, or being tired of always acting tactfully but never experiencing tactfulness towards himself/herself. In the same vein, the levels are presented as implicational scales, which means that skills situated on level 3 cannot be acquired before those at levels 1 and 2 have been acquired successfully, and this is somewhat problematic: Does one need a checklist of the sort of situations one is likely to need to deal with (intermediate competence) before one can deal with intercultural situations adequately (basic competence)? Is intuitively dealing with intercultural situations (full competence) necessarily different from dealing with intercultural situations at the basic level of competence, if one's intuition is correct?

Level 1 – Basic Competence
You are already willing to interact successfully with people of other cultures. You tend to pick things up and learn from them as you go along, but you haven't yet got the experience to work out any system of dealing with intercultural situations in general. You respond to events, rather than planning for them. At this stage you are reasonably tolerant of other values, customs and practices although you may find them odd or surprising and approve or disapprove.

Level 2 – Intermediate Competence
As a result of experience and/or training, you are beginning to view more coherently some of the aspects of intercultural encounters you used to deal with in a "one-off" way. You have a mental "map" or "checklists" of the sort of situations you are likely to need to deal with and are developing your skills to cope with them. This means that you are more prepared for the need to respond and adapt to the demands of unfamiliar situations. You are quicker to see patterns in the various experiences you have and you are beginning to draw conclusions without having to seek advice. You find it easier to respond in a neutral way to difference, rather than approving or disapproving.

Level 3 – Full Competence
Many of the competences you developed consciously at level 2 have become intuitive. You are constantly ready for situations and

encounters in which you will exercise your knowledge, judgement and skills and have a large repertoire of strategies for dealing with differences in values, customs and practices among members of the intercultural group. You not only accept that people can see things from widely varying perspectives and are entitled to do so, but you are also able to put yourself in their place and avoid behaviour you sense would be hurtful or offensive. At this level of operation you are able to intercede when difficulties arise and tactfully support other members of the group in understanding each other. You are confident enough of your position to take a polite stand over issues, despite your respect for the viewpoint of others.
(http://www.incaproject.org/en_downloads/21_INCA_Assessor_ Manual_eng_final.pdf: 8; accessed 14/1/09)

Even if this type of rating scales makes an attempt at making explicit what knowledge, organisation of knowledge, application of knowledge or indicator of a particular attitudinal behaviour will be considered evidence of learning, reliability in the assignment of learners to one level or the other is difficult to achieve, despite the fact that the three levels have been defined very broadly. Because this particular rating scale depends highly on actual experience in intercultural situations, it does not fit well into school education, and even suggests that school learning without direct contact with other cultures cannot lead to the development of intercultural competence.

Conclusion

In this chapter, I have addressed the question of how to assess intercultural competence and have considered some major caveats that need to be taken into account when designing assessment tasks and assessing learners' performance on those tasks from an intercultural point of view. From our discussions, it has become very clear that the assessment of intercultural competence in foreign language education is anything but straightforward.

At a time where there is an implicit commandment in education to promote the acquisition of intercultural competence, it is high time that more systematic, intercultural and international concerted research efforts went into the development and validation of a limited number of reliable and valid assessment tools for different age and ability groups, as well as curricula, which can then exert mutual positive backwash effects from which teachers, learners and society will benefit. In the context of training professionals for intercultural (business) contexts, one

survey identified no less than 86 assessment instruments (Fantini, 2006). A US Army Research Institute study narrowed the list down to 10 quantitative instruments for further exploration into their reliability and validity (Abbe *et al.*, 2007). Specific examples of qualitative assessment instruments, such as scenario-based assessments and simulation games also abound, and are used, together with quantitative measures, to assess a person's competence for international business at home and abroad. In professional training contexts, assessment criteria tend to be defined in culture-specific terms, namely, in relation to the major cultures with which a particular individual will be confronted when working for a particular firm or going to live in a particular culture. It seems like everything is under control in terms of the assessment of intercultural competence in the professional domain, and that the market of ever new inventories and scales has come to a quieting halt. Hectic though the pace towards developing assessment instruments has been in the professional domain, in the educational domain the pace continues to be slow and this is a pity, certainly at a time when education has a more powerful than ever contribution to make to tomorrow's society, which should be composed of people who find it self-evident to function adequately and knowledgeably in intercultural contact situations and who have developed a broad-minded intercultural world view and identity. What are we waiting for?

Note

1. The term 'intercultural competence' will be used here, not 'intercultural communicative competence'. Both terms are used in the context of foreign language education, with the second one underlining explicitly that reference is made to intercultural competence in situations where one is using a language that is not one's mother tongue.

References

Abbe, A., Gulick, L.M.V. and Herman, J.L. (2007) *Cross-cultural Competence in Army Leaders: A Conceptual and Empirical Foundation*. Washington, DC: US Army Research Institute.

Alred, G. and Byram, M. (2002) Becoming an intercultural mediator: A longitudinal study of residence abroad. *Journal of Multilingual and Multicultural Development* 23 (5), 339–352.

Álvarez, I. (2007) Foreign language education at the crossroads: Whose model of competence? *Language, Culture and Curriculum* 20 (2), 126–139.

Arasaratnam, L. and Doerfel, M.L. (2005) Intercultural communication competence: Identifying key components from multicultural perspectives. *International Journal of Intercultural Relations* 29, 137–163.

Assessment Reform Group (2002) *Assessment for Learning: 10 Principles. Research-based Principles to Guide Classroom Practice*. On WWW at http://arg.educ.cam.ac.uk/CIE3.pdf.

Atkinson, D. (1999) TESOL and culture. *TESOL Quarterly* 33 (4), 625–654.

Bartz, W. and Vermette, R. (1996) Testing cultural competence. In A.J. Singerman (ed.) *Acquiring Cross-cultural Competence: Four Stages for Students of French* (pp. 75–84). Lincolnwood, IL: National Textbook Company.

Broadfoot, P. (2005) Dark alleys and blind bends: Testing the language of learning. *Language Testing* 22, 123–141.

Byram, M. (1997) *Teaching and Assessing Intercultural Competence.* Clevedon: Multilingual Matters.

Casse, P. (1999) The four-value orientation exercise using a self-awareness inventory. In S. Fowler and M. Mumford (eds) *Intercultural Sourcebook: Cross-cultural Training Methods* (Vol. 2) (pp. 31–44). Yarmouth, ME: Intercultural Press.

Chen, G-M. and Starosta, W.J. (1998) *Foundations of Intercultural Communication*. Boston, MA: Allyn & Bacon.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Fantini, A.E. (2006) *Exploring and Assessing Intercultural Competence.* On WWW at http://www.sit.edu/SITOccasionalPapers/feil_appendix_f.pdf. Accessed 14/1/09.

Fowler, S. and Mumford, M. (eds) (1995) *Intercultural Sourcebook: Cross-cultural Training Methods* (Vol. 1). Yarmouth, ME: Intercultural Press.

Hammer, M. (1999) A measure of intercultural sensitivity: The intercultural development inventory. In S. Fowler and M. Mumford (eds) *Intercultural Sourcebook: Cross-cultural Training Methods* (Vol. 2) (pp. 61–72). Yarmouth, ME: Intercultural Press.

Humphrey, D. (2007) *Intercultural Communication Competence: The State of Knowledge. Report Prepared for CILT, the National Centre for Languages*. London: CILT. On WWW at http://www.cilt.org.uk/standards/donnareport.pdf.

INCA Intercultural Competence Assessment/LdVII (2004) On WWW at http://www.incaproject.org/en_downloads/21_INCA_Assessor_Manual_eng_final.pdf. Accessed 14/1/09.

Kealey, D.J. and Ruben, B.D. (1983) Cross-cultural personnel selection criteria, issues, and methods. In D. Landis and R.W. Brislin (eds) *Handbook of Intercultural Training. Volume 1: Issues in Theory and Design* (pp. 156–175). New York: Pergamon.

Kelley, C. and Meyers, J. (1999) The cross-cultural adaptability inventory. In S. Fowler and M. Mumford (eds) *Intercultural Sourcebook: Cross-cultural Training Methods* (Vol. 2) (pp. 53–70). Yarmouth, ME: Intercultural Press.

Kim, E.Y. (2002) *Global Intelligence: Seven Pillars for New Global Leaders, A Cross-cultural Reference of Business Practices in a New Korea*. Westport, CT: Quorum Books.

Lado, R. (1961) *Language Testing*. London: Longman.

Lange, D.L. (2003) Future directions for culture teaching and learning: Implications of the new cultural standards and theoretical frameworks for curriculum, assessment, instruction and research. In D.L. Lange and R.M. Paige (eds) *Culture*

as the Core: Perspectives on Culture in Second Language Learning (pp. 337–354). Greenwich, CT: Information Age Publishing.

McNamara, T. (2000) *Language Testing*. Oxford: Oxford University Press.

Oller, J.W. (1979) *Language Tests at School*. London: Longman.

Paige, R.M. (2003) The Intercultural Development Inventory: A critical review of the research literature. *Journal of Intercultural Communication* 6, 53–61.

Phipps, A. and Gonzalez, M. (2004) *Modern Languages: Learning and Teaching in an Intercultural Field*. London: Sage.

Schulz, R.A. (2007) The challenge of assessing intercultural understanding in the context of foreign language instruction. *Foreign Language Annals* 40 (1), 9–26.

Sercu, L. (2002) Autonomous learning and the acquisition of intercultural communicative competence: Some implications for course development. *Language, Culture and Curriculum* 15 (1), 61–74.

Sercu, L. (2004) Intercultural communication competence in foreign language education: Integrating theory and practice. In O. St. John, K. van Esch and E. Schalkwijk (eds) *New Insights into Foreign Language Learning and Teaching* (pp. 115–130). Frankfurt: Peter Lang.

Sercu, L., Bandura, E., Castro Prieto, P., Davcheva, L., Laskaridou, C., Lundgren, U., Méndez García, M. and Ryan, P.M. (2005) *Foreign Language Teachers and Intercultural Competence. An International Investigation.* Clevedon: Multilingual Matters.

Sfard, A. (1998) On two metaphors for learning and the dangers of choosing just one. *Educational Researcher* 27, 4–13.

Sinicrope, C., Norris, J. and Watanabe, Y. (2007) *Understanding and Assessing Intercultural Competence: A Summary of Theory, Research and Practice*. On WWW at www.hawaii.edu/sls/uhwpesl/26(1)/Norris.pdf. Accessed 14/1/09.

Tucker, M.F. (1999) Self-awareness and development using the overseas assignment inventory. In S. Fowler and M. Mumford (eds) *Intercultural Sourcebook: Cross-cultural Training Methods* (Vol. 2) (pp. 45–52). Yarmouth, ME: Intercultural Press.

Chapter 3

Interculturally Savvy or Not? Developing and Assessing Intercultural Competence in the Context of Learning for Business

KAISU KORHONEN

Introduction

Because of the increasing internationalization and globalization of working life in the late 20th century and the beginning of the 21st, new kinds of requirements are being set on employees' task performance and adjustment to other cultures. Advances in transportation and telecommunications as well as economic integration have broken down geographical isolation. When expanding their operations overseas, companies have to balance the prospects of growth and the risk associated with operating in unfamiliar markets. Successful companies, and employees, are those who understand that cultural diversity is more than a language issue; who do not see diversity as a threat but as an opportunity, as something that can be learned, managed and made use of. The *knowledge management* perspective presents culture, not as a source of difference and antagonism, but as a form of organizational, company-specific knowledge. This knowledge can be converted into *tacit knowledge*, which adds value to company activities and is difficult for rivals to copy (Holden, 2002).

All this means that in addition to professional, social and communication competence, today's employees need to achieve intercultural competence (see also Sercu, this volume). The concept of *intercultural competence* incorporates both culture-general and culture-specific knowledge and foreign language skills. It also incorporates a wide range of personal characteristics such as a strong sense of self, low levels of ethnocentrism, openness and empathy, as well as task performance skills such as team and project work skills, management and leadership

skills, presentation skills, negotiation skills and computer literacy. Considering intercultural adjustment, the concept incorporates tolerance of ambiguity and uncertainty connected with new situations and the ability to manage anxiety and stress. In brief, the components of intercultural competence are cognition, affect and behavior.

In the Finnish context, mobility from and to Finland is increasing, including business people travelling throughout the world, expatriates working on international assignments, and immigrants (especially refugees). In many Finnish companies, the working language is no longer Finnish but English. In such intercultural contexts, interlocutors often have a varying command of English and they use different kinds of accents. In spite of this, interlocutors have to be able to interpret both people and events as host nationals do and, when necessary, to negotiate common meanings and understandings. This kind of *participative competence* and *interactive translation* requires willingness to discuss issues in a productive way not only in one's native language but in foreign languages as well (Holden, 2002).

Bennett (1997) uses the term *fluent fool* to describe a person who speaks a foreign language well but does not understand the social or philosophical content of that language. The stereotype of Finnish people is that they are task-oriented, value their own privacy, speak directly and have a high tolerance of silence in conversations (see, e.g. Yli-Vakkuri, 2005). Depending on the context, the Finnish communication style can be a strength or a weakness. The study described in this chapter was an intercultural training experiment with Bachelor of Engineering (BEng) students at the Kajaani University of Applied Sciences, aimed at developing their intercultural competence and prevent them from being 'fluent fools'. This chapter discusses the justification for intercultural training, the implementation of the training study, the outcomes of the experiment and its effectiveness in developing intercultural competence.

Intercultural Training

Intercultural training in classroom conditions

Foreign languages and cultures can be studied in foreign countries and/or in the trainees' native country in classroom conditions. Most students in various institutions of higher education do not participate in exchange programs and do not live through foreign experiences of their own. This means that there is a need for internationalization at home, i.e. intercultural training about the basic concepts, theories and models of intercultural communication within the classroom.

Intercultural training in classroom conditions is not without problems. In a classroom, learning is not based on trainees' real-life experiences of foreign cultures. It is possible that the importance of culture will be trivialized, or that teaching culture will be neglected. The experience of culture can become artificial, rigid and trainer-driven, emphasizing cognition, errors and making corrections. Furthermore, textbooks available for professionally integrated foreign language education have a number of weaknesses. They reflect the worldview of the author(s), and they often focus on teaching customs and conventions to tourists and/or business people without elaborating on what the customs and conventions are based on. They may provide a monocultural perspective, forgetting subcultural differences such as those associated with ethnicity, age, gender, socio-economic status, religion and the different versions of the language. Stereotypes may also be generalized. On the other hand, classroom conditions can provide a safe learning environment in which trainees are allowed to experiment and make errors, too (Kjartansson & Skopinskaja, 2003).

Perspective transformation

One way of developing the various interrelated components of intercultural competence is learning through perspective transformation. *Perspective transformation* is a slow, gradual transformative learning process, including real-life experiences, reflective observation of the experiences, abstract conceptualization, or interpretation, of the experiences, and active experimentation, or behavior. When learning through perspective transformation, *critical reflection* is emphasized. Critical reflection involves how meanings are attributed, i.e. how new experiences are interpreted (Kolb, 1984; Mezirow, 1991; Taylor, 1994).

Being open to the perspectives of others may lead to a wider worldview and being more accepting; it may even lead to a transformation in behavior (Mezirow, 1991). The objective of transformative learning is not, however, to transform a trainee's fundamental personality or basic character, but rather to provide a professional competitive advantage and added value respecting the existing personal and cultural make-up of the trainee (Kealey & Protheroe, 1996).

Intercultural Training Methods

Culture-general, didactic and experiential

In intercultural training, both didactic culture-general and experiential culture-general training methods are favored (Gudykunst *et al.*, 1996).

If training is based on Kolb's (1984) experiential learning cycle, or the perspective transformation process (Mezirow, 1991; Taylor, 1994), trainees' learning style preferences can be taken into account, for example: concrete experience draws on class discussions; in reflective observation use is made of learning diaries; abstract conceptualization involves lectures; and active experimentation requires projects (Bennett, 2003, in Fowler & Blohm, 2004). According to Mendenhall *et al*. (2004), training methods used frequently include lectures, presentations, culture assimilators, class discussions, role play, videos, films, field trips, reading materials, simulations, case studies and self-assessment inventories, in this order. Fowler and Blohm (2004) claim that any method involves cognition, affect and behavior to some extent: the key is to decide which component of intercultural competence will contribute most significantly to the desired outcome, and focus on that.

One increasingly important development in intercultural training is the way in which many traditional intercultural training methods are being brought into the 21st century with computers. Computer-based training refers to all types of learning available through CDs, DVDs and online programs. In training, much of the information transfer can be done online, and time can be used for integrating and applying the information that has been provided online. Class discussions provide the space to share information. Computer-based learning can be used as both pre- and post-classroom work (Fowler & Blohm, 2004; Korhonen, 2002).

Critical incidents and the Culture Assimilator

Critical incidents were first used by Flanagan in the 1950s in psychology for developing job descriptions (Flanagan, 1954). According to Wight (1995), the first use of critical incidents in a cross-cultural context, i.e. intercultural training, was in the 1960s. *Critical incidents* are short descriptions of contexts in which there is a problem arising from cultural differences between the interacting parties, or where there is a problem of intercultural adjustment.

The description of each incident provides information to set the stage, details what happened and may also address the emotions and reactions of the parties involved. The cultural differences that the parties bring to the situation are discovered or revealed as part of the exercise. Critical incidents appeal to learning styles based on concrete experience and reflective observation (Fowler & Blohm, 2004).

According to Albert (1995), the intercultural sensitizer (ICS) was developed by Fiedler, Osgood, Stolurow and Triandis in the early 1960s.

The ICS is an instrument constructed to sensitize trainees to possible cultural differences consisting of a number of critical incidents, approximately 20 incidents being a sufficient number for a single training course. Albert (1995) emphasizes that the construction of an ICS requires culturally valid information.

In the early 1980s, Brislin and his associates decided to call the ICS method a Culture Assimilator, and also introduced the *Culture-General Assimilator* (Cushner & Brislin, 1996). A Culture-General Assimilator is a programmed intercultural learning package, which does not focus on one specific culture (cf. a Culture-Specific Assimilator), but includes critical incidents from a variety of cultures involving members from the trainees' culture and from the target cultures (Cushner & Landis, 1996). Each incident is equipped with four to five alternative attributions, or interpretations, in response to the situation, as well as explanations and feedback. One response is most often preferred by people from the target culture (Fowler & Blohm, 2004).

To make a Culture Assimilator more culture-general, instead of the given alternative attributions, the incident can be followed by a question about the thoughts, emotions and/or behavior of the member of the trainees' culture. The trainees critically reflect on the incident and write an essay, which is then assessed using previously defined criteria (Albert, 1995).

The idea of implementing a Culture Assimilator with computer technology was introduced in the 1960s (Cushner & Landis, 1996; Triandis, 1995). According to Fowler and Blohm (2004), most existing Culture Assimilators are now available electronically.

Assessing the Effectiveness of Intercultural Training

Alternative assessment methods

Empirical assessment should be an intrinsic part of intercultural training to provide information on how effective and satisfactory the training was from both an individual and collective point of view. *Summative assessment* is used to define the extent to which the objectives set on the training were achieved, the focus being on the degree to which the trainees were affected by the training, i.e. its effectiveness (Blake *et al*., 1996; Brislin & Yoshida, 1994; Byram, 1997). In *formative assessment*, information is gathered on trainee satisfaction to perceive whether the training was worthwhile, what contents were important, what materials should be dropped and/or added and how to develop the training methods further. To cover the entire learning process,
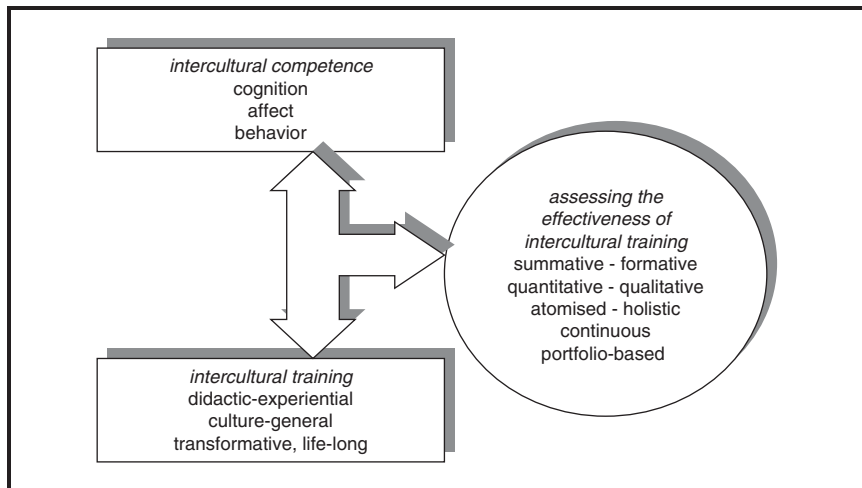
Figure 3.1 Intercultural competence, intercultural training and assessment

including the various interrelated components of intercultural compe-
tence, both *holistic* and *atomic* assessment is needed (Linn & Gronlund,
2000). Figure 3.1 shows the relationship between intercultural compe-
tence, intercultural training and assessment in this area.

Methods for assessing intercultural training and its effectiveness
include self-reports, trainer and peer observations, judgments of out-
siders (e.g. employers, co-workers, host nationals), interviews, question-
naires, tests, experiments, video recording of interaction and measures of
overt behavior. Assessment methods also include learning diaries, critical
incidents, case studies, role plays, simulations, interactive computer
programs and portfolios (Bhawuk & Triandis, 1996; Blake *et al.*, 1996;
Byram, 1997).

In assessment, objectivity is usually emphasized. How culture is
understood and experienced is, however, subjective, and, consequently,
assessing the effectiveness of intercultural training cannot be objective in
the conventional meaning of the word. Besides culture, factors such as the
cultural identity, age, gender, socio-economic status and learning style of
the trainee may reduce the objectivity of assessment (Kjartansson &
Skopinskaja, 2003). Defining the outcomes of intercultural training in
frequencies and percentages is not always possible and in addition
to *quantitative* measures, *qualitative* measures are required (Linn &
Gronlund, 2000). Because it is not clear which training method is most
effective overall in developing intercultural competence and which

methods are most effective in specific contingencies, alternative assessment methods should be applied (Mendenhall *et al.*, 2004).

Effectiveness of the Culture Assimilator

The Culture Assimilator is the most widely investigated method in intercultural training. Previous research includes both the method and its effectiveness in developing intercultural competence and preparing people such as exchange students and various professionals for interaction and adjustment in culturally diverse contexts (Albert, 1995; Blake *et al.*, 1996; Cushner & Brislin, 1996; Fowler & Blohm, 2004; Mendenhall *et al.*, 2004).

The Culture Assimilator is often classified as a cognitive method because it focuses on the acquisition of knowledge or information by the trainee. According to Albert (1995), the process by which the information is acquired is, however, in a sense experiential: information is acquired by a trial-and-error process, which simulates the experience of entering a new culture, but without the risks of failure and embarrassment. Albert (1995) goes on to say that because the materials in Culture Assimilators also cover the affect and behaviors of the people involved, all the components of intercultural competence are brought together in this method, both in the contents of what is learned and the process of learning. Furthermore, the method uses the behavioral techniques of feedback and reinforcement (see also Baxter & Ramsey, 1996; Bennett, 1995; Cushner & Landis, 1996; Wight, 1995). According to Kealey and Protheroe (1996), the method is cognitive even if it aims at some degree of interpersonal skills development (see also Cushner & Landis, 1996; Triandis, 1995).

Frameworks of the Intercultural Training Experiment

The training program

To develop professionally integrated foreign language (English) education at the Kajaani University of Applied Sciences, Finland, to meet the requirements set by today's international and global working life, and to establish the linkage between intercultural training and the outcomes of such training, an intercultural training experiment was implemented in cooperation with Bachelor of Engineering (BEng) students ($n = 117$) during the years 2000–2002 (see Korhonen, 2002). The experiment was part of the students' compulsory English studies. In addition to the course in intercultural communication, the students complete a course in engineering English (including reading comprehension, reporting, summarizing

and technical documentation) and in business English (including presentations, telephoning, written business messages, meetings and negotiations).

The course in intercultural communication was divided into two consecutive periods: face-to-face training and self-paced study. Because there was no appropriate training material available for the face-to-face training period, a handout containing basic concepts, theories and models of intercultural communication was developed. The training methods were reflective class discussions and PowerPoint-supported presentations. The face-to-face training period took eight weeks and it consisted of three 45-minute lessons a week.

For the self-paced study period, a web-based Culture-General Assimilator running in the Windows environment was developed in cooperation with a programmer. The Assimilator consists of 25 critical incidents adapted from Cushner and Brislin (1996). In the USA, critical incidents have a long tradition (Albert, 1995; Flanagan, 1954), but in Finland they were and still are rather an unknown method. Because all the empirical research about the effectiveness of the method had not documented affective and/or behavioral changes, it was decided to test it. There was simultaneously a need to develop a web-based self-paced study and it was decided to implement the Assimilator as a web-based version.

The critical incidents chosen for the Assimilator were customized for the Finnish BEng students to be worked out either at school or at home. When customizing the incidents, a great deal of the content, including names of countries and people, job titles, as well as business conventions, was changed in order to be more suitable for the Finnish audience. Although it is possible that the validation of the original incidents suffered, for practical reasons, it was not possible to have target-culture specialists to assess the validity of the customized incidents.

The *Business or Pleasure?* incident is an example of a customized critical incident.

Business or Pleasure?
Ville, a top sales negotiator in an electronics company, was asked to present the company in Riyadh, Saudi Arabia. Together with a local agent called Ahmed bin Muhammad Ville planned his trip with extreme care. At the airport he was welcomed by Mr Hassan.

As soon as they got in the car, Ville began explaining some of his ideas to Mr Hassan. Mr Hassan, however, kept changing the subject and talked about the weather and asked questions about Ville and his

family. He told Ville that instead of flashy urban life he prefers the traditional ways of the Bedouin: camels, horses, and falconry.

When planning the trip, Ahmed bin Muhammad, the agent, had set up an appointment for the morning after Ville's arrival. Mr Hassan, however, informed Ville that the meeting would be two days later. Ville's hosts wanted him to recover from his journey first; perhaps see some sights and enjoy their hospitality. Ville said that he was quite fit and prepared to give his presentation as agreed. Mr Hassan seemed taken aback at this, but said that he would discuss it with his superiors.

The Saudis agreed to meet Ville. After some chatting and preliminaries with many people coming and going, they suggested that they could continue soon after Ville had some more time to recover. Besides, it was an hour of prayers. During the next few days, they said that they wanted to discuss the details of Ville's presentation but they seemed to spend a lot of time on inconsequentials. This began to annoy Ville as he thought that the deal could have been closed several days ago. It would be Ramadan soon, and he just did not know what the Saudis were driving at.

How would you help Ville to interpret the Saudis' behavior?
The alternative attributions

(1) The Saudis were trying to check on Ville and his company by finding out more information.
(2) The Saudis are not used to working hard and just wanted Ville to relax more.
(3) The Saudis were not really interested in the products of Ville's company and were just putting him off.
(4) Ville's Finnish perspective was concerned with getting the job done, whereas the Saudis had the perspective of building a relationship with Ville and his company.
(Adapted from Cushner & Brislin, 1996)

In the Assimilator, hypertext with hotwords containing both text and picture elements was used. The text, for example, contains an English-Finnish glossary and definitions of intercultural concepts, while the picture elements, for example, contain pictures of cities, people and events.

Critical incidents were not discussed during the face-to-face training period, but the Culture-General Assimilator method was explained to the students before they started the self-paced study period. The students

were also reminded of the dangers of stereotyping and overgeneraliza-
tion. When analyzing the critical incidents and writing their essays and
their rationales, the students were asked to consider the context and to
include the following interrogatives:

- Who: The participants' nationality, gender, age, and role relation-
  ships
- Where: Location and physical conditions
- When: Time
- What: The sequence of events; what is said and done
- How: Thoughts, actions, and emotions of the person(s) from the
  students' culture
- How: Nonverbal aspects of the interaction

  (Adapted from Baxter & Ramsey, 1996; Dant, 1995; Wight 1995)

The self-paced study period started immediately after the face-to-face
training period and it took three weeks. The time individual students
spent on self-paced study varied from 1 to 12 hours, the average time
being 2.3 hours.

Objectives and measuring instruments

The primary objective of the intercultural training experiment was to
assess the effectiveness of intercultural training in developing the BEng
students' intercultural competence (summative assessment). This in-
cluded finding out whether the self-paced study period had any effect on
the students' behavior when compared with the situation after the face-
to-face training period, in other words, whether the self-paced study
period provided any added value. The secondary objectives were to find
out what kind of communicators the students were and would like to be,
whether they were motivated to develop their intercultural competence
as part of their professional qualifications (self- and peer assessment),
and what they thought about the intercultural training course, i.e. their
training satisfaction (formative assessment).

When assessing the students' learning outcomes, a pre- and post-
design was applied, and the students were first assessed before the face-
to-face training period and then after both the face-to-face training period
and the self-paced study period. Because of validity and reliability
concerns, the usefulness of the experiment was limited to measuring
collective characteristics rather than individual ones.

Before the face-to-face training period, the students filled in the *first
questionnaire* with structured and open-ended questions to provide

demographic data as well as data connected with their prior international experiences and motivation to study and/or work abroad. The first questionnaire also assessed the intercultural element of the students' cognition, affect and behavior.

- *Cognition was assessed by means of* declarative statements with a Likert-type scale. Aspects assessed included the students' factual knowledge about the concept of culture, intercultural communication as a process, low and high context communication, verbal and nonverbal communication, concepts used to compare cultures such as values and norms, Geert Hofstede's dimensions of national culture and intercultural adjustment as a process.
- *Affect was also assessed through Likert scales*. The aspects assessed included attitudinal aspects such as the students' worldview, approach- and avoidance-orientedness, contacts with foreigners, prejudices toward immigrants and the degree of ethnocentrism versus ethnorelativism.
- *Behavior*, or skills, was assessed by means of five critical incidents from the Assimilator.

In the first questionnaire, declarative statements with a Likert-type scale were also used to map out the students' opinions of their own communication skills with foreigners. To map out how satisfied the students were with their own communication skills, open-ended questions were used.

After the face-to-face training period, the students filled in the *second questionnaire*. This was identical to the first questionnaire, but omitted the students' background information and self-assessment connected with their communication skills. The students also filled in the *third questionnaire* with open-ended questions to assess the training course as well as their own and other students' participation in the course.

The students filled in the *fourth questionnaire* after the self-paced study period. The questionnaire contained the five critical incidents from the first and second questionnaire. The students did not know that they would be the same incidents as before. Finally, the students filled in the *fifth questionnaire* with declarative statements and a Likert-type scale about the self-paced study period and the Assimilator.

For each critical incident used in the questionnaires, five criteria were defined in advance to be used when assessing the rationales written by the students. One criterion mentioned by the student gave him/her one grade, 1 being the lowest and 5 the highest grade (poor to excellent). For example, the predefined assessment criteria for the critical incident

Business or Pleasure? described above were: (1) *individualism versus collectivism* (Hofstede & Hofstede, 2005); (2) *concept of time* including monochronic versus polychronic time, use of time, task-orientedness versus relationships-orientedness; (3) *concept of space* including physical space, public versus private space, use of space (Hall, 1966, 1976); (4) *business conventions* including the influence of religion; and (5) *low versus high context communication* including socializing, small talk and safe versus unsafe topics for small talk.

Outcomes and Discussion

As explained above, in the training experiment both the effectiveness of the intercultural training course in developing the cognitive, affective and behavioral components of the BEng students' intercultural competence at the collective level (summative assessment) and the training itself (formative assessment) were assessed. In addition, the students' attitudes toward intercultural competence and their motivation to develop it as part of their professional qualifications (self- and peer assessment) were mapped out. The students were pre- and post-tested using a series of five questionnaires. This section provides a summary of the results of the study.

To determine the significance of the tested pre- and post-differences in the students' cognition, percentage scores were used. After the face-to-face training period, on the basis of the percentage scores of the variables measuring the *cognitive* component of the students' competence, there was considerable change, i.e. perspective transformation. For example, considering the variable called *The Concept of Culture* and the statement 'I know different definitions for the concept of culture', before the face-to-face training period more than one third of the students ($n = 42$; 36%) were aware of the everyday aspect of culture, but afterwards nearly three quarters of the students ($n = 86$; 74%) knew how to define culture for the purposes of intercultural communication. When asking the students about the benefits of the training course, more than three quarters ($n = 89$; 76%) also thought that their knowledge (i.e. the cognitive element) had increased.

Percentage scores were also used to determine the significance of the tested pre- and post-differences in the students' affect. In contrast to cognition, there was not much change in the percentage scores of the variables measuring the *affective* component of the students' intercultural competence. For example, considering the variable called *Approach- and Avoidance-Orientedness* and the statement 'Foreign neighbors are nice',

before the face-to-face training period approximately half of the students ($n = 57$; 49%) accepted foreigners as neighbors and afterwards the percentage was only a little higher ($n = 65$; 56%); considering the variable called *Ethnocentrism versus Ethnorelativism* and the statement 'We don't need foreigners in Finland', before the face-to-face training period more than two thirds of the students ($n = 95$; 81%) thought that immigrants are needed in Finland and afterwards the percentage was almost the same ($n = 96$; 82%). When asking the students about the benefits of the training course, one third of them ($n = 36$; 31%) thought that their attitudes had become slightly more tolerant.

To analyze the development of the *behavioral* component of the students' intercultural competence, five critical incidents were used. Each incident was equipped with a set of alternative attributions, one of the attributions being the 'best', i.e. the most suitable, considering the context. The incidents were not discussed during the face-to-face training period. When comparing the students' choices of the alternative attributions and the frequencies of the 'best' attributions before and after the face-to-face training period, the frequencies of the 'best' attributions were higher in two incidents out of five after the face-to-face training period, i.e. there was some positive change.

The face-to-face training period was followed by a web-based self-paced study period with the Culture-General Assimilator consisting of 25 critical incidents. The objective was to find out whether the self-paced study period provides any added value considering the development of the students' behavior and skills. To determine the statistical significance of the tested differences in the mean scores of the predefined criteria from 1 (poor) to 5 (excellent) before and after the self-paced study period, the Wilcoxon test was used. On the basis of the test, the mean scores for all five incidents were somewhat higher after the self-paced study period than before it and there was a quantitative change of less than one grade, from poor to satisfactory minus. Qualitatively, the contents of the rationales written by the students enhanced and/or became more accurate from the theoretical point of view. Therefore, it can be argued that the Web-based self-study period supplemented the face-to-face training period and consequently, it provided added value when aiming at perspective transformation.

On the basis of the percentage scores, most students ($n = 100$; 86%) thought that when communicating in Finnish, they have no major problems; the main factor preventing them from communicating with foreigners is their insufficient knowledge of foreign languages, especially English; potential communication problems with foreigners are very

difficult to solve; interpreting foreigners' emotions is especially difficult; and an ideal communicator is clear, fluent and able to communicate with different kinds of people interactively. Most students ($n = 114$; 97%) accepted intercultural competence as an integral and equal part of the professional qualifications that engineers need in today's working life. They thought that intercultural training was both useful and interesting ($n = 114$; 97%) and self-paced study with the Culture-General Assimilator was serious studying, not edutainment ($n = 93$; 79%). They found themselves self-directed adult learners able to take responsibility for their own learning ($n = 91$; 78%). On the basis of peer assessment, most students ($n = 101$; 86%) were, however, not able to do this.

Mendenhall *et al.* (2004) claim that intercultural training is usually, but not always, effective in enhancing trainees' cognition, has a mixed record in facilitating changes in trainees' affect and is occasionally effective in changing trainees' behavior. Fowler and Blohm (2004) agree that the most common outcome of intercultural training is knowledge acquisition, although attitudinal changes and skills development are possible. According to Bhawuk and Brislin (2000, in Mendenhall *et al.*, 2004), the Culture Assimilator is an effective training tool in terms of cognitive effect, and it has some positive impact on affective and behavioral criteria as well. In the training experiment discussed (Korhonen, 2002), the BEng students' cognition developed considerably, whereas their affect only developed a little. The students' ability to analyze and solve intercultural conflicts occurring in the critical incidents also developed a little. Instead of a quantitative change, it might be preferable to speak about a qualitative change, or a small positive leap that took place in the students' insight.

Conclusion

Taken cautiously, pre- and post-test comparisons are helpful in giving the intercultural trainer, or facilitator, a general idea of the direction of any change or development in trainees' performance. Tests also provide ideas on how to improve training. On the basis of the pre- and post-comparison discussed, intercultural competence can be developed through intercultural training. On the other hand, intercultural training is not necessarily effective in attaining all the objectives associated with it in practice, but its effectiveness varies depending on the type of component under consideration, i.e. whether the component is cognition, affect or behavior (Korhonen, 2002; Mendenhall *et al.*, 2004). To be able to conclude whether the BEng students' behavior in terms of actual

interaction and intercultural adjustment changed, longitudinal measures of their subsequent performance, especially on-the-job in the target cultures, would have been required. However, for practical reasons this was not possible.

The intercultural training process, i.e. perspective transformation, is slow, gradual and difficult to assess (Korhonen, 2002; Mezirow, 1991; Taylor, 1994). Therefore, its development should be started as early as possible at home and at school. Taylor (1994) argues that a culture shock is a necessary precondition to change in the perspective transformation process. At least some change seems to be possible without the experience of entering a new culture and culture shock. A change involving a culture shock may be more prominent, but intercultural training in classroom conditions can also be a catalyst for change.

The Culture-General Assimilator method may produce changes in trainees' behavior, although it seems that the method is cognition-emphasized. An increase of knowledge is a useful variable to measure but, considering today's working life, more emphasis on and exploration of behavioral changes, intercultural adjustment and task performance is required. Behavioral measures provide the most convincing evidence of the effectiveness of intercultural training, i.e. when evidence can be presented that trainees have changed their behavior in desirable directions. To improve the validity and reliability of the measuring instruments, observation and interviews involving host nationals in the target cultures are also required (Korhonen, 2002; Mendenhall *et al.*, 2004).

In spite of the assessment difficulties, intercultural training and its assessment must not be bypassed in the interest of technically reliable measurement. If the assessment of the effectiveness of training is totally neglected, it may lead to a conclusion that culture is secondary and intercultural competence does not matter (Kjartansson & Skopinskaja, 2003; Korhonen, 2002).

The trend in intercultural training is for coaching, consulting and/or advising. Facilitating, guiding and self-directedness, i.e. putting the trainee in charge, are keys to success. If the objective of the training is intercultural competence, in other words, sensitivity, effectiveness, performance in multicultural contexts and establishing relationships, Web-based self-paced study is not sufficient, and face-to-face training is also required (Fowler & Blohm, 2004; Korhonen, 2002).

Teachers of foreign languages should be provided with further training so that they can become intercultural facilitators to help their students to become intercultural actors who are able to engage with other

intercultural actors in communication and interaction, which is different from those between native speakers (Byram, 1997; Korhonen, 2002).

Considering the future of intercultural training, more empirical assessment of training outcomes is required to convince those with financial resources in education and business of the benefits of training (Korhonen, 2002; Mendenhall *et al*., 2004).

References

Albert, R.D. (1995) The intercultural sensitizer/culture assimilator as a cross-cultural training method. In S.M. Fowler and M.G. Mumford (eds) *Intercultural Sourcebook: Cross-cultural Training Methods* (Vol. 1) (pp. 157–167). Yarmouth, ME: Intercultural Press.

Baxter, J. and Ramsey, S. (1996) Improvising critical incidents. In N.H. Seelye (ed.) *Experiential Activities for Intercultural Learning* (Vol. 1) (pp. 211–218). Yarmouth, ME: Intercultural Press.

Bennett, M.J. (1995) Critical incidents in an intercultural conflict-resolution exercise. In S.M. Fowler and M.G. Mumford (eds) *Intercultural Sourcebook: Cross-cultural Training Methods* (Vol. 1) (pp. 147–156). Yarmouth, ME: Intercultural Press.

Bennett, M.J. (1997) How not to be a fluent fool: Understanding the cultural dimension of language. In A.E. Fantini (ed.) *New Ways in Teaching Culture* (pp. 16–21). Alexandria, VA: TESOL.

Bhawuk, D.P.S. and Triandis, H.S. (1996) The role of culture theory in the study of culture and intercultural training. In D. Landis and R.S. Bhakat (eds) *Handbook of Intercultural Training* (2nd edn) (pp. 17–33). Thousand Oaks, CA: Sage Publications.

Blake, B.F., Heslin, R. and Curtis, S.C. (1996) Measuring impacts of cross-cultural training. In D. Landis and R. Bhakat (eds) *Handbook of Intercultural Training* (2nd edn) (pp. 165–182). Thousand Oaks, CA: Sage Publications.

Brislin, R.W. and Yoshida, T. (1994) *Intercultural Communication Training: An Introduction*. Thousand Oaks, CA: Sage Publications.

Byram, M. (1997) *Language Teachers, Politics and Cultures*. Clevedon: Multilingual Matters.

Cushner, K. and Brislin, R.W. (1996) *Intercultural Interactions. A Practical Guide* (2nd edn). Thousand Oaks, CA: Sage Publications.

Cushner, K. and Landis, D. (1996) The intercultural sensitizer. In D. Landis and R.S. Bhakat (eds) *Handbook of Intercultural Training* (2nd edn) (pp. 185–202). Thousand Oaks, CA: Sage Publications.

Dant, W. (1995) Using critical incidents as a tool for reflection. In S.M. Fowler and M.G. Mumford (eds) *Intercultural Sourcebook: Cross-cultural Training Methods* (Vol. 1) (pp. 141–146). Yarmouth, ME: Intercultural Press.

Flanagan, J.C. (1954) The critical incident technique. *Psychological Bulletin* 51 (4), 327–358.

Fowler, S.M. and Blohm, J.M. (2004) An analysis of methods for intercultural training. In D. Landis, J.M. Bennett and M.J. Bennett (eds) *Handbook of Intercultural Training* (3rd edn) (pp. 37–84). Thousand Oaks, CA: Sage Publications.

Gudykunst, W.B., Guzley, R.M. and Hammer, M.R. (1996) Designing intercultural training. In D. Landis and R.S. Bhakat (eds) *Handbook of Intercultural Training* (2nd edn) (pp. 61–79). Thousand Oaks, CA: Sage Publications.

Hall, E.T. (1966) *The Hidden Dimension* (2nd edn). Garden City, NY: Anchor Press Doubleday.

Hall, E.T. (1976) *Beyond Culture*. New York: Anchor Press Doubleday.

Hofstede, G. and Hofstede, G.J. (2005) *Cultures and Organizations. Software of the Mind* (2nd edn). New York: McGraw-Hill.

Holden, N.J. (2002) *Cross-cultural Management. A Knowledge Management Perspective*. Harlow: Prentice Hall.

Kealey, D.J. and Protheroe, D.R. (1996) The effectiveness of cross-cultural training for expatriates: An assessment of the literature of the issue. *International Journal of Intercultural Relations* 20 (2), 141–165.

Kjartansson, R. and Skopinskaja, L. (2003) Incorporating intercultural communicative competence in language teacher education. European Centre for Modern Languages. Council of Europe Publishing.

Kolb, D.A. (1984) *Experiential Learning. Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice Hall.

Korhonen, K. (2002) Intercultural competence as part of professional qualifications. A training experiment with bachelor of engineering students. Jyväskylä Studies in Communication 17. PhD thesis, University of Jyväskylä.

Linn, R.L. and Gronlund, N.E. (2000) *Measurement and Assessment in Teaching* (8th edn). Mahwah, NJ: Prentice Hall.

Mendenhall, M.E., Stahl, G.K., Ehnert, I., Oddou, G., Osland, J.S. and Kühlmann, T.M. (2004) Evaluation studies of cross-cultural training programs: A review of the literature from 1988 to 2000. In D. Landis, J.M. Bennett and M.J. Bennett (eds) *Handbook of Intercultural Training* (3rd edn) (pp. 129–143). Thousand Oaks, CA: Sage Publications.

Mezirow, J. (1991) *Transformative Dimensions of Adult Learning*. San Francisco, CA: Jossey-Bass.

Taylor, E.W. (1994) A learning model for becoming interculturally competent. *International Journal of Intercultural Relations* 18 (3), 389–408.

Triandis, H.C. (1995) Culture-specific assimilators. In S.M. Fowler and M.G. Mumford (eds) *Intercultural Sourcebook: Cross-cultural Training Methods* (Vol. 1) (pp. 179–186). Yarmouth, ME: Intercultural Press.

Wight, A.R. (1995) The critical incident as a training tool. In S.M. Fowler and M.G. Mumford (eds) *Intercultural Sourcebook: Cross-cultural Training Methods* (Vol. 1) (pp. 127–140). Yarmouth, ME: Intercultural Press.

Yli-Vakkuri, V. (2005) Politeness in Finland. Evasion at all costs. In L. Hickey and M. Stewart (eds) *Politeness in Europe* (pp. 180–201). Clevedon: Multilingual Matters.

Chapter 4

Eliciting the Intercultural in Foreign Language Education at School

ANTHONY J. LIDDICOAT and ANGELA SCARINO

Introduction

Intercultural competence in the context of language education imposes a number of challenges for assessment. In part, the difficulty lies in the diverse understandings of the construct to be assessed (see, for example, Sercu, this volume), but even then, eliciting the performances of interculturality has proved to be problematic and assessment approaches may elicit only part of the overall construct, considering this to reflect the whole. Byram *et al.* (2002) argue that while assessing facts or knowledge and understanding about facts is relatively straightforward, any assessment of knowledge will not capture what it means to be intercultural (Liddicoat, 2002, 2005; Liddicoat *et al.*, 2003). Assessing knowledge does not capture the 'ability to make the strange familiar and the familiar strange' (Byram *et al.*, 2002: 14), nor does it capture the affective and ethical dimensions of what it means to be intercultural.

The most elaborated model of intercultural competence and its assessment is the model of *savoirs* (*savoir comprendre*, *savoir apprendre*, *savoir faire*, *savoir s'engager*, *savoir être*) developed by Byram and Zarate (Byram, 1997; Byram & Zarate, 1994) and its use as a basis for assessment as elaborated by Byram (1997) and extended by Sercu (2004). Sercu (2004: 76) proposes that the Byram model of *savoirs* be extended to include 'a meta-cognitive dimension', that is self-regulating mechanisms that enable students to plan, monitor and evaluate their own learning processes. In addition to the limitation noted by Sercu, the model of *savoirs* does not elaborate on the important ways in which language affects culture and culture affects language and how this is understood by the learner.

The difficulty with any modelling is that, inevitably, it involves some form of categorisation or breaking down into component parts. This

componential conceptualisation then mitigates against a holistic under-standing of the capability and equally, a holistic view of assessment.

The Assessment Cycle

While this chapter is centrally concerned with issues of eliciting, the practice of eliciting is one of a set of interrelated processes in an overall assessment cycle (Figure 4.1).

The initial starting point for the assessment cycle is the conceptua-lisation of what will be assessed and for what purpose. This conceptualisation involves multiple assumptions, which may be more (or less) clearly articulated by the assessor, about the construct being assessed. In any investigation of the assessment of the intercultural in languages education, it is the conceptualisation of the construct itself that will influence (and be influenced by) the process of elicitation. Jaeger (1999) argues that the problem with assessing the intercultural may lie more in problems of conceptualising what is being assessed than the assessment itself. The first issue, therefore, in dealing with eliciting the intercultural is to articulate what the intercultural itself is for the purposes of language education.

The conceptualisation of the construct and the purpose of the assessment inform the process of eliciting. The elicitation itself will consist of tasks or procedures designed to make the construct visible in the performance of a learner. Choices made in the design of elicitation
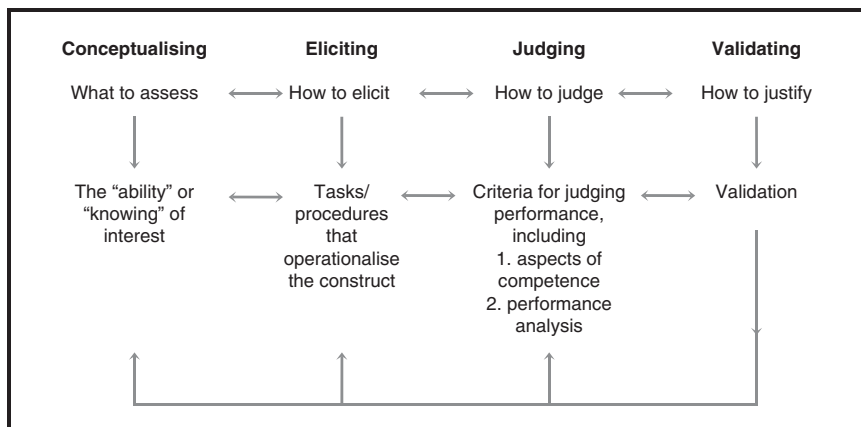

Figure 4.1 The assessment cycle
(*Source*: Scarino, 2006)

will determine how well the construct can be assessed and how much of the construct will be made visible. Judgment is based on the assessor's construction or understanding of the criteria that make performance judgeable. The elicitation process must be designed to reflect and capture the criteria on which judgment is to be made. Moreover, these criteria themselves will be informed by the overall conceptualisation of the purpose of assessment and the construct being assessed. Finally, the judgments made on the learner's performance must be justified as accounts of the construct being assessed. This means that performance must be judged not only against criteria, but also the inferred relationship drawn between performance and criterion must be defended and defensible. Inferences based on observations of performance must be shown to be adequate and appropriate given the empirical evidence of the learner's performance and the theoretical perspectives brought to the assessment process (Messick, 1989).

Understanding the assessment as a cycle means that it is not a linear movement from conceptualisation to validation. It is instead a cyclical process in which each element informs, interrogates and modifies each of the other processes.

The Construct of the 'Intercultural' and Conceptions of Assessment

The construct of the intercultural

As argued above, eliciting any dimension of student learning depends on the conceptualisation of the construct of interest and the intercultural itself is subject to varied conceptualisations (Crichton & Scarino, 2007; Sercu, this volume). One way of understanding the construct of the 'intercultural' is as factual, objective knowledge, which is removed from people as constructors and users of that knowledge (Liddicoat, 2002). Within this conception, language use for intercultural communication is treated in a decontextualised way. The process of learning is understood as a process of acquisition of facts, abstracted and generalised – the 'acquisition metaphor' (Sfard, 1998). The assessment paradigm best aligned with this conception of the intercultural is the psychometric paradigm (see Mabry, 1999, for a discussion of assessment paradigms), with its focus on standardised testing of content through objective procedures, in which the basic approach to understanding student learning is through comparison with either the performance of other students (norm-referenced testing) or a predetermined standard or cut-score (criterion-referenced testing).

Another way of understanding the 'intercultural' is as contextual and personalised knowledge. Within this understanding, the 'intercultural' is understood as involving an ongoing, contextualised process of developing and elaborating personalised knowledge. An integral part of this understanding involves the examination of one's own linguistic and cultural practices as a participant member of human society, because these practices cannot be divorced from the contexts and processes of developing and using one's own language(s) and culture(s) as an ongoing project. The process of learning within this conception is understood as occurring through a process of participation to construct understanding in interaction with more knowledgeable others – the 'participation metaphor' as described by Sfard (1998). The assessment paradigm that is best aligned with this conception of the 'intercultural' is contextual and personalised with a focus on assessing curriculum-related, authentic and student-sensitive content through both objective and subjective, constructed-response procedures. Such procedures are evaluated subjectively and student learning is understood as personally unique for each individual. Students are assessed on the curriculum that they actually experience and that is personally meaningful to them, rather than on standardised content; it is performed in ways that allow for the best opportunity for students to show what they know. Within this assessment paradigm, understood as 'alternative' assessment, personal or self-evaluation is critical because of the importance, ultimately, of self-knowledge as the basis for all understanding.

The conceptualisation of the intercultural proposed here is multi-faceted. It is manifested as and through an ethical commitment to the acceptance and valuing of language and culture within and across languages and cultures. This ethical commitment involves reacting to encounters with diverse others in constructive ways. It is not a passive, observational approach to difference, but rather an active 'being in diversity', in which diversity is not an external reality, but rather a community in which one lives and acts. The intercultural is manifested through language in use, through interpreting and expressing meaning across cultural boundaries in dialogue with self and others, drawing on awareness and knowledge gained through previous experience, recognising the possibility of multiple interpretations of messages and the culturally embedded nature of meaning. Understanding can only emerge from interpersonal dialogue and external observation of the self and other.

As an interpersonal phenomenon, interculturality is predicated on the development of reciprocity in interaction, which recognises one's own multiple roles and responsibilities and is sensitive to, and

accommodating of, those of one's interlocutors. This means that to be intercultural involves the continuation of intercultural learning through experience and critical reflection. There can be no final end point at which the individual achieves the intercultural state, but rather to be intercultural is by its very nature an unfinishable work in progress of action in response to new experiences.

Assessing the intercultural

One approach to assessing interculturality is through the assessment of intercultural communicative behaviours. In this way, intercultural competence is treated as if it were a separable language skill, analogous to the conventional macro-skills of speaking, listening, reading and writing (Damen, 1987). The tasks involved are forms of enactment tasks (Sercu, 2004) in which a learner is asked to perform the role of a communicator in an intercultural situation. These enactment tasks elicit performance in an intercultural situation, but it is often difficult to judge on the basis of such tasks whether or not this performance is in fact intercultural. Conformity to a set of cultural norms of another group shows very little about the knowledge and dispositions, i.e. the interpretive resources, that the learner brings to the task and which contribute to the performance. In fact, in assessment contexts in which conformity is rated highly, such conformity may be based on little more than achieving completion of the task. In order to become intercultural, such tasks need to add to the elicited performance elements of reflection that are at play.

A second tradition in assessing the intercultural is the elicitation of the learners' intercultural dispositions or understandings. There are three different models proposed to assess intercultural skills:

(1) Attitudinal tests (e.g. Cadd, 1994).
(2) Culture assimilator tests (e.g. Brislin *et al.*, 1986).
(3) Cultural awareness tests (Byram & Morgan, 1994).

These approaches to eliciting the intercultural understanding of learners recognise the dimension of the intercultural that is related to values, values clarification, understanding different positions, etc. They therefore address what is missing in many of the communicatively focused tasks in that they seek to understand the dispositions the learner brings to constructing the intercultural. For languages education, however, the problem with such tests is that what they elicit is decontextualised and generalised dispositions that are little connected

with action or with the language being learnt as an element of the intercultural experience and repertoire of the learner. Moreover, what is elicited tends not to have a developmental perspective in that learners either possess or do not possess the attitudes, values or knowledges being elicited. Once these have been demonstrated, it can be difficult to see what educative progression there can be in such assessment.

The intercultural in communication is therefore not a simple equation of performance and competence, but rather a reflective recognition of what is at play in a particular instance of communication. This means that language assessment comes to include the critical perspective the learners bring to interaction as language users. The important educational outcome of intercultural learning is not just the clarification of one's own values, but also a process of understanding and negotiating values in interpersonal contexts across cultures. The dispositions of the interculturally aware individual are not intercultural simply because they have been clarified and understood, but rather because they are seen as a relative positioning among a field of possible values with which the individual needs to engage. Rather than eliciting values, an intercultural approach to assessing language learners needs to elicit the ability to gain feedback on one's own initial positioning from others, and being able to respond to this feedback and to the initial positioning in such a way as to show the ability to reflect on one's own values and those of others from multiple perspectives.

In languages education, reflection on dispositions and understandings needs to be located within the context of the students' linguistic and cultural repertoire that includes their own language and the additional one being studied. The clarification of values and understanding does not come from a general exposure to another language and culture, but from targeted reflection on language(s) and culture(s) as they emerge from meaningful communication. For example, Liddicoat (2006), in discussing Australian learners' understandings of the interpersonal as represented through the French pronouns *tu* and *vous* over time, used a series of instances of language use that depicted an increasingly nuanced use of these pronouns in communication as a way of both promoting and eliciting an intercultural response to this element of French. This study showed that learners construct representations of greater complexity about the nature of the interpersonal relationship as they reflect on instances of authentic language use. In some cases, learners began their acquisition of French with a view of the *tu/vous* distinction as anomalous or problematic as constructions of interpersonal relationships, but increasingly came to see their own construction of the interpersonal as a cultural phenomenon. This

enabled them to see the cultural constructedness of the interpersonal in French as having value and an internal logic, even though it is different from their own construction.

Eliciting the Intercultural

In line with Sfard's (1998) notion of holding both the acquisition and participation metaphors in play, with regard to different understandings of the intercultural, it is necessary to assess not only subject matter knowledge, but also students' personal engagement with that knowledge. What becomes of interest, then, is students' ability to negotiate language use as participants within diverse contexts (participation metaphor). This process cannot be managed successfully without students knowing what is the same and different in different languages and cultures (acquisition metaphor). Elicitation, then, involves not just seeking to draw out students' understanding of intercultural experiences as observers of others' participation in interaction, but most importantly, drawing out students' understandings of themselves as participants in learning and using language(s) in the context of culture(s), that is, in learning and using languages as actual intercultural experiences.

Students embody their own particular linguistic, social, cultural and learning history, which informs the way they interpret the world. Elicitation of students' intercultural capability entails seeking to tap into students' frameworks of interpretive resources, encouraging them to explore their own understandings of the way they interpret the world through language and culture and how that, in turn, influences how they understand others. It entails assessing students' language use in the target language as well as their own understanding of what is at stake in communicating interculturally, that is, how what they say is received by others and how this influences the way others see them, and that this is always variable. For example, in students' use of particular forms of greeting, it is not only recognising and knowing the different forms and their use in so-called 'formal' or 'informal' contexts, but knowing how to decide and how their decision is likely to influence other participants in the interaction. Thus, in assessing the intercultural in languages learning and using, what is to be elicited is students' capacity to integrate, within interaction, an understanding of themselves as culturally and linguistically situated and to consider the impact of their situatedness and that of others on their interaction with others. Students, therefore, need to perform their language learning and use in culturally appropriate ways, take account of how fellow participants respond (what others 'make of

me' in interaction and conversely) and reflect upon how participants understand each other on the basis of the language used, in diverse interactions, in diverse contexts. In this way, they are invited to use language appropriately in relation to their understanding of the cultural/intercultural context, which in turn adds the potential for engaging with variability.

A central dimension of the elicitation of students' performance is how teachers *position* students as performers in interaction and the nature of the range of opportunities the students are given to perform and further develop their understanding in an ongoing way. Each new opportunity and experience of interaction depends on students' prior conceptions and expectations in relation to their own and the additional language and culture. These experiences are cumulative, each one adding to the students' repertoire of participation. It is for this reason that elicitation cannot be episodic, but rather it needs to be ongoing, tapping into the ever-developing diversity of scope and complexity of intercultural understanding.

In both single episodes and cumulatively, intercultural communication involves understanding how to engage with the *variable context of communication* to interpret and construct meaning, in interaction with others across diverse cultures. From an assessment perspective it is important to elicit communicative performance and it is equally important to elicit how students understand the distinctive social and cultural context of that communicative performance, that is, eliciting how they explain the intercultural to themselves. It is not possible to evidence the intercultural without an explanatory/elaborating dimension. This is because an instance of language use for intercultural communication does not reveal students' capacity to draw connections between the decisions they make in using language to make meaning and the social and cultural context of use. In eliciting linguistically and culturally appropriate language use, and *intercultural* understanding specifically, it is necessary to elicit students' interpretation of and participation in the social and cultural context.

Case Studies in Eliciting the Intercultural in Language Teaching and Learning

Case study 1: Eliciting intercultural awareness

The case study[1] established to explore the process of eliciting the 'intercultural' in language teaching and learning involved five experienced teachers (two teachers of French, two of Indonesian and one of

German) and a small group of their Year 10 students. (Year 10 in Australian education marks the end of the junior secondary and the beginning of the senior secondary cycle of schooling. Students are approximately 15 years of age and, in the main, will have been learning the target language for three years.) The focus was specifically on aspects of meta-awareness of language and culture as a central dimension of evidencing the intercultural, rather than eliciting this awareness in the context of intercultural interaction. Two elicitation devices were developed. The first focused on the teacher and involved the video-recording of a teaching sequence where the 'intercultural' was considered by the teacher to be a focus and where assessment took place informally through classroom interaction. Three weeks after the initial video-recording, each teacher participated in an in-depth retrospective discussion and interview that was audio-recorded and subsequently transcribed. Each teacher was asked to select episodes to discuss with the researchers, and, after focussing on these episodes, the discussion extended to a general exchange about the ways teachers mediate intercultural language learning in classroom interaction. The second device focused on the students and was developed collaboratively by a member of the research team and each of the teachers. The whole class of students was asked to participate in a writing task, which involved (1) responding to questions based on a text in the target language and (2) writing a short composition in English. The writing task was followed by a student interview with a group of five students from each class, selected on the basis of their responses to the first task. The interviews were conducted by the researchers; they invited students to reflect upon and discuss their responses to the writing tasks. Both sets of devices were designed to elicit teachers' and students' understandings of the intercultural in language teaching and learning.

The multiple sources of data were analysed thematically with a view to documenting student and teacher conceptions of the 'intercultural' (i.e. the construct of interest) and to examining procedures for eliciting intercultural language learning.

Teachers' constructions of the intercultural

To understand the ways in which each teacher constructed the notion of the intercultural as a prelude to understanding how they elicited the construct, each teacher was asked to respond to the following three questions:

- What were you trying to do in this particular lesson with respect to the integration of language and culture?

- What did you want your students to learn from this lesson?
- How does it relate to the rest of your programme?

The general nature of the questions represents a deliberate attempt to allow scope for teachers to elaborate. The first question was designed to elicit teachers' understanding of the integration of language and culture, understood by the researchers as a way of articulating the intercultural. The second question was intended to elicit the goals/purposes for learning and using an additional language and the third question was intended to elicit the ongoing nature of developing intercultural understanding through the learning programme.

Follow-up questions included questions about the teachers' own values and beliefs, conceptions of language, culture and intercultural learning, the personal education of young people and ethics in interaction.

In general, although all the teachers in the case study explored cultural themes in their teaching and learning programmes, these were not always connected deliberately to linguistic and cultural concepts, or to the lives of students themselves, and they were not elaborated continuously over time. Commenting on a cooking experience he had created for his students, for example, the Indonesian teacher states 'I don't think that they were using expressions as well. They would have been talking about the ingredients but they wouldn't have been actually using Indonesian', thereby indicating a disconnection between language and culture. Similarly, the French teacher had introduced a video clip of French rap, the purpose of which, he explains, was 'to look at it from a cultural perspective as distinct from a linguistic one'. The separation between language and culture is partly a question of teacher conceptions, and/or of desired emphases; it can also be the result of the disjunct between students' level of conceptual understanding and the limits of their level of development in the target language. For example, the Indonesian teacher observes that an intercultural discussion can only be managed in English, that is, the students' first language rather than the target language. He states:

> Because you are working at such a level, it is very difficult to get the students, unless it is something probably English-based working at that higher order thinking skills, like analyse. In a (foreign) language I find that quite difficult.

In response to questions about long-term planning for intercultural language learning creating an opportunity to assess understanding over

time, neither teacher treated the intercultural in a sustained way. The French teacher indicates that the cultural focus on the video clip is episodic, but makes a highly generalised connection with the programme as a whole at a highly generalised level. He states:

> On the one hand it (the lesson incorporating French video clip) doesn't relate at all in so far as it was an isolated, one-off activity; it certainly had nothing in particular to do with the unit of work that we were involved with. But, on the other hand, it is related to the course that they are doing in so far as that course also avoids stereotypes... in one way showing the culture as being ordinary as opposed to being exotic.

In response to a similar question, the Indonesian teacher recognises that cultural learning is incidental rather than planned within a developmental perspective.

Reflecting on outcomes of learning, the French teacher recognises the intricacy of determining what it is, in fact, that students learn. He observes:

> With that song (of the video clip) the students were greatly taken by it. But what do they get out of it? Do they go away with just the music and the feel? Or do they go away with the images. And if they go away with the images, is it the multicultural, the tolerant image or is it the bald heads and the quirky behaviour, the ''cool'' side of it?

The teachers seemed to be aware of the lack of connection or integration and related this to issues of time and choices about what is central to language-and-culture-teaching. Some expressed the view that they see themselves as teachers of language and not teachers of history, geography, science and so on, a comment that reflects their conception of teaching language and culture. They all perceived it as useful to see several 'layers of meaning' in teaching languages, going from surface to deeper treatment of language and culture in communication, that is, developing meta-awareness, but they found it difficult to connect this conceptually with students' personal experiences and lives and concepts from other areas of the curriculum. Additional themes that emerged from the analysis of the transcripts include the implicit versus explicit integration of language and culture learning; the perception that 'culture' contributes substance to language learning, but also takes away the focus from language; and the treatment of intercultural understanding as input rather than as integral to interaction and intercultural learning.

In eliciting teacher conceptions, there is always and inevitably the effect of the presence of the video-recorder and the interviewer because performance is necessarily a social, cultural and interactive accomplishment. Recognising that no elicitation procedure can ever be neutral, the nature of the elicitation questions themselves in this context is also influential.

Students' responses

Students undertook two pieces of writing, one related to the latest theme and texts they had explored in the languages classroom and another that was more general. In the Indonesian class, for example, students had explored the theme of Indonesian food and cooking, in the French class the language of rap music, and in the German class the theme of fairytales. The students were asked to address explicit questions relating to these topics. For example, for Indonesian these questions were:

> Part A: 'Indonesia has a wider variety of food venues with more attractive food available to them than is available to Australian young people'. In at least one paragraph, state whether or not you agree with this statement and give your reasons why.
> Part B: (i) List three words in Indonesian that relate to food/cooking. Explain what each word means. Describe how Indonesians use these words. Why are they important words in Indonesian culture?
>   (ii) Think about everything that you know about Indonesian food and culture. What has interested you most about it? Describe this.
>   (iii) What has learning about Indonesian food and culture made you think about your own way of life in Australia?

These questions, although framed entirely in English for the purposes of both elicitation and response, were intended to elicit students' overall 'take-home' understandings of their exploration of the theme. In part B, there was a focus on particular words in Indonesian. The intention here was both to elicit specific understanding of word-concepts and to include the deliberate elicitation of aspects of the target language. Information was elicited about the way in which words are used in the target language and culture and why the word-concept is important, but students were not asked to use the words actively themselves in interaction to communicate meaning. The final question (B. iii) invited a connection with students' way of life in Australia. The focus in this set of questions as a whole was on eliciting knowledge derived from engagement in a thematic learning sequence. As devices for the

elicitation of learning, these questions invite students to demonstrate understanding from the perspective of analysts of language and culture rather than as participants. The focus was specifically on aspects of meta-awareness and overall reflection.

The second, more general piece of writing was designed to invite students to compare and reflect upon the target language and culture and their own. The question used for Indonesian was:

> Imagine this is your final year of studying Indonesian. Your class has decided to make a collection of impressions of Indonesia for a class album. You have been asked to write a summary describing your impressions of Indonesian culture: what is it, what makes it the same or different from your culture, why do you think it is the way it is? Include examples of Indonesian language and culture, your opinions and comparisons to justify or explain your points.

Again, these tasks were intended to elicit generalised understandings of language and culture, positioning students deliberately in the role of analysts of language and culture, rather than as participant users of the target language in interaction. Both tasks specifically required of students that they de-centre from their primary linguistic and cultural context to consider the context of users of the target language. Depending on the purpose of assessment, this style of questioning can be seen as contributory (at least to some extent) to understanding students' understanding or meta-awareness of language and culture. The purpose here is to elicit and assess 'knowing that and why' rather than eliciting socio-culturally appropriate target language use on the part of students in intercultural interactions.

In the student interviews, groups of five students were asked about their responses to the writing tasks through two key questions:

- What was your initial reaction to the task? Is it similar/different to others you usually do?
- What kinds of opportunities do you have in your course to learn about culture?

In addition, students were asked to elaborate on some of the generalisations that emerged from their own writing in relation to four broad areas: people, place/society, language, comparisons. For example some of the questions for French included:

> **People:** 'They are expressive people'. 'They are confident people'. What gives you this impression? Is it the case for all?

> **Place/society:** 'Fashions in France are important to know about – so as not to judge by Australian standards'. What does that mean?
> **Language:** 'The language illustrates a ''laissez faire'' approach to life'. What do you mean by this?
> **Comparison:** 'French is more specific than English'. What do you mean? What does this say about Australian culture?

As with the teachers, eliciting student conceptions is shaped by the presence of the interviewer and by the nature of the elicitation procedures or devices used. This shaping process needs to be taken into account when making inferences from students' responses.

Students' responses indicated that they were not used to the style of questioning used here for elicitation. While they were asked to respond in English rather than the target language, thereby being positioned as analysts rather than users of the target language and culture, their responses revealed valuable evidence to the teachers (and to themselves) as well as to the researchers, about the way students construct language and culture, their meta-awareness, their conceptions and misconceptions, and most importantly, the cumulative picture they develop of their own and other languages and cultures. Some of the comments from the German students included:

> Language and culture are not really connected. I've never really thought of language being part of the culture. They're separate things. Culture is pyramids . . . and that doesn't have anything to do with what people say to each other. Language is what people use to communicate.
> We (in Australia) don't have a culture because we don't really come from one group of people.

Some reflections on Indonesian culture include:

> the Australian culture has less emphasis on politeness and respect for elders than in Indonesia. Also that religion plays less of a part in our lives.
> We had to be considerate about their beliefs and rules that we were not used to in Australia.

While highly generalised and often stereotypical, these comments provide a window into students' understanding of language, culture and the intercultural. For the teacher, the students' responses also reflect back on their teaching and the messages that their students form and take away from the process.

What was elicited in this case study was not students' capability to negotiate their use of the target language in interaction across diverse cultural contexts, that is, intercultural communicative interaction, but rather, the sense that the students themselves were making of the key concepts in their learning and the development of their awareness of their own culture and the target culture(s). Students can only participate in intercultural interaction if they have a sense of the cultural context, particularly, what is the same and different from their own. Beyond language use, eliciting the intercultural requires eliciting students' interpretations of the cultural context and what is going on in interaction, all of which are part of their developing meta-awareness not only of the role of language and culture in interaction, but also the nature of intercultural interaction itself. This meta-awareness is an integral part of the communication process in general. For students, the processes of analysing, explaining and elaborating on their meta-awareness are integral to evidencing the intercultural.

The case study focused specifically on the analysis dimension of the role of language and culture in communication, separately from the process of intercultural interaction. It would be both feasible and desirable to integrate both aspects in a single elicitation procedure, but the decision about the focus of elicitation depends on the teachers'/ assessors' purposes, which at any particular moment may or may not be integrative.

Case study 2: Eliciting intercultural behaviours

The case study for eliciting intercultural behaviours was conducted with 10 second year university level French students in the *ab initio* stream, studying in Australia (see Liddicoat & Crozet (2001) for a full discussion of the study). These students had completed one year of study of French, beginning at tertiary level. The students participated in role play tasks based around the question *'T'as passé un bon weekend?'* both prior to instruction and immediately after instruction. The focus was on cultural differences in the values associated with this question in French and Australian English and the resulting interactional differences in ways conversations were constructed in response to the question.

In cross-cultural studies of interaction (Béal, 1992), the key differences found for this question were that in Australia, the question has the function of an access ritual at the beginning of the working week, while in France, it is a non-ritual information-seeking question. The

interactional result of these different evaluations is that Australian interactions tend to involve short generalised, formulaic answers and the conversation usually moves to other topics or to closing quickly. In many interactions, individuals are referred to by their affiliation to the speaker (my partner, my sister, etc.) rather than by name. The interactions usually involve minimal feedback from the listener, as the social role involved is expressed through reciprocity rather than involvement in the topic. In French interactions, the question leads directly to talk on the topic, which is specific and detailed and involves the expression of opinions and feelings about the events that have transpired. Interactions tend to place emphasis on using the answer to the question as an opportunity to interest and entertain one's interlocutor. Moreover, the interlocutor is treated as knowing family relationships and friends, who are discussed by name rather than affiliation. French interactions are characterised by high levels of involvement from the listener, with frequent feedback being given, repetition of elements of talk and frequent overlap, especially in the form of collaborative completions, to show engagement. The situation is therefore contextually and interactionally different between the two cultures.

Prior to receiving instruction, Australian students demonstrated a strong adherence to Australian interactional expectations in constructing talk about the topic (see Table 4.1), with almost no dyads producing any

Table 4.1 Features of performance in role plays without instruction

| | | Dyads | | | | | | | | | |
| | | A | | B | | C | | D | | E | |
Feature		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Content	Leads to talk on topic	×	×	×	×	×	×	×	×	×	×
	Detail	×	×	×	×	×	×	×	×	×	×
	Opinions/feelings	×	×	×	×	×	×	×	×	×	×
	Lively	×	×	×	×	×	×	×	×	×	×
	Knowing	×	×	×	×	×	×	×	×	×	×
Form	Feedback	✓	×	×	×	✓	×	×	✓	×	×
	Repetition	×	×	×	×	×	×	×	×	×	×
	Overlap	×	×	×	×	×	×	×	×	?	×

Source: Liddicoat and Crozet (2001)
Note: ✓ = feature present; × = feature absent; ? = ambiguous result

of the salient features of interaction. What this task elicited, in terms of intercultural behaviours, was a lack of awareness or a lack of use of French culturally based interactional norms.

Following interculturally focused instruction that involved raising awareness of the interactional and conceptual difference relating to questions about the weekend, and opportunities to experiment with new norms and interactional expectations, the students' behaviours differed significantly (see Table 4.2). In this case, the task elicited an ability to reproduce the interactional expectations of another culture, although different dyads performed differently on this task.

As was the case with case study 1, case study 2 is again only a partial eliciting of the intercultural. This study, like many studies of students' production in interlanguage pragmatics, elicited culturally more appropriate interaction as the result of interculturally focused language learning. It did not gauge what motivated the use of the practices, although it did elicit from students their affective reactions to using these practices. In a language learning context, it is probable that the interactional patterns were motivated by a desire to perform in an appropriate way in a language learning task rather than from purely intercultural motivations. In fact, what happens in such tasks is that the interaction is not really considered as being intercultural, except at a very superficial level – that of the behaviours of people from one culture in response to learning about the behaviours in another. In such

Table 4.2 Features of performance in post-instruction role plays

Feature		A		B		C		D		E	
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Content	Leads to talk on topic	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Detail	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Opinions/feelings	✓	✓	?	×	✓	✓	✓	✓	×	✓
	Lively	✓	✓	✓	✓	?	✓	✓	✓	✓	?
	Knowing	?	×	✓	✓	✓	✓	?	✓	✓	✓
Form	Feedback	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Repetition	×	✓	✓	✓	×	×	?	✓	✓	✓
	Overlap	×	×	✓	✓	×	×	×	×	×	?

Note: ✓ = feature present; × = feature absent; ? = ambiguous result

studies, the non-native speaker is, in fact, being measured against native speakers' culturally embedded interactional norms, with the level of accuracy in using these norms being the criterion for judging the performance. Intercultural communication is not identical with an assimilation or approximation to others' norms; rather it is a principled engagement with diverse assumptions, expectations and understandings. For this reason, it is possible that authentically intercultural behaviour may be less native-speaker-like, with individuals negotiating their own identities between cultures and manifesting these by selecting elements from their whole linguistic and cultural repertoire (Kramsch, 1999; Liddicoat, 2002). With this in mind, in Liddicoat and Crozet's (2001) study it was the presence of particular features that was observed, not the accuracy with which they were reproduced, allowing individuals to mediate their behaviour between different cultural expectations and assumptions.

Concluding Comments: Eliciting and Assessing the Intercultural

While the case studies described above focused only on eliciting some aspects of language and cultural awareness as the basis for the analysis dimension of intercultural understanding, the elicitation of the intercultural entails a fuller set of features. Some of these include:

- that the elicitation procedure or task involves interaction in the target language with other users of that language, which requires that students de-centre from their own language and culture;
- that it involves eliciting students' meta-awareness of the language-culture nexus in such interactions and that for the purposes of assessment, students be able to analyse, explain and elaborate this awareness;
- that within the elicitation procedure the student be positioned as both participant and analyst in interaction, though in any individual elicitation task one or the other role may be foregrounded, depending on purpose;
- that the procedures elicit students' understanding of different perspectives in the interaction, as an ethical concern;
- that a combination of a range of elicitation procedures be used to capture (1) the diverse and multidimensional nature of intercultural interaction and related meta-awareness, and (2) the recognition of the need to manage the variability of context, given that each

episode of intercultural interaction presents a new set of contextual considerations to be negotiated in communication;

- that elicitation procedures be considered within a long-term perspective that allows development and progress to be taken into account, for example, through the use of portfolios;
- that the elicitation procedures include self-assessment that recognises learning as a personal process and a personally unique accomplishment.

Fundamentally, the elicitation procedures are no more than a device, an artefact for capturing students' ever developing interpretive framework that is brought to bear in participating in and reflecting on their own role in intercultural interaction, simultaneously as an interactant, a performer *and* an analyst. This poses challenges in that the procedure can only be a proxy for eliciting the knowledge and understanding of interest, a complex construct that is understood in diverse ways. To capture the complexity, elicitation must include opportunities for students to interact interculturally and to analyse, appraise, reconsider and learn from the interaction in an ongoing way, as is characteristic of human communication in general and in a more intense way in communication across diverse cultures. Further challenges emerge in seeking to identify and judge the evidence of the development of this dynamic understanding and warranting the inferences made about students' developing understanding (see Figure 4.1).

It is not likely that any single task, whether as acts of communicating or as the articulation of dispositions and insights, will in itself fully elicit the intercultural as understood and experienced by any one learner. Assessment of the intercultural, at least in languages education, needs to be based on demonstrations of both intercultural communication and the insights and dispositions towards diversity, with an orientation to assessing the individual's learning over time. Such an approach to assessment cannot be captured in a single episode, which cannot capture either the longitudinal nature of learning or the dynamic and multi-layered nature of the intercultural (Liddicoat, 2002). What is needed is a combination of communicative and reflective tasks in the form of a portfolio collected over time.

The notion of a portfolio of tasks for assessing intercultural competence is not new, but the ways in which portfolios are understood in language assessment varies greatly. At its simplest and weakest formulation, a portfolio is no more than a collection of individual episodes, which may or may not have internal coherence and connectivity. While such a

portfolio provides multiple exemplars of various aspects of the intercultural, it presents only a weak eliciting of the developmental nature of the intercultural, whether from the perspective of the learner or that of the assessor. Each of the episodes that the portfolio represents is a means of eliciting, but the portfolio itself does not elicit; it evidences. A stronger version of the portfolio would add to the collection of tasks commentaries on the tasks themselves in which the learner engaged with the task as an instance of intercultural experience. However, while such an approach captures a stronger perspective on the learners' own understandings of the learning experience, it is still weak from a developmental perspective. What is needed still is a set of procedures, designed or selected to capture diverse instantiations of the intercultural over time so that these are less restricted to particular (partial) constructions of the intercultural.

Portfolios can be further strengthened as a resource for eliciting the intercultural by including the learners' reflections on the process of learning to be intercultural itself. Such a reflection is an on-going self-evaluation of the individuals' perceptions of their own progress towards an intercultural stance through the process of language learning. This is more than a language biography; it is a guided reflection on the specific learnings that have emerged over a course of study. The aim of the reflection is to encourage the development of insight into oneself as an intercultural actor. Ideally, this should be accompanied by a culminating commentary in which the individual learner reflects on his/her own journey towards interculturality as an on-going and unfinished project. Such portfolios, we argue, allow longer-term development to emerge as the key focus of assessment and this development is where the intercultural can be seen and judged. As the case studies above show, no experience, communicative engagement or reflection on culture or on learning is itself 'interculturality'. Each episode is no more than that – an episode – and interculturality, as we have described it, is not an episodic phenomenon. The portfolio has to be designed around tasks that involve features such as those discussed above in order for the portfolio to present the most complete and most coherent account of learners as intercultural beings.

Note

References

Béal, C. (1992) Did you have a good weekend: Or why there is no such thing as a simple question in cross-cultural encounters. *Australian Review of Applied Linguistics* 15 (1), 23–52.

Brislin, R., Cushner, K., Cherrie, C. and Yong, M. (1986) *Intercultural Interactions: A Practical Guide*. New York: Sage.

Byram, M. (1997) *Teaching and Assessing Intercultural Communicative Competence*. Clevedon: Multilingual Matters.

Byram, M., Gribkova, B. and Starkey, H. (2002) *Developing the Intercultural Dimension in Language Teaching: A Practical Introduction for Teachers*. Strasbourg: Council of Europe.

Byram, M. and Morgan, C. (1994) *Teaching and Learning Language and Culture*. Clevedon: Multilingual Matters.

Byram, M. and Zarate, G. (1994) *Définitions, objectifs et évaluation de la compétence socio-culturelle* [*Definitions, Objectives and Evaluation of Socio-cultural Competence*]. Strasbourg: Report for the Council of Europe.

Cadd, M. (1994) An attempt to reduce ethnocentrism in the foreign language classroom. *Foreign Language Annals* 27 (2), 143–160.

Crichton, J.A. and Scarino, A. (2007) How are we to understand the 'intercultural dimension'? An examination of the intercultural dimension of internationalisation in the context of higher education in Australia. *Australian Review of Applied Linguistics* 30 (1), 4.

Damen, L. (1987) *Culture Learning: The Fifth Dimension in the Language Classroom*. Reading, MA: Addison-Wesley.

Jæger, K. (1999) Intercultural competence and the problem of assessment. In T. Vestergaard (ed.) *Language, Culture and Identity* (pp. 71–81). Aalborg: Aalborg Universitetsforlag.

Kramsch, C. (1999) The privilege of the intercultural speaker. In M. Byram and M. Fleming (eds) *Language Learning in Intercultural Perspective: Approaches through Drama and Ethnography* (pp. 16–31). Cambridge: Cambridge University Press.

Liddicoat, A.J. (2002) Static and dynamic views of culture and intercultural language acquisition. *Babel* 36 (3), 4–11, 37.

Liddicoat, A.J. (2005) Teaching languages for intercultural communication. In D. Cunningham and A. Hatoss (eds) *An International Perspective on Language Policies, Practices and Proficiencies* (pp. 201–214). Belgrave: FIPLV.

Liddicoat, A.J. (2006) Learning the culture of interpersonal relationships: Students' understandings of person reference in French. *Intercultural Pragmatics* 6 (1), 55–80.

Liddicoat, A.J. and Crozet, C. (2001) Acquiring French interactional norms through instruction. In K.R. Rose and G. Kasper (eds) *Pragmatic Development in Instructional Contexts* (pp. 125–144). Cambridge: Cambridge University Press.

Liddicoat, A.J., Papademetre, L., Scarino, A. and Kohler, M. (2003) *Report on Intercultural Language Learning*. Canberra: DEST.

Mabry, L. (1999) *Portfolios Plus: A Critical Guide to Alternative Assessment*. Thousand Oaks, CA: Corwin Press.

Messick, S. (1989) Validity. In R.L. Linn (ed.) *Educational Measurement* (3rd edn) (pp. 13–104). New York: Macmillan.

Scarino, A. (2006) An introspective and retrospective study of teacher judgment of student performance. Unpublished manuscript.

Sercu, L. (2004) Assessing intercultural competence: A framework for systematic test development in foreign language education and beyond. *Intercultural Education* 15 (1), 73–89.

Sfard, A. (1998) On two metaphors for learning and the dangers of choosing just one. *Educational Researcher* 27, 4–13.

Part 2

Autonomy

Chapter 5

Measuring Autonomy: Should We Put Our Ability to the Test?

PHIL BENSON

Introduction

When the editors of this volume first asked me to contribute a chapter to this volume, my first reaction was to ask why on earth would anyone want to test a foreign language student's autonomy? Surely, our aim is to foster autonomy, not for its own sake, but to help students become better language learners? If we have to test anything at all, let us test students' language learning, not their autonomy. But the editors also pointed out that, although very little had been published on the topic to date, I had already written a short piece on measuring autonomy (Benson, 2001: 51–54). This was true, but what I had written was mainly concerned with showing how problematic the idea of measuring autonomy was. More importantly, my comments were speculative, rather than empirically grounded, and I wondered what I might have to add to them at this point. Indeed, in a more recent review of the literature, I found that I had nothing to say about measuring autonomy at all (Benson, 2007). This was, however, only my first reaction, and on deeper reflection I felt that there was perhaps more to be said in two respects.

Nunan's (1997: 92) observation that autonomy is not an 'all-or-nothing concept' but a matter of degree has been widely cited in the literature. We also frequently read of students becoming 'more autonomous' as a consequence of their participation in a particular programme or activity. In other words, we seem to have a strong intuitive sense that students may be 'more' or 'less' autonomous. But what exactly is this sense based on? Linking the construct of autonomy to observable dimensions of control over learning, in Benson (2001: 51) I suggested that, 'if we are able to define autonomy and describe it in terms of various aspects of control over learning, we should also in principle be able to measure the extent to which learners are autonomous'. This was intended to be no more than a hypothesis. But although studies concerned with the measurement of

autonomy remain few, a number of instruments designed for the purpose of measuring aspects of autonomy have subsequently been reported in the literature. There has also been at least one report of a credit-bearing university language course in which student autonomy is assessed (Ravindram, 2000). Through a review of this work, I hope to move one or two steps forward with the hypothesis that we *can* measure autonomy.

A second issue that appears worthy of more attention concerns the reasons why, in spite of our ability to design instruments for the measurement of autonomy, we may not actually *want to* use them to assess student learning. In Benson (2001), I was primarily interested in what a better understanding of autonomy as a measurable construct might contribute to our ideas about how autonomy develops over time and how it interacts with different contexts of teaching and learning. But in a later paper, I explored a nagging feeling that this was not perhaps the 'right' way to think about autonomy (Benson, 2003). Developing this feeling in this chapter, I want to explore the risks that might be involved in delving too deeply into questions of the measurement of autonomy in educational climates in which the 'unmeasurable' often seems to lack value.

Autonomy and Control

Autonomy is a complex construct in the sense that it is, like the more widely measured construct of foreign language proficiency, made up of many things, none of which are quite the same as autonomy itself. When we judge intuitively that students are either more or less autonomous, what we appear to be doing is observing certain behaviours and associating them with the broader construct of autonomy. For example, we may observe that a student is making a study plan, read the plan and judge that it is a good one. From this, we may infer that the student is to some degree autonomous. But planning is by no means identical to autonomy. At most, it is a component of autonomous learning, and possibly not a necessary component (for a discussion of this point, see Benson, 2001: 77–78). An initial problem in measuring autonomy, therefore, is determining what autonomy's necessary or most important observable components are. Autonomy may, of course, have important non-observable components as well and a second problem is to determine whether these non-observable components are so important that we cannot really measure autonomy at all.

Educational research has borrowed the idea of autonomy from the field of political philosophy, where it primarily refers to the self-determination of the affairs of individuals, groups and polities. Individuals are

personally autonomous to the extent that they 'author' (Young, 1986), or 'chart the course of' (Wall, 2003), their own lives within social and environmental constraints. There is a sense in which personal autonomy may be a desired outcome of education generally and foreign language learning specifically (Littlewood, 1996). But this is not the sense of autonomy that we usually have in mind when we talk about autonomy in foreign language learning, which refers more to a certain kind of relationship between the student and the learning process. The term that I have found most useful in describing this relationship is 'control', largely because it has wide resonances in the educational literature. Autonomous language learners are, therefore, learners who are in some sense 'in control' of important dimensions of their learning, which might otherwise be controlled by others or by nobody at all.

I also find the notion of control useful because it seems to be present within a range of observable behaviours that we readily associate with autonomous learning. A student who creates a workable and appropriate study plan, for example, begins to exercise control over factors such as the location, timing, pace, sequencing and content of learning. These factors could be directly or indirectly controlled by others – a teacher, an institution, a textbook, an examination syllabus etc. – or they may not be controlled at all. The same could also be said of quite different aspects of the foreign language learning process. When reading or listening to a text, students may, through control of attentional processes, choose the elements of linguistic input that they will pay most attention to, or these may be selected by others or not at all. Students may also exercise control over the kinds of learning activities they participate in and the extent of their participation, and once again this may be controlled by others or not controlled at all. As Little argues:

> It is true, of course, that we recognize autonomous learners by their behaviour: but that can take numerous different forms, depending on their age, how far they have progressed with their learning, what they perceive their immediate learning needs to be, and so on. Autonomy, in other words, can manifest itself in very different ways. (Little, 1991: 4)

But how exactly do we recognise autonomous learners by behaviours that are essentially diverse? We do so, I would argue, by recognising them as behaviours in which control over one or more dimensions of the learning process is manifest.

As a possible framework for the measurement of a student's autonomy, therefore, I want to suggest that there are essentially three

poles of attraction in regard to control over learning: student control, other control or no control. In regard to any particular dimension of control – for example, study planning – these three poles of attraction could be represented as the points of a triangle, with the distribution of control at a particular moment in time being represented by a cross somewhere within the triangle. In Figure 5.1(a), the position of the cross indicates that the student is largely in control of study planning. Similar triangles could be drawn for other dimensions of control. Figure 5.1(b), for example, indicates a lack of control over attention to input on the part of both the student and others. My assumption here is that, for any given dimension of learning at any particular moment, there will always be some degree of control, which will usually be shared between the student and others. Imagining the various dimensions of the learning process as a series of overlapping triangles of this kind, the student's overall degree of autonomy could be represented by the configuration of the points within the triangles. Figure 5.2 hypothetically illustrates this model for five dimensions of learning, although in practice there are likely to be more than five dimensions at issue. This may not give us a workable measurement scale, but it does offer a way of visualising the sense in which students may be either more or less autonomous in different ways.

   One of the strengths of this framework, in my view, is that it helps us chase away one or two red herrings. First, autonomous learning is not exactly the same thing as freedom from the influence of teachers, institutions, materials and so on, or learning by oneself. Whereas autonomous learning implies, by definition, that the student is in control of the learning process to some degree, learning by oneself does not carry
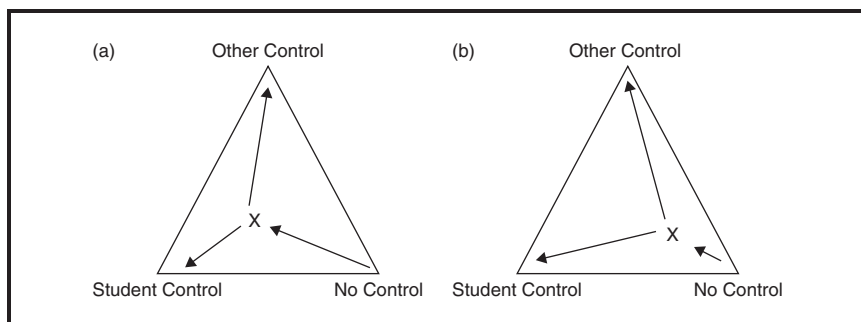

Figure 5.1 (a) Control over study planning. (b) Control over attention to input
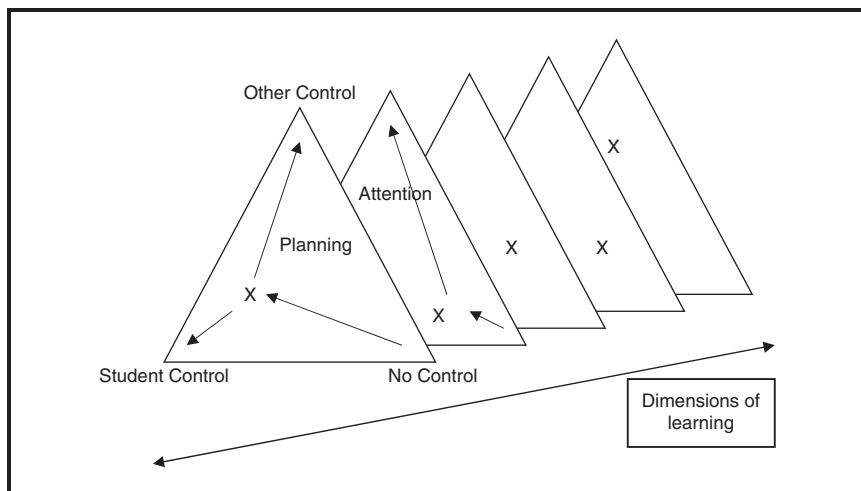
Figure 5.2 Control over language learning

the same implication. It is possible, for example, that the learner will be 'drifting' or 'lost', because nobody is really in control of the learning process. A helpful analogy here, perhaps, is a situation in which a person is driving, but not fully in control of the vehicle. Similarly, even under circumstances where learning is primarily other-controlled, notably in classrooms, students are likely to exercise some control over their learning. The overall control exercised by the institution and teacher may even help the student exercise this control. If other control is relaxed, the consequence may be that the students' learning becomes uncontrolled. We hope, of course, that the students will step in and take control for themselves, but we also know that there is no guarantee that this will happen. If we view autonomy as a measurable construct, in other words, what we need to measure is not the degree to which students are independent of the influence of other-controlling agents, but the degree to which they are actually in control of their learning.

Second, it seems that autonomy is often confused with constructs that have their own integrity. In the field of self-directed learning, there have been a number of attempts to produce taxonomies and scales based on the components of learner autonomy (Brockett & Hiemstra, 1991; Candy, 1991; Confessore & Park, 2004; Guglielmino, 1977; Knowles, 1975). One problem that has been identified with such taxonomies and scales is that they often cover too much, such that high scores on Guglielmino's

Self-directed Learning Readiness Scale, for example, 'seem to represent a positive attitude toward learning in general and not specifically toward the kind of learning called self-directed' (Bonham, 1991: 92). Similarly, Candy (1991: 459–466) lists more than 100 competencies associated with autonomy in the educational literature, but in many cases – being able to organise data, having a taste for learning, being amiable and peace-loving, being emotionally stable, objective and impartial, and so on – it is difficult to see the sense in which they belong to autonomy as opposed to other constructs. Within the field of foreign language education, we also have evidence of relationships between autonomy and strategy use (Wenden, 1991), certain kinds of learner beliefs (Cotterall, 1995), metacognitive knowledge (Wenden, 1998) and motivation (Ushioda, 1996). But it seems important that we keep these constructs distinct. We cannot, for example, reasonably measure a student's autonomy using an instrument designed to measure learning strategy use; or at least, if we were to do so, we would imply that there is no distinction to be made between autonomy and strategy use. It is through the idea of control, therefore, that we are apparently most able to assert the integrity of autonomy as a usable construct within language education research.

Measuring Autonomy: Some Problems

Having suggested a possible framework for the measurement of autonomy based on observable behaviours in which control over learning processes is manifest, I now want to turn to four major problems with the implementation of such a framework.

The multidimensionality of autonomy

Autonomy is a multidimensional construct in the sense that students control their learning in a variety of ways. Elsewhere, I have discussed these kinds of control under three headings: the management of day-to-day learning, the cognitive processes involved in foreign language learning, and the content of learning (Benson, 2001: 50). The three examples that I used in the previous section – study planning, control over attentional processes, and control over learning activities and participation – were intended to illustrate these three levels of control. It would be unreasonable, of course, to suggest that learners need to control every conceivable aspect of their learning in order to count as autonomous learners. The problem we have, however, is that autonomy does not seem to be reducible to any particular combination of controlling behaviours. The three instances that I have given, for

example, do not necessarily go together, nor do we have any strong grounds to say that one is more fundamental to autonomy than the others. Indeed, behaviours in which control over learning is manifest often seem to be non-comparable, such that it is difficult to say that one student is 'more autonomous' than another. In this sense, the recognition that autonomy can take different forms would appear to preclude an international 'test of autonomy', but not necessarily context-specific tests designed to assess individuals' gains in autonomy over time.

Autonomy as a capacity

Perhaps the most significant problem with the framework I have outlined is that autonomy is not, in fact, generally considered to be a matter of observable behaviour. As Holec (1981: 3) puts it, autonomy describes 'a potential capacity to act in a given situation – in our case – learning, and not the actual behaviour of an individual in that situation'. Sinclair (1999: 95–96) illustrates this idea with an anecdote about a business management English student working in a self-access centre. The student comes across the phrase *power distance*. Knowing the meaning of the words *power* and *distance*, but not the collocation, he gets up from his seat, walks over to the tutor on duty and asks her what it means. Sinclair's question is: How do we know whether this student is demonstrating autonomy or not? An irritated tutor, she suggests, 'might feel that he is taking the lazy teacher-dependent way out', but does the tutor know what he has done to find out the meaning before he approached her? Or has he perhaps considered the alternatives and concluded that asking the tutor will be the best course of action? To the extent that he has made an informed decision, Sinclair argues, it could be that the student 'has been using his capacity for autonomy but the tutor cannot see this process, only the outcome'. This anecdote appears to reinforce the point that we should be concerned with the sense in which students are in control of their learning, and not their dependence on, or independence from, teachers. But it also suggests that inferences drawn from observable behaviours may be an unreliable guide to underlying behavioural intentions (Confessore & Park, 2004). We need, in other words, somehow to get at the meaning of behaviours, and as we will see later, Sinclair does have a possible solution to this problem.

Autonomy as a developmental process

One step forward, two steps backward. Our third problem concerns what we know, or perhaps do not know, about the nature of autonomy as

a developmental process. We know, for example, that autonomy tends to be 'domain-specific' and that it can be 'lost' as well as 'gained'. As Little argues,

> The fact is that autonomy is likely to be hard-won and its permanence cannot be guaranteed; and the learner who displays a high degree of autonomy in one area may be non-autonomous in another. (Little, 1991: 5)

Little also argues that the development of autonomy in institutional contexts is rarely a smooth process:

> Indeed, it is a common experience that attempts to make learners conscious of the demands of a learning task and the techniques with which they might approach it, lead in the first instance to disorientation and a sense that learning has become less rather than more purposeful and efficient. (Little, 1991: 21)

This is perhaps because, as Holec (1985) argues, the development of autonomy often involves processes of 'psychological deconditioning', in which entrenched habits of dependency acquired through institutionalised learning are challenged and overcome. Breen and Mann (1997: 143) also comment that in situations where learners have been socialised into relations of dependency, autonomy may initially be manifested in 'individualistic and non-co-operative or competitive ways of being'.

The 'mask of autonomous behaviour'

Breen and Mann use this metaphor to signal the possibility that students will learn how to display autonomy, without necessarily becoming more autonomous in a deeper sense:

> Learners will generally seek to please me as the teacher. If I ask them to manifest behaviours that they think I perceive as the exercise of autonomy, they will gradually discover what these behaviours are and will subsequently reveal them back to me. Put simply, learners will give up their autonomy to put on the mask of autonomous behaviour. (Breen & Mann, 1997: 141)

There is perhaps a risk of overstating the case here. A desire to 'please the teacher' does not necessarily signal a lack of autonomy, and it may lead to the later adoption of autonomous behaviours of the students' own volition. The problem here, however, is partly related to the distinction between autonomous behaviour and autonomy as a capacity. In general,

the adoption of a behaviour signals the possession of some underlying capacity. Yet, in the case of autonomy, it seems that the relationship between the two can be hidden, and even that autonomous behaviour can be simulated. If we are really to measure the degree to which students are autonomous, we will somehow have to capture both the meaning of behaviours and their authenticity as behaviours deriving from a capacity for autonomy. And presumably, students would be especially inclined to wear the 'mask of autonomous behaviour' in situations where they were actually tested on their autonomy.

To some degree, these four problems are the same kinds of problems that we confront in measuring any complex construct. We might also make a tentative comparison between problems in the measurement of autonomy and foreign language proficiency. Foreign language proficiency, for example, is a multidimensional construct and the difficulties of capturing all of its dimensions in tests are notorious. Foreign language proficiency can also be conceptualised in terms of underlying capacities that are problematically related to test performances. But the third and fourth problems seem to be more specific to autonomy. Although the achievement of foreign language proficiency is not exactly an incremental process – error and uncertainty often being necessary precursors to progress – our assumption that it is measurable is, in part, based on an assumption that linguistic knowledge and skills accumulate in the learner. The same assumption cannot be made about autonomy. In fact, we know relatively little about how autonomy 'develops' or how it is transferred from one situation to another. Most damaging to the idea that we may be able to test autonomy, however, is the possibility that students may simulate autonomous behaviour for the purposes of a test. Clearly, autonomous behaviour is more easily simulated than foreign language proficiency. We should also acknowledge, however, the time and energy that has gone into the conceptualisation of foreign language proficiency and the design of measurement scales in the field of language testing. It may simply be the case that the problems that we foresee in the measurement of autonomy appear more acute because we have, to date, largely failed to address them.

What Can We Assess in Autonomous Learning Programmes?

Having outlined a potential framework for the measurement of autonomy and identified a number of problems with it, I now want to turn to a number of studies reporting practical attempts to measure

degrees of autonomy. These studies have very different purposes and cannot really be described as a coherent body of research addressing the question of how we might measure autonomy. What I am mainly interested in, however, are the ways in which the construct of autonomy is operationalised in terms of control over various dimensions of the learning process. I will, therefore, mainly be concerned with what exactly is being measured in each study, and how it is measured.

Autonomously controlled tasks

Rowsell and Libben's (1994: 672) study 'sought to determine whether high achievers and low achievers are distinguishable in terms of what they *do* in the course of independent learning'. The study was conducted in the context of a second language acquisition course at a Canadian university, in which the students were asked to study a foreign language of their choice over a six-month period. The study focused on 30 students, out of a total of 54, who elected to use self-instructional materials. Based on diaries that the students kept over the period of study, Rowsell and Libben (1994) used self-ratings of proficiency and the extent to which initial goals had been achieved to divide the group into 'high' and 'low' achievers. They then examined the diaries for reports of 'autonomously controlled tasks' (ACTs), or examples of 'self-determined volitional behaviour that the learner engages in to affect the course of his or her learning' (Rowsell & Libben, 1994: 672–673). These were coded as 'pedagogical' or 'functional' ACTs. Pedagogical ACTs, or 'actions that took control of the pedagogical activities associated with language learning' (Rowsell & Libben, 1994: 673), were divided into five operations: 'addition', 'deletion', 'transposition', 'repetition' and 'change'; addition referring to the insertion of a new task into the sequence prescribed by the materials, deletion to a decision not to carry out a task and so on. Functional ACTs involved attempts to achieve functional or communicative use of the language while learning in isolation and were of two types: creating some form of meaningful communicative interaction (often with a dummy partner, and in one case with a teddy bear) and creating a meaningful context or background (often an imaginary one, as the case of a student who placed her learning of German in the context of a proposed visit to tennis star Boris Becker). Rowsell and Libben (1994) found that, although the high and low achievers did not differ in their use of pedagogical ACTs, the high achievers reported far more functional ACTs.

Decision-making episodes

Simmons and Wheeler's (1995) study was conducted in the context of an attempt to implement the Process Syllabus (Breen, 1987). Although not explicitly oriented toward the development of autonomy, the Process Syllabus is strongly oriented toward student control of teaching and learning content (Benson, 2001: 164–165); and in this instance it was viewed by Simmons and Wheeler (1995: 15) as 'an opportunity to enable full learner participation in the decision-making processes associated with selection of content, agreement on procedures, choice of activities and tasks, direction of working and ongoing evaluation'. The participants were a multilingual group of migrants to Australia following an advanced-level intensive English course for students with professional qualifications. The study addressed several research questions, of which two are relevant to the measurement of autonomy: Were the students able to adopt syllabus design roles that are traditionally reserved for teachers? What types of decisions did the students make? The main source of data on these questions consisted of transcripts of weekly Monday morning 'action meetings', at which the students planned the timetable and activities for the coming week. These were analysed in terms of decision-making episodes, the number of words and turns taken by individual students and their roles as 'emitters' or 'targeters' of messages. Using this data, Simmons and Wheeler (1995: 55) found that 'collectively the learners were able to take responsibility for their own learning and to share in the decision-making process of the group as a whole'. They found that the action meetings tended to address group needs, while individual needs were addressed through active negotiation with teachers for independent learning time and small-group projects. They also observed that the more fluent students tended to dominate meetings at first, but that the less fluent students took on more decision-making roles as the course developed.

Self-directed learning behaviours

Rivers (2001) studied the self-directed learning behaviours of 11 adult students attending intensive courses in Georgian and Kazakh at a US university. In contrast to the previous two studies, Rivers investigated these behaviours on courses that did not intentionally involve self-directed learning. The study used student evaluation data gathered at regular intervals in the form of responses to an open-ended question inviting general comments on the course. Rivers analysed these data using a 'grounded theory' approach (Strauss & Corbin, 1990) and

identified evidence of self-direction in three areas: (1) self-assessment of conflicts between learner and teacher styles, learning style conflicts within the groups, and the students' own learning styles and strategy preferences; (2) learner autonomy in the form of demands for the modification of aspects of the courses, including methodology, teacher feedback, classroom environment, sequencing and activities; and (3) self-directed language learning behaviours, including prioritising classroom and homework assignments, selection of tasks and inclusion of an independent study day in the programme.

Micro and macro-level process control

Lai (2001: 35) reports two rating scales designed to 'facilitate objective measurement of how learners' capacity for autonomy in language learning has developed over a course term'. The first scale aims to measure 'process control at the task or micro level' and the second 'self-direction at the overall process or macro level'. In order to develop these scales, Lai asked students using self-access listening resources on an independent learning course to evaluate their learning at the end of each session, using printed forms with headings to guide reflection. The students conducted 15 listening sessions in all and Lai designed a rating scale to assess the 'Task Aims' and 'Self-Assessment' sections of the first three and last three evaluations for each student. Defining process control as 'the learner's ability to self-monitor and self-evaluate her learning tasks and/or learning strategies employed for each learning activity' (Lai, 2001: 35), Lai developed four statements on which raters assessed what the students had written under these two headings on a 5-point scale:

(a) The task aim(s) is/are realistically set for the type of programme chosen.
(b) The aim(s) are directly related to specific aspects of listening skills or strategies.
(c) The self-assessment directly addresses the set aim(s).
(d) The self-assessment specifically addresses the learner's learning process or performance. (Lai, 2001: 37)

Two raters assessed 99 journal entries and statistical analysis showed a high level of inter-rater reliability. Lai also asked the students to design a 'personal course for self-directed language learning' at the beginning and end of the course and developed 17 statements assessing these designs to be rated on a 7-point scale (Lai, 2001: 39). For this scale, three raters were

used and statistical analysis showed internal consistency among the test items and the inter-rater reliability was again high.

Metacognitive awareness

Sinclair (1999) reports ongoing research concerned with the measurement of autonomy, based on data from student-teacher consultation sessions in an independent learning programme at Temasek Polytechnic in Singapore. As we have noted, Sinclair (1999: 102) finds observable behaviour to be a poor indicator of autonomy and the method that she reports addresses this problem by assessing metacognitive awareness, or 'the ''capacity'' for making informed decisions about language learning'. As a means for developing assessment criteria in the context of an independent language learning programme, she poses the following questions:

Can the student:
• provide a rationale for his/her choice of learning activities and materials?
• describe the strategies he/she used?
• provide an evaluation of the strategies used?
• identify his/her strengths and weaknesses?
• describe his/her plans for learning?
• describe alternative strategies that he/she could have used? (Sinclair, 1999: 103)

These are translated into questions that can be put to students when teachers discuss their work with them, with the most important question being: 'What else could you have done/ could you do?'. Sinclair (1999: 105) also provides a provisional assessment scale with three levels: 'largely unaware', 'becoming aware' and 'largely aware'. Lastly, she suggests that the ways in which students talk about their learning may also point to different levels of metacognitive awareness: anecdotal evidence, introspection, metaphor, epiphanies, questions and metalanguage, for example, appear to be characteristic of the 'becoming aware' stage, while the 'largely aware' stage is characterised mainly by description of alternatives.

The five studies that I have reviewed here seem to be supportive of the potential framework for the measurement of autonomy that I proposed earlier, in that they all relate to control of some dimension of learning or another. Rowsell and Libben (1994), for example, succeeded in designing an ingenious method of identifying and categorising

controlling behaviours within a self-instructional context. Simmons and Wheeler (1995) devised an equally ingenious method of measuring control over decision-making and learning processes in a classroom context. Rivers (2001: 287) also focused on control, arguing that 'the accurate use of metacognitive, affective, and social strategies to control the language learning process and the learning environment is the hallmark of self-directed learning'. Lai's (2001) rating scales focus on planning and self-assessment as key dimensions of control over the learning process. Assuming that the session records accurately reflected what the students had done, the first scale also measured both behaviour and the thinking that underlies it, while the second scale measured only the students' thinking. Addressing the capacity for autonomous behaviour, rather than autonomous behaviour itself, Sinclair's (1999) proposal focuses on control in a particular sense. In contrast to Rowsell and Libben (1994) and Simmons and Wheeler (1995), for example, Sinclair is not concerned simply with the students' ability to make decisions, but the extent to which their decisions are 'informed' by an awareness of options. Although the sources of data differ, her procedures are akin to those adopted by Rivers (2001), in that they work backwards from observed behaviour to its underlying rationale. They are also akin to Lai's (2001) procedures in that they address students' ability to plan and assess their behaviours, although again the sources of data differ.

In looking at these studies as examples of the measurement of autonomy, we should take care to note that in each case the measurement instrument was essentially designed for research purposes. The procedures used by Rowsell and Libben (1994), Simmons and Wheeler (1995) and Rivers (2001) seem too labour intensive for practical use in pedagogical assessment. Sinclair's (1999) and Lai's (2001) instruments, on the other hand, were designed with pedagogical applications in mind. While Sinclair's was essentially a preliminary proposal, Lai did use her rating scales with groups of students in a naturalistic research context. Unfortunately, she did not report the results of these assessments. It would be interesting to know, for example, whether the scores improved from the beginning to the end of the course and whether there was correspondence between individual scores on each of the scales. The high inter-rater reliability scores reported, however, suggest reliability in the sense that different teachers would be likely to give similar scores on both scales – an important consideration in any measurement used for assessment purposes.

What Should We Assess in Autonomous Learning Programmes?

My brief in this chapter is essentially to discuss the *possibility* of testing autonomy, and at this point I want to draw a line under what I have written so far by concluding that autonomy does appear to be a 'testable' construct, in the sense that we are able to find ways of measuring it. I have outlined a theoretical framework for the measurement of autonomy based on the idea that autonomy is fundamentally a matter of control over learning processes and I have also raised a number of potential problems with this framework. The studies I have reviewed all approach the larger construct of autonomy through the idea of control, although they do so in very different ways. They point both to the difficulty of measuring autonomy in a global sense and the importance of selecting dimensions of control that are salient to the context and mode of learning. We have also seen how several studies have begun to address problems concerned with the conceptualisation of autonomy as a capacity by attempting to dig into students' rationales for their learning behaviours.

At this point, however, I have a picture in my mind of readers who are committed to the idea of fostering autonomy in foreign language education holding their heads in their hands and wondering where all this talk of measuring autonomy is leading. The studies that I reviewed in the previous section form only a small part of the literature on autonomy in language learning, and perhaps there is a good reason for this. As a field, we are reluctant to engage with the question of testing autonomy, because many of us, and I include myself here, believe that 'testing itself is anti-autonomy' (Champagne *et al.*, 2001: 49). The implications of pursuing research agendas concerned with the measurement of autonomy need, therefore, to be thought through carefully. As a way into these implications, I want to review three papers in which problems of evaluation and assessment in programmes oriented toward autonomy are considered from a broader perspective: Ravindram's (2001) description of the Certificate of Independent Language Learning (CILL) at Temasek Polytechnic, Singapore; Morrison's (2005) study of self-access centre (SAC) evaluation; and Champagne *et al*'s. (2001) study of an action research project concerned with assessment in a taught programme. My concern here is with the different ways in which the authors view the role of assessment and what should be assessed in the context of these programmes.

The CILL is, to the best of my knowledge, unique as a credit-bearing language course in which students are awarded credits for the development of their autonomy, with language proficiency development playing only a minor role in assessment. As Ravindram (2000: 61) describes it, the CILL is based in a self-access centre and involves modules taken over a period of up to three years, structured so that 'each module increasingly demands learners demonstrate their abilities to take greater responsibility in their learning'. The assessment criteria for the course consist of 20 items derived from Knowles's (1975) 'key skills of self-directed learning', most of which can be interpreted as skills concerned with control over learning processes. Credits are awarded based on 'the levels of awareness and understanding students display of these skills, and the extent to which students apply these skills in practice while working through the Certificate programme' (Knowles, 1975: 62) using a grade scale from A to F. The course also involves a variety of assessment tools, including:

- profile of strengths and weaknesses, observations made and outlined in the consultation records drawn up by CILL helpers,
- quality of reflection on learning and task as gleaned from learners' learning logs, learning reviews and contracts,
- the quality of language work submitted,
- learners' ratings of their self-directed learning skills at the end compared to ratings at the end of each module, and,
- a team decision on the allocation of possible credits for the learner. (Ravindram, 2000: 66)

The CILL clearly presents complex problems of assessment, including that of ensuring consistency among teachers. Due to the number and variety of modes of assessment involved, this problem also seems to be more complex than it was in Lai's (2001) study where high inter-rater scores were achieved using relatively straightforward assessment instruments. Nevertheless, Ravindram (2000: 66) reports that calibration and benchmarking exercises had 'thus far shown minimal inter-rater/intra-rater discrepancy in the award of credits for completed learning contract cycles'.

Morrison (2005: 270), writing as the director of a university self-access centre, argues that the issue of learning gain is crucial to the evaluation of the operation of self-access centres, but that 'the variety of types and scope of learning makes any attempt at its definition, analysis and measurement problematic'. Based on interviews with 16 stakeholders in self-access, including teachers and students, Morrison observed a tension

between a 'perceived need to include learning gain as an evaluation focus' and 'a recognition of the practical problems of doing so', with one interviewee, a language centre director saying, 'I don't know how you do it, I really don't' (Morrison, 2005: 276). Twelve of the interviewees saw 'teacher identification of learners' *perceptions* of gain as a meaningful alternative to more easily quantifiable criteria such as test scores' (Morrison, 2005: 280). Morrison also cites one teacher's comment that 'progress could be a feeling of making headway in your proficiency, it might also be a feeling of greater involvement and motivation' (Morrison, 2005: 280). He concludes that 'it is possible to evaluate learning gain in a SAC but not in the expectation of observable measurements and scores that reflect quantifiable instances of learning by the student body as a whole which are validated by an evaluator external to the learning process' (Morrison, 2005: 286–287).

Co-authored by nine participants in a year-long action research project concerned with student assessment in the Talkbase programme at the Asian Institute of Technology in Thailand, Champagne *et al*. (2001: 46) begin by stating the group's original 'intention to develop both quantitative and qualitative measures to assess learner autonomy and language improvement'. Initially designed as a pre-masters language programme, Talkbase adopts an experiential learning approach, in which there is a movement from teacher-defined tasks at the beginning of the programme to student-defined tasks toward the end. Sources of data used for assessment include entry and exit interviews and self-assessments, portfolios of work and ongoing teacher observations and student self-assessments. The group also experimented with conventional language proficiency tests, noting that 'tests seem to be the most accepted evidence of second language acquisition, with test scores more likely to satisfy administrators and funders' (Champagne *et al*., 2001: 49). What is of interest here, however, is that the reported outcomes of the project did not, in fact, include a clear cut set of assessment criteria and procedures. One outcome was an attempt to clarify the learner autonomy and language proficiency goals of the programme. Another was an understanding that the programme needed to move more toward self-assessment and to 'maintain an integrity around the compatibility of testing with an educational approach that aims to be pro-autonomy' (Champagne *et al*., 2001: 54). Champagne *et al.* also describe how their efforts to address issues of assessment revealed 'differences over issues of testing, qualitative-quantitative, goals-objectives' within the group, and raised 'issues that are fundamental to what learner autonomy might

be within our particular language learning context' (Champagne *et al.*, 2001: 46).

In reviewing these three studies, the point I want to highlight is that assessment in autonomy-oriented programmes tends to be a complex matter, involving a kind of 'ecology' between assessment content and procedures, and accountability requirements, on the one hand, and programme goals, on the other. As Sinclair (1999: 96) puts it, 'developing greater independence is of little benefit unless it culminates in successful language learning' and, in practice, external accountability is usually a matter of meeting language proficiency targets. We accept, in other words, that we must in some sense test language proficiency. But at the same time we are reluctant to employ standardised assessment tests that fail to address the divergent outcomes that autonomy-oriented programmes aim at. We are also suspicious of conventional entrance and exit tests, which serve to 'reinforce (on the crucial first and last days of the participants' experience on the program) traditional notions of teacher control and student accountability' (Champagne *et al.*, 2001: 49). We are primarily concerned, therefore, to balance accountability requirements with programme goals through assessment procedures that reward both language proficiency and autonomy, to various degrees, and involve a good deal of self- and negotiated assessment (e.g. Dam, 2000; Karlsson *et al.*, 1997).

But why exactly, in this context, would we want to assess students' *autonomy*? In the case of the CILL, Ravindram (2000: 60) argues that this is in line with 'recognition of the importance of independent learning accorded at national level' in Singapore. In my experience, however, although autonomy-oriented programmes are often justified by policy-level initiatives, they are usually held accountable for the achievement of language proficiency rather than the achievement of autonomy. It is possible, however, that we want to assess students' autonomy because we orient toward a broader kind of accountability. Trim (1997: v–vi), for example, argues:

> During the period of general education at school, before the career threshold is reached, it is not possible for a young person to know which language he or she will need for what purposes in adult life. The opportunities offered for mobility, both educational and occupational, are becoming wider and more frequent. Accordingly it is an increasingly important part of the responsibility of schools, not only to enable pupils to reach a worthwhile level of proficiency in the particular language or languages being taught, but also to equip

learners with the attitudes or skills which will enable them to continue to plan, carry out and monitor their own learning once all the supporting and disciplinary structures of institutionalised learning are withdrawn.

There is, in other words, a growing expectation that foreign language education will produce autonomous language learners. Although this expectation has not yet filtered down into accountability mechanisms, it is frequently found in policy documents. And it is possible that, in an educational climate in which there is a close relationship between the value of educational achievements and their measurability, we are encouraged to think of autonomy as both a measurable and testable construct.

The risk in pursuing research agendas concerned with the measurement of autonomy is, then, that we might find ourselves trapped in a logic that leads from the idea that autonomy can be measured to the construction of tests, and from the construction of tests to their implementation in student assessment. In this context, we might ask what the likely consequences of 'autonomy testing' are. Students are naturally concerned with the consequences of passing and failing tests. Would they quickly learn what they are required to do in order to achieve high grades? Would their teachers find that they were no longer engaged in fostering autonomy, but in helping their students put on Breen and Mann's (1997) 'mask of autonomous behaviour' or, perhaps, in devising ever more ingenious ways of forcing their students to remove the mask?

Conclusion

In this chapter, I have argued that autonomy is a testable construct in foreign language education contexts in the sense that it can reasonably be reduced to measurable behaviours in which control over aspects of the language learning process are displayed. But there are also certain difficulties with the measurement of autonomy, among which assessing the meaning and authenticity of learning behaviours seem to be the most intractable. The idea of measuring autonomy is of most value, in my view, in the context of research, because it has the potential to place our understanding of the ways in which autonomy develops under various circumstances on firmer empirical ground. At the beginning of this chapter, however, I suggested that this may not be the 'right' way to think about autonomy. What I meant by this is that we aim, in general, to foster autonomy because we believe it brings benefits to our students, including

foreign language proficiency. This aim does not imply any need to assess the autonomy of our students, and it is, in fact, only fairly recently that we have begun to perceive this as a need. This has happened, I believe, largely because broader educational discourses encourage us to view everything we do as being potentially measurable. And we should perhaps consider whether in thinking of autonomy as a measurable construct, we are not simply internalising discourses of this kind.

References

Benson, P. (2001) *Teaching and Researching Autonomy in Language Learning*. London: Longman.

Benson, P. (2003) A Bacardi by the pool. In A. Barfield and M. Nix (eds) *Learner and Teacher Autonomy in Japan 1: Autonomy You Ask!* (pp. 275–282). Tokyo: Learner Development Special Interest Group of the Japan Association of Language Teachers.

Benson, P. (2007) Autonomy in language teaching and learning. State of the Art Article. *Language Teaching* 40 (1), 21–40.

Bonham, L.A. (1991) Guglielmino's self-directed learning readiness scale: What does it measure? *Adult Education Quarterly* 41 (2), 91–99.

Breen, M.P. (1987) Contemporary paradigms in syllabus design. *Language Teaching* 20 (3), 157–174.

Breen, M.P. and Mann, S. (1997) Shooting arrows at the sun: Perspectives on a pedagogy for autonomy. In P. Benson and P. Voller (eds) *Autonomy and Independence in Language Learning* (pp. 132–149). London: Longman.

Brockett, R.G. and Hiemstra, R. (1991) *Self-direction in Adult Learning: Perspectives on Theory, Research, and Practice.* London: Routledge.

Candy, P.C. (1991) *Self-direction for Lifelong Learning.* San Francisco, CA: Jossey-Bass.

Champagne, M.F., Clayton, T., Dimmitt, N., Laszewski, M., Savage, W., Shaw, J., Sroupe, R., Thien, M.M. and Walter, P. (2001) The assessment of learner autonomy and language learning. In L. Dam (ed.) *Learner Autonomy: New Insights/Autonomy de l'apprenant: Nouvelle pistes*. Special issue of *The AILA Review* 15, 45–55.

Confessore, G.J. and Park, E.M. (2004) Factor validation of the learner autonomy profile (Version 3.0) and extraction of the short form. *International Journal of Self-Directed Learning* 1 (1), 39–58.

Cotterall, S. (1995) Readiness for autonomy: Investigating learner beliefs. *System* 23 (2), 195–206.

Dam, L. (2000) Evaluating autonomous learning. In B. Sinclair, T. Lamb and I. McGrath (eds) *Learner Autonomy, Teacher Autonomy: Future Directions* (pp. 48–59). London: Longman.

Guglielmino, L.M. (1977) Development of the self-directed learning readiness scale. PhD thesis. University of Georgia.

Holec, H. (1981) *Autonomy in Foreign Language Learning*. Oxford: Pergamon (first published 1979, Strasbourg: Council of Europe).

Holec, H. (1985) On autonomy: Some elementary concepts. In P. Riley (ed.) *Discourse and Learning* (pp. 173–190). London: Longman.

Karlsson, L., Kjisik, F. and Nordlund, J. (1997) *From Here to Autonomy*. Helsinki: Helsinki University Press.

Knowles, M. (1975) *Self-directed Learning: A Guide for Learners and Teachers*. New York: Cambridge, The Adult Education Company.

Lai, J. (2001) Towards an analytic approach to assessing learner autonomy. In L. Dam (ed.) *Learner Autonomy: New Insights/Autonomy de l'apprenant: Nouvelle pistes*. Special issue of *The AILA Review* 15, 34–44.

Little, D. (1991) *Learner Autonomy. 1: Definitions, Issues and Problems.* Dublin: Authentik.

Littlewood, W. (1996) Autonomy: An Anatomy and a Framework. *System* 24 (4), 427–435.

Morrison, B. (2005) Evaluating learning gain in a self-access language learning centre. *Language Teaching Research* 9 (3), 267–293.

Nunan, D. (1997) Designing and adapting materials to encourage learner autonomy. In P. Benson and P. Voller (eds) *Autonomy and Independence in Language Learning* (pp. 192–203). London: Longman.

Ravindram, R. (2000) Towards learner autonomy: Evolution in practice: The certificate in independent language learning. In B. Sinclair, T. Lamb and I. McGrath (eds) *Learner Autonomy, Teacher Autonomy: Future Directions* (pp. 60–72). London: Longman.

Rivers, W.P. (2001) Autonomy at all costs: An ethnography of metacognitive self-assessment and self-management among experienced language learners. *The Modern Language Journal* 85 (2), 279–290.

Rowsell, L.V. and Libben, G. (1994) The sound of one-hand clapping: How to succeed in independent language learning. *Canadian Modern Language Review* 50 (4), 668–688.

Simmons, D. and Wheeler, S. (1995) *The Process Syllabus in Action.* Sydney: National Centre for English Language Teaching and Research.

Sinclair, B. (1999) Wrestling with a jelly: The evaluation of learner autonomy. In B. Morrison (ed.) *Experiments and Evaluation in Self-access Language Learning* (pp. 95–109). Hong Kong: Hong Kong Association for Self-Access Learning and Development.

Strauss, A. and Corbin, J. (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques* (3rd edn). Newbury Park, CA: Sage Publications.

Trim, J.L.M. (1997) Preface. In H. Holec and I. Huttunen (eds) *Learner Autonomy in Modern Languages: Research and Development* (pp. v–vi). Strasburg: Editions du Conseil de l'Europe.

Ushioda, E. (1996) *Learner Autonomy 5: The Role of Motivation.* Dublin: Authentik.

Wall, S. (2003) Freedom as a political ideal. In E.F. Paul, F.D. Miller and J. Paul (eds) *Autonomy* (pp. 307–334). Cambridge: Cambridge University Press.

Wenden, A. (1991) *Learner Strategies for Learner Autonomy.* London: Prentice Hall International.

Wenden, A. (1998) Metacognitive knowledge and language learning. *Applied Linguistics* 19 (4), 515–537.

Young, R. (1986) *Personal Autonomy: Beyond Negative and Positive Liberty.* London: Croom Helm.

Chapter 6

Assessment of *Autonomy or Assessment* for *Autonomy? Evaluating Learner Autonomy for Formative Purposes*

TERRY LAMB

Introduction

This chapter argues for assessment in relation to learner autonomy, but in a formative rather than a summative sense. Building on principles of formative assessment (Black & Wiliam, 1998), it examines its links to learner autonomy, and argues that, just as learning can be enhanced by assessment *for* learning, so can autonomy (and hence, it suggests, learning) be enhanced by assessment *for* autonomy.

After defining elements that underpin learner autonomy, namely, metacognitive knowledge and beliefs about learning, the chapter explores a method of enabling young language learners to bring these elements to the surface in order to find ways of enhancing their autonomous behaviours and in order to facilitate assessment for learning. Drawing on a research project designed initially to explore the relationships between learner autonomy and motivation, a methodology is proposed to engage young language learners in talking about their own constructions of learning. Although the research was not intended to be interventionist, the chapter argues that the quality of data produced suggests that the method, focused group conversations (FGCs), can enable learners to reflect on their autonomy, as a form of assessment for autonomy, and take greater control over and responsibility for their learning. Following the principle that such forms of assessment should also provide feedback to teachers to enable them to reflect on and adapt the classroom experience, examples are shown of ways in which a range of learners' voices offer clear but varied suggestions for teacher intervention with a view to empowering the

learners to have control over their learning in ways that are sensitive to their diverse needs. The research thus offers tools that can enable teachers themselves to investigate their own learners' learning and to adapt their teaching in ways that are appropriate to their own contexts.

Assessment for Learning

The publication of Black and Wiliam's (1998) paper, 'Assessment and classroom learning' and the subsequent booklet, *Inside the Black Box* (Black *et al.*, 2002), had a profound influence on schools in England through its incorporation by official bodies into classroom teaching and learning reforms. Black and his colleagues' work drew on 250 papers on assessment and led to a shift in official understandings of its role in learning. Since the Education Reform Act of 1988, with its focus on enhanced, measurable quality and the encouragement of market forces and parental choice, there had been an escalation of highly public, summative forms of assessment (examinations and tests) that over-shadowed ongoing assessment in the classroom. However, in the early years of the new millennium, the Department for Education and Skills (DfES), the Qualifications and Curriculum Authority (QCA) and the Office of Standards in Education (OFSTED) officially introduced the notion of assessment for learning (the term used officially in England instead of formative assessment) into the curriculum (Looney & Wiliam, 2005: 129–132).

This move has had parallels in other countries. An Organisation for Economic Cooperation and Development (OECD) report in 2005 ana-lysed ways in which formative assessment was promoted in education reforms in eight countries (Australia, Canada, Denmark, England, Finland, Italy, New Zealand and Scotland), finding that such develop-ment was motivated by a desire to improve pupil achievement and equity of outcomes through increased focus on learning to learn (CERI, 2005). Central to such a form of assessment is the pupil's own active involvement in the learning process through the development of metacognition, defined by Flavell *et al.* (2002: 164) as 'any knowledge or cognitive activity that takes as its object, or regulates, any aspect of any cognitive enterprise... Metacognitive territory includes both what you know about cognition and how you manage your own cognition'. Black and Jones (2006: 8) add another dimension to this definition when they describe it as 'the power to oversee and steer one's own learning so that one can become a more committed, responsible and effective learner'.

Assessment for learning (rather than assessment *of* learning), has been succinctly defined by Black and Jones:

> Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting pupils' learning...
> An assessment activity can help learning if it provides information to be used as feedback, by teachers, and by their pupils in assessing themselves and each other, to modify the teaching and learning activities in which they are engaged. Such assessment becomes ''formative assessment'' when the evidence is actually used to adapt the teaching work to meet learning needs. (Black & Jones, 2006: 4)

Key characteristics of assessment for learning are described in support materials for secondary teachers as follows:

- Sharing learning objectives with pupils.
- Helping pupils to know and recognise the standards they are aiming for.
- Involving pupils in peer and self-assessment.
- Providing feedback that leads pupils to recognising their next steps and how to take them.
- Promoting confidence that every pupil can improve.
- Involving both teacher and pupil in reviewing and reflecting on assessment information. (DfES, 2004: 3–4)

Clearly implied in this is a shift in teacher-pupil roles, with teachers supporting pupils in the development of skills and knowledge to enable them to become more aware of themselves as learners and to understand that they have a responsibility for their own learning (Black & Wiliam, 2005: 232–233).

There are obvious overlaps between such principles of assessment for learning and definitions of learner autonomy in language learning. Though such definitions are multifarious, they usually involve elements of self-management (planning, monitoring and evaluating one's own learning) and self-regulation (the cognitive factors involved in managing one's own learning that make self-management effective) (Benson, 2001: 48–49; Lamb & Reinders, 2005). It could indeed be said that assessment for learning is designed to develop the necessary capacities for becoming an autonomous learner with a view to improving learning through better self-monitoring and self-evaluation leading to better planning.

There is, however, evidence to suggest that secondary pupils do not always demonstrate such behaviours. Lee *et al.*'s (1998) research into

teenage language learners revealed an apparent lack of awareness of the process of language learning, including the purpose of learning activities and how they relate to progression in learning. Similarly, after finding evidence of poor ability to discuss learning strategies or to plan work, Williams *et al.* (2002) called for further research into the development of cognitive and metacognitive strategies as well as learners' ability to control the process of learning in order to 'provide insights into ways to enable learners to see value in and take control of their learning in school' (Williams *et al.*, 2002: 525). Such studies suggest then that assessment for learning could usefully be enhanced by a specific focus on *assessment for autonomy* that, drawing on Black and Jones' (2006) definition of assessment for learning above, would mean any assessment for which the first priority in its design and practice is to serve the purpose of promoting pupils' *autonomy.*

Assessment for Autonomy

The purpose of assessment for autonomy, therefore, would not be to measure autonomy for its own sake, with a view to defining levels of ability or ranking pupils, but to increase learners' self-awareness of their own autonomy and teachers' awareness of what constitutes such autonomy and how they may adapt their teaching in order to enhance it.

How, though, can autonomy be usefully described in such a way as to provide a useful tool for assessment for autonomy? A way forward can be found in the autonomy literature, which suggests that in order for learners to be able to manage their own learning, there is a need for self-regulation; in other words, there is a need for learners to have control over their cognitive processes. Such control requires the development of metacognitive knowledge (Jiménez Raya, 1998; Lamb, 2006a; Victori & Lockhart, 1995; Wenden, 2001), defined by Flavell (1985) as knowledge about the self as learner (person knowledge), the tasks involved in learning (task knowledge) and the strategies that can be called into play in order for learning to take place (strategy knowledge).

Wenden suggests that metacognitive knowledge has the following properties:

- it is stable
- it is acquired unconsciously or consciously
- with cognitive maturity comes the ability to reflect on the learning process and develop new assumptions
- it can be brought to consciousness and talked about
- it is a system of related ideas. (Wenden, 1999: 435)

Wenden (1999: 436) reminds us that metacognitive knowledge is not the same as metacognitive strategies, since the former is knowledge acquired about learning, whereas the latter are specific ways in which learners manage, direct, regulate and guide their learning. In fact, it has been claimed that the development of metacognitive knowledge is more important to learning than attempting to teach strategies 'because this knowledge can form the basis for selecting and activating one strategy over another' (Rubin, 1987: 19) [though strategy training has been by far the major focus of learner training (Wenden, 1996: 246)].

Learner beliefs are often used as a synonym for metacognitive knowledge (e.g. Victori, 1999). Yang (1999), however, views beliefs as including metacognitive knowledge, but proposes that they consist of a broader theoretical construct that encompasses motivational beliefs (the latter defined by Garcia and Pintrich (1995) as learners' beliefs about their ability to learn a language, their expectations regarding level of difficulty of the tasks, their goals and reasons for learning a language and their emotional reactions to second language learning). Since then, other researchers have demonstrated how learner beliefs influence aspects of language learning (e.g. Benson & Lor, 1999; Cotterall, 1999; Mori, 1999).

What is clear from this literature is that there is still a great amount of work to be done to explore the nature and role of metacognitive knowledge and learner beliefs in language learning, particularly in specific contexts (such as with young language learners and learners of languages other than English). Indeed, Wenden (2001) has described metacognitive knowledge in second language acquisition as 'the neglected variable'. Gaining access to such knowledge is not easy, of course, since learners are not always accustomed to discussing it (Rudduck *et al.*, 1997). However, given the connections between learning, learner autonomy and assessment for learning, and the crucial role played by metacognitive knowledge in learner autonomy, it would seem appropriate to consider ways in which accessing it might offer a framework for assessing autonomy in a formative way in order to enhance learning.

If we are to consider this as a framework for assessing autonomy formatively, the next question is how to gain access to young learners' metacognitive knowledge and their beliefs about learning in ways that will be meaningful both to themselves and the teacher, so that teaching and learning can be adapted in order to develop greater autonomy. A way forward will now be proposed, drawing on an ethnographic study originally designed to explore the relationships between autonomy and

motivation among teenage learners of French and German in an inner-city school in Yorkshire, England. This research used the tool of the focused group conversation in order to encourage the construction of and reflection on these elements of learners' autonomy.

Autonomy and Motivation: Background to the Research

The background to this research project was the lack of popularity of language learning in English schools and the need to find ways of enhancing secondary language learners' motivation. This need has become even more acute following the rapid drop in the number of 14–16 year olds learning languages since it stopped being a compulsory part of the curriculum in 2004 and changed to an 'entitlement'. Indeed, in more than half of all maintained schools in England, less than 50% of the learners are now studying a language between the ages of 14 and 16, and in many schools less than 25% of this age group is still learning a language. On a regional level, Yorkshire schools have fewer of these learners studying languages than anywhere else apart from the North East of England (CILT, ALL, ISMLA, 2006).

My interest in the relationship between autonomy and motivation originated in my own experiences as a teacher of French and German in secondary schools, when I had had to contend with some learners' lack of motivation to learn my subject. As pupils' concentration levels seemed to be reducing year by year, and individual differences within the same class seemed to be becoming more marked, I had begun to explore new ways of organising my classroom in order to create a learning environment where I could provide differentiated learning activities in which pupils could be actively involved. The resultant system of 'flexible' learning (Lamb, 1996, 1998) offered pupils the opportunity to make decisions about which learning objectives to focus on and which tasks to do in order to achieve these objectives, as well as to assess themselves. This appeared to be successful in that learners expressed a preference for working autonomously, motivation levels appeared to increase and examination results improved considerably (Lamb, 1998). The research project described in this chapter was designed to explore this further, to understand the relationships better and to build on the experience.

Autonomy and motivation: Previous studies

The work on self-determination carried out by Deci and colleagues (e.g. Deci, 1980; Deci & Ryan, 1987; Deci *et al.*, 1991) has revealed a strong link between intrinsic motivation and autonomy, with autonomy

postulated as one of three innate drives or psychological needs to which intrinsic motivation is a response (the other two are competence and relatedness). For Deci *et al.* (1991: 327) autonomy is about 'being self-initiating and self-regulating of one's own actions'; individuals need to feel that they are able to act on the world around them, making choices and determining their actions in ways that are appropriate given their own strengths and weaknesses and the constraints placed on them. Furthermore, external control of such behaviour will affect intrinsic motivation; Deci and Ryan (1987) have summarised research into what exactly curtails self-determination, itemising such contextual features as rewards used as a control, threats, deadlines, evaluation, surveillance, limited choice, and other forms of feedback.

Further connections with autonomy can be found in Rotter's (1966) notion of 'locus of control', which concerns the extent to which behaviours influence events and where these behaviours are located. Learners with external locus of control, he argued, believe that their actions have little impact on events. Conversely, those with internal locus of control believe that they have control over events. This influenced a plethora of research studies (e.g. Deci, 1975; Deci & Ryan, 1985; Findley & Cooper, 1983; Pintrich & Schunk, 1996), which indicate broad agreement that having a sense of control over what is happening (and, hence, a sense of personal responsibility) is a key factor in both initial and continuing motivation.

Closely related to this is attribution theory (e.g. Weiner, 1984), which is concerned with the ways in which students explain success or failure, and what they do as a result. Work relating to this theory tends to analyse causal attributions for success or failure according to whether the cause is perceived to be external or internal, and whether it is considered to be stable (e.g. due to ability) or unstable (e.g. due to effort). To summarise, such work considers it to be good for motivation if success is perceived to be due to internal, stable causes or if failure is put down to unstable causes. In other words, a learner will be more motivated if she/he views success as resulting from high ability (internal, stable) rather than luck (external, unstable), whereas failure will be easier to cope with if it is attributed to short-term illness (internal, unstable) rather than lack of ability (internal, stable).

There has also been a small but growing body of research specifically into the relationship between language learning motivation and autonomy. Although this research is still 'relatively new' (Dörnyei, 2001: 59), there is increasing evidence that the two are related, even if the nature of this relationship is unclear. It cannot be claimed, for example, that an

increased use of learning strategies leads to enhanced motivation (e.g. O'Malley & Chamot, 1990; Simmons, 1996). On the other hand, other studies (e.g. Oxford, 1990) suggest that it is possible to develop affective strategies (such as strategies for lowering anxiety, encouraging yourself, and taking your emotional temperature), shifting the focus away from motivating our learners to finding ways of helping them to motivate themselves.

A further question relates to which comes first; is motivation needed for autonomy to develop, or vice versa? Spratt *et al.*'s (2002) quantitative research among university students in Hong Kong suggested that motivation was needed for learners to be able to behave autonomously, whereas Harris and Noyau's (1990) and Lamb's (1996, 2001; Lamb & Fisher, 1999) secondary classroom interventions have suggested that increased autonomy can lead to increased motivation. However, this is a relatively new area of research, which needs further empirical exploration (Benson, 2001: 70).

Exploring Metacognitive Knowledge and Beliefs about Learning

Accessing (and assessing?) metacognitive knowledge through focused group conversations

Returning to my own research, I decided to explore learner autonomy as manifested in the metacognitive knowledge and beliefs about learning of young language learners, and the ways in which these appeared to relate to their motivation to learn a language. The intention was to build on my previous studies (Lamb, 1996, 2001; Lamb & Fisher, 1999), and find a way of helping young learners to describe and reflect on aspects of their autonomy.

In designing the research, I was influenced on an ontological level by theories in urban education, which led me to a commitment to avoiding assumptions of deficit. In other words, I was committed to exploring ways in which social and educational structures contributed to the emergence of disadvantaged groups, and, as a small manifestation of such disadvantage, to a poor uptake of language learning among these groups. The concept of disaffection was thus for me more than a lack of engagement (e.g. Rudduck *et al.*, 1996), but rather a result of the disenfranchisement experienced by certain pupils, leading to a desire for agency that, drawing on resistance theories (e.g. Giroux, 1983; Mirón & Lauria, 1998; Sarup, 1991), may be expressed actively or passively, and in constructive or destructive ways. I thus reconceptualised disaffection as a *search for a voice*

in a context of disenfranchisement (Lamb, 2000a, 2000b). An implication of this was that I needed to explore language learning from the perspective of the learners. I therefore decided on ethnographically oriented research, adopting a hermeneutic approach to understanding language learners' experiences rather than traditional, positivist approaches to motivation research. My research took the position that the learners are the experts in voicing their own understandings of learning.

Given this ontological position, my research was designed to explore the learners' metacognitive knowledge and beliefs about learning with a focus on specific motivational beliefs relating to control and responsibility, and was guided by the following sub-questions:

- Person knowledge: How do the pupils construe the role of the learner (including themselves) in language learning, specifically in terms of control over and responsibility for learning?
- Task knowledge I: How do they construe the task (i.e. the nature and purpose) of language learning, both generally and in their own experience?
- Task knowledge II: How do they construe the individual tasks involved in language learning, and how might this inform the ways in which they take control of their learning?
- Strategic knowledge: How do they construe the strategies they use in order to learn, and how might these contribute to their learning?

I did not originally set out to distinguish between metacognitive knowledge and beliefs in the research design. After analysing the pupils' person knowledge, however, I realised that it was important to distinguish between two types of knowledge in this context: the first was knowledge gained about the realities of learning in this particular context, and related specifically to what pupils knew about themselves and their own learning contexts; the second was related to their knowledge about learning in general. As it became clear that the two were not necessarily the same, and that the level of congruence or incongruence was significant, it seemed useful to refer to the former as knowledge and the latter as beliefs. For example, based on their experience of using English, some learners had an unrealistic expectation of the task of language learning, believing that they should be able to communicate fluently after a short period of time; of course, the reality was that they were unable to use the language fluently during the early years, and this sometimes led to frustration and demotivation.

In order to address the methodological challenge of accessing the learners' metacognitive knowledge and beliefs about learning, I decided

to develop a form of focus group interview, described by Mertens (1998: 174) as 'group interviews that rely, not on a question-and-answer format of interview, but on the interaction within the group'. Drawing on a range of interview techniques, such as Roy's (1991) cognitive interviewing, Holstein and Gubrium's (1995) 'active interviewing' and Krueger's (1994) focus groups, I designed a series of FGCs. Each FGC had a different focus whilst attempting to retain a spirit of openness, dynamism, spontaneity and interaction in order to support the articulation of what may be difficult constructions. As an interviewer, I saw myself as a facilitator, creating an enabling environment. It was also important for the FGCs to be enjoyable and non-threatening for participants.

My research focused on pupils in Y9 moving into Y10 (aged 13 at the beginning of the research, which continued for 18 months), firstly because these pupils would have had sufficient experience of learning languages in the school (two years) to be able to discuss their experience, and secondly, because Y9 had been identified as a 'problem year' by the languages department. Four different groups of pupils took part, two drawn from the higher achieving half of Y9 and two from the lower achieving half. In each half, I worked with a group of motivated pupils and a group of unmotivated pupils as identified by their teachers. The groups were therefore constituted as follows:

9A1 – lower achievers, motivated
9A2 – lower achievers, not motivated
9B1 – higher achievers, motivated
9B2 – higher achievers, not motivated

I decided on a group size of six, and planned six 50–60 minute meetings with each group in the hope that they would become increasingly comfortable, encouraging a more in-depth exploration in an atmosphere of trust and confidence over time. The 18-month period of research also enabled insights to be gained into learners' changing experiences as they moved closer to their GCSE examinations (taken in England at the age of 16), as well as allowing aspects of the data to be revisited in different ways, enhancing the reliability and validity of the research.

Interview protocol design

Detailed interview protocols were designed, containing questions and stimuli that would be meaningful and appropriate to the pupils. In order

to minimise my influence, I drew on Tomlinson's (1989) hierarchical focusing, a form of questioning that is common in the languages classroom. It consists simply of moving from open-ended questions, which give pupils the opportunity to provide their own answers in unrestricted ways, to more defined questions for those not confident enough to cope at that level.

In order to stimulate interaction, a wide range of varied activities were included, which were very similar to classroom activities in nature if not focus (though the success of the activities in engaging the pupils in talking about learning suggests that they may well form the focus of classroom-based assessment for autonomy activities to enhance assessment for learning). Activities included:

- Brainstorms, either as a warm up or as a way of establishing a range of ideas, e.g. what is enjoyable (or not) about language learning; activities to practice listening, speaking, reading and writing, followed by a discussion of purpose, usefulness and enjoyability.
- 'Concept mapping' (Powney & Watts, 1987: 30), in which pupils' ideas would be placed onto the board in a certain order, e.g. to trace the language learning cycle from the introduction to the end of a new unit. (This is useful when the issue is about sequence of activities or ideas rather than just the ideas themselves, since it allows a visual support to the discussion.)
- Contextualised questions, encouraging pupils to think back to specific learning experiences (the last module, an occasion where they did well, a specific activity they enjoyed). For example, when I wanted to establish what had actually happened rather than what might happen, I would contextualise the question in this way.
- Direct questions (e.g. What makes a good teacher/learner? How do you learn vocabulary? Have you ever done it in a different way?), since some pupils respond better to direct approaches, and I wanted to make sure that my group discussions were differentiated enough to allow these voices to emerge.
- Projective techniques (e.g. designing a poster to persuade people to learn languages; advice page on how to address specific language learning issues; pretending they are the teacher writing a report on their progress). These techniques stimulate the imagination, thereby opening up the mind to deeper forms of expression, facilitating access to ideas that are abstract (LeCompte & Preissle, 1993: 164). They also enable pupils to give 'unsafe' opinions in a 'safe' way (as they are projecting them onto a different person or a different context).

- Use of drawing (e.g. in pairs, drawing an imaginary languages classroom of the future), after which the pupils described the thinking behind the drawings.
- Questionnaires involving, for example, sentence completion, e.g. 'In French/German I think/don't think I'm very good at . . .'; 'I'd do better if I . . .', used as a stimulus for discussion. Another example involved filling in a table, asking pupils to rate (then discuss) the importance of different purposes for learning languages.
- Self-rating scales (e.g. a Likert scale), not in order to get statistical precision, but to stimulate discussion – for example, to help students clarify perspectives before sharing them.

Summary of the Data

The research design produced rich data, and all groups were positive about the conversations. Pupils found them an opportunity to discuss new and interesting issues, with Helen[1] (A1) stating that she 'thought they were good because we spoke about things that we've never spoke about before'. Others found them an opportunity to 'speak your mind' (Candice, A2) and felt that they should talk about such issues with their teachers as a way of influencing the lessons, 'so they know how you feel' (Mick, B1). Among others, Helen (A1) suggested that the conversations may have had some impact on their learning (though the research intention had been to explore rather than to intervene):

> Yeah. I've started to listen a bit more and do things, you know like when we discussed about things you could do to improve.

It is, of course, beyond the main focus of this chapter to present the many individual voices of the pupils directly (see Lamb, 2005, 2006a, 2006b). However, on analysing the data, I discovered that each of the four groups taking part in the FGCs had certain characteristics in common, and I found myself referring to them using names that reflected their overall group characters. The rest of this chapter will therefore be devoted to brief summaries of elements of the different groups' autonomy as revealed through the FGCs. These summaries also include implications for future teacher interventions, thus illustrating the potential of the FGC to be a sensitive tool that can access diverse needs and inform future learning and teaching in the spirit of assessment for autonomy. This leads us to conclude that both the framework of metacognitive knowledge and beliefs about learning and the tool of the FGC offer a way forward in facilitating assessment for autonomy.

The grafters (A1)

These pupils believe strongly that the learner should have control over learning, but, because they attribute their failure to maintain this control on internal failings, such as poor concentration, they recognise that they let themselves down, meaning that there is incongruence between their person beliefs about what constitutes a good learner and their knowledge of themselves as learners. Unfortunately, they believe these constraints on learning to be stable and unchangeable, as they do not know how to control themselves better. The internal constraints also include little evidence of self-management skills, and naïve levels of self-regulation, particularly in areas of strategic knowledge that would enable them to practise and remember the language more effectively.

Because they have modest expectations of language learning as a task, the way in which they see it manifested in the school is congruent with these beliefs, though they would appreciate more opportunities for independent learning, as they see this as an opportunity for them to concentrate on their learning more easily. However, they display only a rudimentary knowledge of specific tasks, which limits the effectiveness of the choices that are fundamental to independent learning.

These pupils are mainly satisfied with the school language learning context, but recognise to some extent that it could be more engaging. Again, however, they perceive this as stable or unchangeable as they have no influence over their teachers. However, they respond to this weakness passively, for example through loss of concentration, rather than through any strategic action, which only exacerbates their difficulties with learning.

Most of all, this group would benefit from development of their strategic knowledge, in particular their self-regulatory strategies, and from increasing opportunities to manage their own learning accompanied by awareness of the nature and purpose of specific tasks.

The angry victims (A2)

The angry victims have strong beliefs that the learner has no control over learning, that it is something that is done to them. Because of this, they attribute any failure to learn to others, mainly their teacher, and believe themselves powerless to influence this, thus rendering these constraints on learning unchangeable or stable. This suggests that their person knowledge and their beliefs about the learner's role are congruent, and that their experiences in school are perceived as disempowering and mystifying. There is some evidence of internal

attribution of failure, but again this is attributed to stable causes, such as lack of confidence or ability. They also reveal little evidence of any knowledge about self-regulation or self-management, which means that their failure and powerlessness in the face of learning is reinforced for them.

Their unrealistic beliefs about what they should be able to do in another language (e.g. speak it like they speak English) are incongruent with their knowledge about the nature of language learning in the school, which is thus condemned as 'crap'. Their only positive experiences are when they are working in small groups, as they are able at best to support each other's learning, at worst (and more commonly) to become allies against the teacher. Possibly, for this reason they are offered few opportunities to work in this way as the teacher tries to keep control of the class through teacher-centred, undemanding copying activities. Their disconnection from the business of learning languages is reflected in their confusion over the purposes of specific tasks.

These pupils are completely disaffected by school in general, and by language learning specifically, and believe that they are being victimised. In order to maintain some element of control over the situation, their only form of resistance is active disruption, which of course leads to negative outcomes in the form of an increasingly serious breakdown in communication with their teacher, and escalating failure and disaffection.

What they would benefit from most of all is more guided reflection on the nature of learning and the active role of the learner, more strategic knowledge and awareness of the purpose of specific tasks to allow them successfully to take on a more active role, and reflection on what can realistically be expected from language learning in school.

The sophisticates (B1)

The sophisticates have very strong beliefs in the active role that the learner should play in learning, and in the idea that the learner is ultimately responsible for learning outcomes. They are largely confident that they are able to assume this role, revealing sophisticated levels of strategic knowledge in the areas of self-regulation and self-management (including self-monitoring and self-evaluation), and are thus able to address their own perceived weaknesses (internal constraints) when they arise. Consequently, they attribute most constraints on learning to external factors such as the teacher, who sometimes does not live up to

their high standards, though they are prepared to discuss this with the teacher (who sometimes listens and responds). Both external and internal constraints on their control over learning are thus to a certain extent perceived as unstable.

Because of the ambitious nature of their beliefs about the nature and purposes of language learning, their experiences of learning languages in school do not live up to their expectations, especially in terms of content (which they consider to be unchallenging compared with what they are able to discuss in other subjects). This means that their general task knowledge and beliefs are incongruent, and could potentially lead to demotivation. They appreciate opportunities to learn independently, either as individuals or groups, as choice affords them opportunities to take control of their learning. Their chances of working effectively in this way are enhanced by their strong knowledge of specific tasks, which includes rigorous evaluation criteria.

The group needs genuine opportunities to continue to manage their own learning and make real choices about what they learn and how. They need open structures in which their concerns can be heard and acted on.

The frustrated (B2)

This group believes that learners should be able to take control of their own learning, and they attribute any failure to do so mainly to themselves. They have high expectations of themselves, but do not know how to meet them, thus considering their weaknesses to be stable. Their person knowledge and their beliefs about the learner are therefore highly incongruent, contributing to a sense of frustration and help-lessness. This is exacerbated by weak knowledge of self-regulation strategies, particularly in areas that contribute to memorising and activating language, areas they value most highly.

These pupils strongly appreciate the opportunity to work independently as it means that their brains need to be engaged, thus offering some control over learning. The group does reveal self-management knowledge, but they appear to have little awareness of the purposes of different specific tasks.

The frustration of these pupils is largely turned inward, though there is some external attribution that emerges later, whereby they reveal their dissatisfaction with the way in which teachers fail to help them achieve what they expect to achieve. Unfortunately, they feel unable to influence this either, as they do not believe that their voices will be listened to.

Their resistance is thus passive, taking the form of quiet disengagement in the classroom or even of absenteeism, which of course has no positive influence over the situation.

The most obvious benefits to this group would be gained from a curriculum that enables them to reflect on the expected (realistic) outcomes of language learning in school, and which develops their strategic knowledge, whilst offering them more self-management opportunities with real choice so that they can influence the content, and practise what they need to practise. They would of course need to have the opportunity to learn more about the nature and purposes of specific tasks.

Conclusions

The research found that the differences between the groups are so complex that any comparison, in terms of the relationship between autonomy and motivation, can only be tentative. What did emerge, however, is that there are clear relationships between motivation and autonomy, but they are not clear cut, and it is more in the combinations of different types of metacognitive knowledge and beliefs, the congruence or incongruence between knowledge and beliefs, and external responses to these that the relationships appear to lie (Lamb, 2005). This suggests that reflection on these factors with a view to both pupil and teacher adaptation could potentially enhance both motivation and, in turn, learning.

The main purpose of this chapter, however, was to explore a method of bringing young learners' metacognitive knowledge and beliefs about learning to the surface in order to explore ways of enhancing their autonomous behaviours to support assessment for learning. The research that informed the chapter was designed not to *assess* the pupils' autonomy, but to attempt to understand the ways in which their constructions of language learning and their motivational beliefs relating to control and responsibility interrelated to affect their motivation. Nevertheless, what emerged was a rich picture of language learning from the pupils' perspective, which endorsed the use of FGCs as a way of engaging learners in talk about learning.

Of course, it is not always possible for teachers to organise FGCs with their pupils in this way. Nevertheless, the need to offer opportunities for pupils to talk about their learning is clear, and this can be integrated into everyday lessons. I have argued that assessment for learning needs to be complemented by assessment for autonomy in order to enable learners to

be able to take responsibility for their learning, and I have also argued that the development of and reflection on metacognitive knowledge and beliefs about learning offer a useful framework. With this in mind, I would like to conclude by revisiting the characteristics of assessment for learning described above (DfES, 2004), but this time with a focus on pupils' understanding of *how* to learn rather than *what* to learn, in order to raise their awareness, knowledge and understanding of their own autonomy. The amendments appear in italics, and each characteristic is followed by examples of classroom practice:

- Sharing learning *process* objectives with pupils: encouraging pupils to consider objectives relating to person, task and strategic knowledge as well as linguistic objectives; for example, pointing out that they are developing listening strategies as well as listening to the content of a particular text.
- Helping pupils to know and recognise the *autonomous behaviours* they are aiming for: this includes discussions about aspects of metacognitive knowledge in the classroom using techniques drawn from the FGCs; for example, discussing what makes a good learner and negotiating a teacher-pupil contract (person knowledge), creating a poster about the reasons for language learning (task knowledge), discussing which task would be most appropriate to achieve a particular objective (task knowledge).
- Involving pupils in peer and self-assessment *of how they have learnt*: providing autonomy-related criteria to be used by pupils when assessing.
- Providing feedback that leads pupils to recognising their next steps and how to take them, *in order to become more autonomous*: organising mini plenary sessions throughout and at the end of each lesson to ask not only what pupils have learnt, but also how they have learnt it and how they might learn more effectively next time; for example, asking groups to feed back not only on the content of their groupwork, but also on how they organised themselves in order to complete it, and how they might do it differently next time.
- Promoting confidence that every pupil can improve *and become increasingly autonomous*: for example, offering pupils the opportunity to choose tasks, but discussing which task is most appropriate for their own individual objectives and levels.
- Involving both teacher and pupil in reviewing and reflecting on *assessment for autonomy* information: developing a variety of

evaluation formats to allow pupils to feed back to the teacher on lesson content, including ways in which they may have been more involved in the lesson.

Of course, there may also be opportunities to organise small FGCs in order to explore particular issues in greater depth. Teachers could occasionally arrange their lessons so that they, or someone else (classroom assistant, student teacher, older pupil, etc.) could facilitate FGCs. Eventually, if pupils became used to such talk in the early stages of learning, they may, through the use of scaffolding questions, be able to discuss without teacher facilitation. They may also be encouraged to reflect individually on ways in which they could improve their control over the learning process, by means of learning journals. What is crucial, however, is that teachers listen to the learners' voices, no matter how diverse, learn what is supporting or constraining their control over and responsibility for learning, and respond in ways that are sensitive to and appropriate for individual needs.

The strong relationship between learner autonomy and assessment for learning, and between learner autonomy and motivation (and learning) suggests that it is crucial to engage with the development of learner autonomy. In order to do so, this chapter has suggested a way of assessing fundamental elements of autonomy for formative purposes. There is clear justification for such assessment for autonomy, and clear evidence that effective assessment for autonomy activity is feasible.

Note

1. In the interest of anonymity, all names are pseudonyms.

References

Benson, P. (2001) *Teaching and Researching Autonomy in Language Learning.* Harlow: Pearson.

Benson, P. and Lor, W. (1999) Conceptions of language and language learning. *System* 27 (4), 459–472.

Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2002) *Working Inside the Black Box: Assessment for Learning in the Classroom.* London: King's College.

Black, P. and Jones, J. (2006) Formative assessment and the learning and teaching of MFL: Sharing the language learning road map with the learners. *Language Learning Journal* 34, 4–9.

Black, P. and Wiliam, D. (1998) Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice* 5 (1), 7–74.

Black, P. and Wiliam, D. (2005) Changing teaching through formative assessment: Research and practice. In Centre for Research and Innovation *Formative Assessment: Improving Learning in Secondary Classrooms* (pp. 223–240). Paris: OECD.

CERI (Centre for Educational Research and Innovation) (2005) *Formative Assessment: Improving Learning in Secondary Classrooms.* Paris: OECD.

CILT, ALL, ISMLA (2006) Language Trends KS4 Survey 2006. On WWW at http://www.cilt.org.uk/key/languagetrends/2006/trends2006_ks4_data.pdf.

Cotterall, S. (1999) Key variables in language learning: What do learners believe about them? *System* 27 (4), 493–513.

Deci, E.L. (1975) *Intrinsic Motivation*. New York: Plenum.

Deci, E.L. (1980) *The Psychology of Self-determination*. Lexington, MA: D.C. Heath.

Deci, E.L. and Ryan, R.M. (1985) *Intrinsic Motivation and Self-determination in Human Behavior*. New York: Plenum Press.

Deci, E.L. and Ryan, R.M. (1987) The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology* 53, 1024–1037.

Deci, E.L., Vallerand, R.J., Pelletier, L.G. and Ryan, R.M. (1991) Motivation and education: The self-determination perspective. *Educational Psychologist* 26 (3 and 4), 325–346.

Department for Education and Skills (2004) *Pedagogy and Practice: Teaching and Learning in Secondary Schools (Unit 12: Assessment for Learning)*. Norwich: HMSO.

Dörnyei, Z. (2001) *Teaching and Researching Motivation*. Harlow: Pearson Educational.

Findley, M.J. and Cooper, H.M. (1983) Locus of control and academic achievement: A literature review. *Journal of Personality and Social Psychology* 44 (2), 419–427.

Flavell, J.H. (1985) *Cognitive Development* (2nd edn). Englewood Cliffs, NJ: Prentice Hall.

Flavell, J.H., Miller, P.H. and Miller, S.A. (2002) *Cognitive Development* (4th edn). Upper Saddle River, NJ: Prentice Hall.

Garcia, T. and Pintrich, P.R. (1995) Assessing students' motivation and learning strategies: The motivated strategies for learning questionnaire. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. ERIC Document Reproduction Service No. ED 383 770.

Giroux, H.A. (1983) *Theory and Resistance in Education: A Pedagogy for the Opposition*. London: Heinemann.

Harris, V. and Noyau, G. (1990) Collaborative learning: Taking the first steps. In I. Gathercole (ed.) *Autonomy in Language Learning* (pp. 55–64). London: Centre for Information on Language Teaching.

Holstein, J.A. and Gubrium, J.F. (1995) *The Active Interview*. Thousand Oaks, CA: Sage Publications.

Jiménez Raya, M. (1998) Training language learners to learn. In W. Gewehr (ed.) *Aspects of Modern Language Teaching in Europe* (pp. 13–29). London: Routledge.

Krueger, R.A. (1994) *Focus Groups: A Practical Guide for Applied Research* (2nd edn). Thousand Oaks, CA: Sage Publications.

Lamb, T.E. (1996) Self-management strategies in the secondary school languages curriculum. In *Proceedings of the International Conference'AUTONOMY 2000': The Development of Learning Independence in Language Learning*, held at King Mongkut's Institute of Technology Thonburi, Bangkok, Thailand, in association with the British Council (pp. 101–115). Bangkok: KMITT.

Lamb, T.E. (1998) Now you are on your own! Developing independent language learning strategies. In W. Gewehr (ed.) *Aspects of Modern Language Teaching in Europe* (pp. 30–47). London: Routledge.

Lamb, T.E. (2000a) Reconceptualising disaffection – issues of power, voice and learner autonomy. In G. Walraven, C. Parsons, D. Van Veen and C. Day (eds) *Combating Social Exclusion through Education* (pp. 99–115). Louvain, Belgium and Apeldoorn, Netherlands: Garant.

Lamb, T.E. (2000b) Finding a voice – learner autonomy and teacher education in an urban context. In B. Sinclair, I. McGrath and T.E. Lamb (eds) *Learner Autonomy, Teacher Autonomy: Future Directions* (pp. 118–127). Harlow: Addison Wesley Longman.

Lamb, T.E. (2001) Metacognition and motivation – learning to learn. In G.N. Chambers (ed.) *Reflections on Motivation* (pp. 85–93). London: CILT.

Lamb, T.E. (2005) Listening to our learners' voices: Pupils' constructions of language learning in an urban school. PhD thesis, University of Nottingham.

Lamb, T.E. (2006a) Supporting independence; students' perceptions of self-management. In T.E. Lamb and H. Reinders (eds) *Supporting Independent Learning: Issues and Interventions* (pp. 97–124). Frankfurt-am-Main: Peter Lang.

Lamb, T.E. (2006b) Listening to learners' voices: On task knowledge in language learning. *Scottish Languages Review*. On WWW at http://www.scilt.stir.ac.uk/SLR/Current%20Issue/Index.htm.

Lamb, T.E. and Fisher, J. (1999) Making connections: Football, the Internet and reluctant language learners. *Language Learning Journal* 20, 32–36.

Lamb, T.E. and Reinders, H. (2005) Learner independence in language teaching: A concept of change. In D. Cunningham and A. Hatoss (eds) *An International Perspective on Language Policies, Practices and Proficiencies* (pp. 225–239). Belgrave, Australia: FIPLV.

LeCompte, M.D. and Preissle, J. (1993) *Ethnography and Qualitative Design in Educational Research* (2nd edn). London: Academic Press.

Lee, J., Buckland, D. and Shaw, G. (1998) *The Invisible Child*. London: CILT.

Looney, J. and Wiliam, D. (2005) England: Implementing formative assessment in a high stakes environment. In Centre for Research and Innovation *Formative Assessment: Improving Learning in Secondary Classrooms* (pp. 129–147). Paris: OECD.

Mertens, D. (1998) *Research Methods in Education and Psychology: Integrating Diversity with Quantitative and Qualitative Approaches.* Thousand Oaks, CA: Sage Publications.

Mirón, L.F. and Lauria, M. (1998) Student voice as agency: Resistance and accommodation in inner-city schools. *Anthropology and Education Quarterly* 29 (2), 189–213.

Mori, Y. (1999) Epistemological beliefs and language learning beliefs: What do language learners believe about their learning? *Language Learning* 49, 377–415.

O'Malley, J.M. and Chamot, A.U. (1990) *Learning Strategies in Second Language Acquisition*. New York: Cambridge University Press.

Oxford, R.L. (1990) *Language Learning Strategies: What Every Teacher Should Know.* Boston, MA: Heinle and Heinle.

Pintrich, P.R. and Schunk, D.H. (1996) *Motivation in Education: Theory, Research and Applications*. Englewood Cliffs, NJ: Prentice Hall.

Powney, J. and Watts, M. (1987) *Interviewing in Educational Research.* London: Routledge and Kegan Paul.

Rotter, J.B. (1966) Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs* 80 (1, whole no. 609).

Roy, D. (1991) Improving recall by eyewitnesses through the cognitive interview. *The Psychologist* 14 (9), 398–400.

Rubin, J. (1987) Learner strategies: Theoretical assumptions, research history and typology. In A. Wenden and J. Rubin (eds) *Learner Strategies in Language Learning* (pp. 15–30). London: Prentice Hall.

Rudduck, J., Chaplain, R. and Wallace, G. (eds) (1996) *School Improvement: What Can Pupils Tell Us?* London: David Fulton.

Rudduck, J., Day, J. and Wallace, G. (1997) Students' perspectives on school improvement. In A. Hargreaves (ed.) *Rethinking Educational Change with Heart and Mind* (pp. 73–91). Alexandria, VA: The Association for Supervision and Curriculum Development.

Sarup, M. (1991) *Education and the Ideologies of Racism*. Stoke-on-Trent: Trentham Books.

Simmons, D. (1996) A study of strategy in use in independent learners. In R. Pemberton, E.S.L. Li, W.W.F. Or and H. Pierson (eds) *Taking Control: Autonomy in Language Learning* (pp. 61–75). Hong Kong: Hong Kong University Press.

Spratt, M., Humphreys, G. and Chan, V. (2002) Autonomy and motivation: Which comes first? *Language Teaching Research*, 6 (3): 245–266.

Tomlinson, P. (1989) Having it both ways: Hierarchical focusing as research interview method. *British Educational Research Journal* 15 (2), 155–176.

Victori, M. (1999) Methodological issues in research on learners' beliefs about language learning. Paper presented at AILA 1999, the 12th World Congress on Applied Linguistics, Tokyo, Japan, 1–6 August 1999.

Victori, M. and Lockhart, W. (1995) Enhancing metacognition in self-directed language learning. *System* 23 (2), 223–234.

Weiner, B. (1984) Principles for a theory of student motivation and their application within an attributional framework. In R. Ames and C. Ames (eds) *Research on Motivation in Education, Vol. 1* (pp.15–38). Orlando, FL: Academic Press.

Wenden, A.L. (1996) Learner representations in language learning: Relevance and function. In *Proceedings of the International Conference'AUTONOMY 2000': The development of learning independence in language learning*, held at King Mongkut's Institute of Technology Thonburi, Bangkok, Thailand, in association with the British Council (pp. 234–253). Bangkok: KMITT.

Wenden, A.L. (1999) An introduction to metacognitive knowledge and beliefs in language learning: Beyond the basics. *System* 27 (4), 435–442.

Wenden, A.L. (2001) Metacognitive knowledge in SLA: The neglected variable. In M.P. Breen (ed.) *Learner Contributions to Language Learning* (pp. 44–64). Harlow: Pearson Educational.

Williams, M., Burden, R. and Lanvers, U. (2002) French is the language of love and stuff: Student perceptions of issues related to motivation in learning a foreign language. *British Educational Research Journal* 28 (4), 503–528.
Yang, N.D. (1999) The relationship between EFL learners' beliefs about language learning. *System* 27 (4), 515–535.

Chapter 7

Learners Reflecting on Learning: Evaluation versus *Testing in Autonomous Language Learning*

LENI DAM and LIENHARD LEGENHAUSEN

Introduction

It is clearly a paradox to claim that one can 'test the untestable' (cf. the title of this book). To make sense of such an undertaking, we need to examine the meaning of testing and distinguish a narrow technical sense of the word as defined by test theoreticians from a more general notion. Testing in the technical sense of the word requires that the criterion can be measured, i.e. quantified, so that scorer and test reliability can be calculated. In a more general sense, it simply implies that one tries to find out what somebody knows about a subject or what somebody can do in a particular field.

The purpose of evaluation in language education is also to gain knowledge, but without the restrictions imposed by the technical sense of testing. In other words, there is a functional overlap between testing and evaluation, but the way data are collected in evaluation is less restrictive than in technical approaches to testing. All types of data that lead to a more comprehensive picture of the behaviour under discussion qualify for the process of evaluation. As a rule, it is mainly qualitative data that is made use of; however, quantitative data are not at all excluded. Evaluations may thus also include 'test data'. The basic idea for many test theoreticians is that evaluation implies decision making. Bachman (1990: 22), for example, quotes a definition suggested by Weiss (1972): 'Evaluation can be defined as the systematic gathering of information for the purpose of making decisions'. It is especially apt in a pedagogical setting that tries to involve students in decision making. Any form of decision making presupposes the learners' ability to stand back and attain a cognitive distance from the processes and working procedures, which in turn presupposes a certain level of reflectivity and

critical thinking. This feature, which indirectly links evaluation with reflection, is, for example, highlighted in the *Oxford Advanced Learner's Dictionary*. The entry for the verb *to evaluate* reads: 'to form an opinion of the amount, value or quality of something after thinking about it'. In other words, evaluation implies that learners and teachers reflect on the experiences gained in language learning and teaching, which will lead to awareness raising and prepare the ground for decision making. This concept of evaluation also underlies the following definition:

> The function of evaluation is on the one hand to ensure that work undertaken is discussed and revised, and on the other to establish a basis of experience and awareness that can be used in planning further learning. (Dam, 1995: 49)

The interrelationship between reflective thinking on the one hand, and evaluative processes and decision making on the other, as prerequisites for learners to become more autonomous, is also discussed in Ridley (1997).

This chapter will outline areas in autonomous language learning that cannot be tested in the narrow technical sense of the word, but which need to be evaluated in order to optimise learning processes from the learners' as well as from the teacher's point of view. It will be argued that this type of evaluation is solely dependent on the learners' voices. We will present examples of these voices collected partly from questionnaires and partly from learners' logbooks used regularly in the autonomous classroom, in English classes in Danish comprehensive schools (see the section 'The data and the learners').

Since, however, major issues of testability relating to language education in general also constitute concerns in autonomous language learning, they will be outlined in the following section.

The Problematicity of Testing in Language Education

The history of testing in language pedagogy has shown that testing objectives that seem most worthwhile turn out to be recalcitrant to testing. This state of affairs has to do with the complexity of the constructs involved, on the one hand, and with the constraints that testing formats impose on the undertaking, on the other.

Key constructs in language education that are not easily testable include:

- the notion of *language proficiency* or *communicative competence*;
- conceptualisations of *learning processes* and the variables contributing to learning success.

Theoretical approaches to both areas are subject to controversial discussions. Communicative competence, for example, has been conceived of by some researchers as an indivisible construct that is largely determined by a global proficiency factor (cf. Oller, 1976, 1979). Others prefer analytical models that try to define variables constituting the construct and the interrelationships between them (cf. Bachman, 1990). These diverging views are then reflected in communicative testing – with integrative formats aligning with the former and discrete-point tests with the latter (cf. the discussions in Oller, 1983; Weir, 1990).

When it comes to learning processes, it is generally agreed that they involve an array of interrelated sub-processes – such as, for example, language learning as a problem-solving cognitive process, as a behavioural automatisation process or as a social process. This implies that these sub-processes interact differentially with the host of variables that constitute communicative competence. For example, the acquisition of grammatical patterns requires problem-solving skills, whereas the development of conversational fluency is largely dependent on automatisation processes. Social skills also come into play when developing strategic conversational competencies, and so forth.

Knowledge of these various sub-processes and their relationship to aspects of communicative competence would be a prerequisite for supporting the learners *systematically* in their learning and for devising forms of intervention in this process. Therefore, apart from just testing communicative competence – which is difficult enough – language educators also have an interest in assessing the learning processes that lead to certain outcomes.

This turns out to be an even greater challenge than testing communicative competence. The following list constitutes a partial list of variables influencing the learning processes:

- the teacher's role (e.g. her personality and teaching style);
- the learner's role (e.g. preferred learning styles, strategies and motivational attitudes);
- the teaching method and/or the activities enacted in the classroom;
- the media and materials used (e.g. the use of authentic texts, the role of audio-visual media and accessibility of computers and the internet);
- the impact of the social setting (e.g. the use of group work and peer tutoring);
- the various roles and forms of evaluation.

The number of variables involved and the intricate ways in which they interact make conventional 'objective' testing difficult, even impossible.

Evaluation rather than Testing in Autonomous Language Learning

All the problems outlined so far also come into play when considering 'testing' in autonomous language learning. In that context, though, a score of additional features that cannot be tested in the conventional way either come into play.

The construct of learner autonomy

There is widespread agreement that the learners' willingness and capacity to take over responsibility for their own learning process is a criterial feature of learner autonomy [cf. Holec (1981) and the Bergen definition, which was developed during the 5th Nordic conference on developing learner autonomy in Bergen in 1989 (see Dam, 1995: 1)]. The capacity to take over responsibility includes a capacity for critical *reflection*, which leads to an *awareness* of all relevant aspects of the learning/teaching undertaking. Reflection and awareness constitute some of the prerequisites for the learners' *involvement* in all the decisions that need to be made and for being able to take cognitive as well as pragmatic *control* of the procedures.

Of particular interest is the learners' *awareness* of

- their own linguistic competence;
- their acquisitional needs;
- preferred activities and ways of working;
- social aspects of learning.

The learners' capacity for *control* of procedures involves

- a capacity for *planning*, i.e. setting goals, suggesting and negotiating approaches to achieving goals, making decisions, setting realistic time frames etc.;
- a capacity for *organising working procedures*, i.e. finding appropriate materials and data, setting up working groups, allocating functions in their working group, managing time etc.;
- a capacity for *evaluating* procedures and products, such as the effectiveness of working procedures, the quality of interactions and the working atmosphere as well as the assessment of outcomes against the goals of official guidelines.

(See also the discussion of control by Benson, this volume.)

These central features of the construct 'autonomy' – or key variables forming prerequisites for autonomy – do not lend themselves to quantification either and thus cannot be measured as required by conventional testing procedures as indicated above. They are, so to speak, 'untestable'. This is where the concept of evaluation and the need to work with qualitative data from the learners come into play. Gaining *real* insight into the above-mentioned variables can only be achieved through the voices of the learners – their self-evaluations as well as peer evaluations.

The reliability and validity of the learners' evaluations

If we have to rely on the learners' evaluations, i.e. the way they reflect on their learning, comment on the procedures and self-evaluate their performance as well as their progress, then the question of the reliability and validity of the data suggests itself. Introspective data of this type have in the past also been referred to as 'soft' data (cf. Kohonen, 1988), in order to indicate that their subjective character might be less objective and reliable than data from carefully administered formal tests. Evidence has, however, accumulated in the past that the learners' self-evaluations can be as reliable and valid as the results of 'objective' tests, especially in contexts in which they have been systematically encouraged to monitor their learning over a longer period of time. This can be demonstrated by correlational studies carried out in 'testable' areas, in which, for example, the linguistic skills of learners have been under investigation. Dam and Legenhausen (1999) showed that the autonomous learners' self-evaluations as regards their reading and writing abilities correlated with the teacher's assessments and with 'objective' tests to a surprisingly high extent. Correlation coefficients between self-assessments and C-tests or teacher assessments ranged between $r = 0.74$ and $r = 0.82$, which means they were as high as the correlations between C-tests and teacher assessments. They thus corroborated earlier studies that have been reviewed by, for example, Oskarsson (1984, 1988).

The awareness of linguistic competence also includes the capacity to self-assess the level of achievement in terms of the official marking scheme. The data we have examined for this chapter show that learners are quite capable of supporting their assessments by convincing arguments. The examples below are taken from learners in their

4th year of English (first term). A below average learner who got the mark 7 wrote, for example:

> I think I'm going to have 6 or 7 because I not read very good. And I not write very well too. From now will I read more and so will I try to speek more english. I will listen to tape too because I can leand words with it, but it is up to you [= the teacher] what you will give me.

Another example comes from an above average learner, who wrote:

> I've got a 10 and I am very satisfied with it, but I also think it is a fair grade, because last time I had a 9, and I have improved since, especially in my articulation, but also in my vocabulary, which now is containing much more grown-up language. So the conclusion must be that is a good grade. (cf. also Dam, 2006: 277)

The data thus support the view that the learners' self-evaluations in 'testable' areas are reliable (as measured against teacher marks) and valid. This allows us to propose with confidence that the learners' voices are similarly valid in the domains that constitute the construct of autonomy as discussed above.

The Voices of the Learners

In this section, we will adduce evidence that the learners' reflections as regards the learning processes and procedures allow them to develop a high degree of awareness, which in turn is a prerequisite for making principled decisions, thereby taking control over their own learning.

These reflections are stimulated and developed via regular written evaluations in logbooks and questionnaires, as well as by an on-going oral negotiation process between teacher and learners or between the learners themselves.

The data and the learners

The 'evidence' will derive from two main sources of 'soft' data, namely, questionnaires and extracts from learners' logbooks. The specific data for this chapter were extracted from a larger body of data that was elicited over a span of 20 years – from 1982 to 2002 – from three different 'autonomous' classes all taught by Leni Dam. The school is a comprehensive school, a 'folkeskole', with mixed ability classes. The students' ages range between six (kindergarten class) and 16, which

is the ordinary school-leaving age after the tenth form. Students can then continue at A-level colleges or move on to other forms of further education. The school is located in Karlslunde, a suburban area 25 km south of Copenhagen. The area is considered to be middle- and working-class. In 1982, English education started in the fifth form at the average age of 11. The learners had four 45-minute lessons in grades 5 and 6. In grades 7–10, the number of lessons per week was reduced to three. In 1993, the students began learning English in grade 4 (at the age of 10) with only two lessons in the first school year. Marks are given three times a year from the eighth grade onward. In 2002, a compulsory oral examination was administered in the ninth grade, and an oral as well as written examination could be taken after the tenth grade. Since then, there have been a number of changes in the Danish school system. English has been taught from the third grade with two lessons a week since August 2008. New types of oral examinations as well as written examinations were introduced at the end of the ninth grade in summer 2007. Furthermore, computerised tests are being tried out in the seventh grade.

Up till 1984, only questionnaires were used for the learners' written evaluations of the learning process as well as of the outcome of the process. From 1984 onward, learners' logbooks took over the 'daily', on-going written evaluation, but were occasionally supplemented with questionnaires relating to areas of special interest for the teacher in conducting the teaching/learning process. In 2000, portfolios were introduced and used together with logbooks in class 3 (see Table 7.1). However, the portfolios contained only products collected at certain times. This means that the learners' 'daily' reflections included in this chapter derive from the logbooks.

The reflections in the learners' logbooks were initiated and guided by demands set up by the teacher (see Dam, 2000: 30–33). Learners' 'daily' entries into their logbooks were expected to include:

- day and date at the beginning of the lesson;
- a report on the activities taking place during a lesson, including personal outcomes of these activities such as new vocabulary, useful expressions, but also comments on and suggestions for distribution of labour when working in groups;
- decisions made as regards homework;
- reflections/evaluations concerning the day's work.

Logbook data and questionnaire data from this period often overlap, as learners frequently base their answers in the questionnaires on

Table 7.1 Overview of classes and data

Name of class	Years with English	Grades/age	Type and focus of data

Class 1	1982–1987	5th to 10th/11–16	Questionnaire data
			    Awareness of linguistic competence (Figure 7.1)
			    Awareness of linguistic competence and how to proceed (Figure 7.2)
			    Awareness of preferred/dispreferred activities (Figure 7.3)
			    Awareness of ways of working (Figure 7.4)

Class 2	1992–1997	5th to 10th/11–16	Logbook data[a]
			    Capacity for setting their own learning goals (Figure 7.5)
			    Capacity for forming groups (Figure 7.6)
			    Capacity for evaluating the effectiveness of learning processes (cf. the social context when learning) (Figure 7.7)
			    Capacity for peer-evaluation (Figures 7.8 and 7.9)

Class 3	1997–2002	4th to 9th/10–15	Questionnaire data Capacity for assessing personal linguistic outcomes and needs (Figure 7.10)

Note: [a] A detailed description of the learners' capacity for (the) control of procedures can be seen in Dam (2006). Some of the examples in this chapter are taken from there. However, the examples included here are the ones supposed to be 'untestable'.

reflections in their logbooks. We have divided the following sections according to the classes involved. The section on learners' awareness of variables involved when learning shows examples of evaluations from class 1; the section on learners' capacity for control of procedures shows examples from class 2; and the section on capacity for assessing personal linguistic outcomes and needs shows examples from class 3. Table 7.1 gives an overview of the data discussed in these sections.

Learners' awareness of variables involved when learning

In this section, we have included data concerned with the learners' *awareness* of variables involved when learning.

Awareness of linguistic competence

Figure 7.1 is taken from class 1 at the end of their first year of English. It shows the 11-year-olds' awareness of their linguistic competence at this point.

This example highlights a number of points. It shows awareness of linguistic competence both in specific terms (e.g. counting from zero to one thousand) and in general terms (e.g. learning many words). It also includes evaluative comments (e.g. an evaluation of progress in pronunciation, or feeling that what has been achieved is 'not enough').

Martin's comments are remarkably varied. He can relate the acquisition of difficult words to a specific activity, and draws self-confidence from the fact that he can read all the books available in class. He has started to think in terms of a compensatory strategy when lacking a word in English – a feature of learner language that is also commented upon in much more detail at a more advanced learning stage by Lene (cf. Figure 7.2).

'What do you feel you have learned this year?'
Lene: – 'I think I have learn much words and sentences and count from zero to tousind. ...'

Martin: – 'I have learn:
– of understanding composing
– when I not has English can I just lighten [= 'use easier'] words
– reading English, write English, and speak English
– I have too learn strong [= 'difficult'] words of play 'Meet the Family' f.eks. office, heavy, parcel, clothe, yourself
– I can reading all them books we have in the class
– I can now outspeak [= 'pronounce'] English words much better'

The following statements were written in the learners' Danish mother tongue:
Jeanette: – 'I have learned to act'

Tina: – 'sometimes I talk to my friends in English'

Peter: – 'I have learned to understand some English but not enough'

Figure 7.1 Evaluation after one year of English – awareness of linguistic competence

Awareness of linguistic competence and how to proceed

Whereas the teacher's question in the class of beginners is fairly general, and thus also provokes general answers, two and a half years later, the teacher's questions as well as the answers are much more elaborate. Figure 7.2 makes it obvious that the teacher's more specific questions

---

'*How did you cope in England with your English?*
When did you cope well, and when poorly?
How are you going "to use" this experience in your English lessons?'

Lene:
I coped well: 'When I was talking to people, who I knew, I did well. I didn't meet any problems, when I was shopping. I found it easier to talk to Jackie and the younger brother, than to talk to Lisa'.
I coped poorly: 'When they were telling jokes, I didn't understand the point. If somebody, who I didn't know, addressed me, I found it difficult to catch the questions. If they spoke to fast or used words, which I didn't understand, I didn't always understand it if I was about to tell something, I sometimes meet a word, which I couldn't explain'.

How am I going to use this experience in my English lessons? 'We are going to explain an english word (which we don't know) in other english words. I'm going to learn something about english humor. Read some jokes'.

Jeanette:
I coped well: 'I coped very good with my english when I was talking with my freinds. In the shops when I were asking things I also cope well. I think that I coped very well on the hole trip'.
I coped poorly: 'The first day I mumbled a lot. And when I was explaining things and experiences to Valerie Parry and Julie Tuz. I think it was because I thought that if I talked slowly they wouldn't listen to me. I was talking very fast and it was silly of me.
How am I going to use this experience in my English lessons? In the next lessons I will try to talk with somebody and when I talk I will try to open my mouth a bit more and talk a bit slower. I will worke together with Gitte. First we will record a discussion about the things we have been doing for the weekend'.

Figure 7.2 Evaluation after two and a half years of English – awareness of linguistic competence and how to proceed

after a two-week trip to England push learners to be more precise and get them to think in terms of different communicative situations.

The example shows how the eighth graders have learned to distinguish different communicative situations in which variables such as familiarity with the speaker, topic of discussion, speed of delivery etc., play a crucial role for them. They have experienced the lack of words in certain situations and devise strategies for coping with linguistic deficits in the future by practising paraphrasing techniques and by doing background reading in specific topic areas (e.g. English humour). Jeanette realises that her strategy of making her conversational contributions more acceptable by speaking very fast did not work, which shows a remarkable depth of insight, reflectivity and the ability to evaluate her experiences. Statements like these provide evidence that learners develop a sense of their acquisitional needs by constantly being forced to reflect on and thus evaluate their linguistic experiences.

Awareness of preferred and dispreferred activities and ways of working

The answers to questions referring to the learners' preferred and dispreferred activities bring to light an awareness that peer-tutoring is also beneficial to the learner who is doing the tutoring. They furthermore stress the importance of the products that result from completing an activity, which has recently been established as a criterial feature in task-based learning (cf. Ellis, 2003). Dispreferred activities seem to derive from a sense of not knowing how to start project work and how to go about it (Figure 7.3).

The 'freedom' in autonomous classrooms is in reality fairly constrained by the curricular guidelines, which limit the possibilities of choice (cf. Dam, 1995). However, Figure 7.4 shows that the learners obviously have a feeling that they can 'work with what we want to work with'. This potential for self-determination has without doubt an enormous impact on their motivation.

Learners' capacity for control of procedures

In this section, the examples included focus on the learners' *capacity* for control of procedures, which include planning, organising working procedures and evaluating procedures and products (see the section 'The construct of learner autonomy'). The examples in this section all derive from class 2.

Capacity for setting their own learning goals

The following logbook entries from grade 8, fourth year of English, refer to goal setting following the evaluation of a previous working

'*Which activity/activities did you like best? Why?*'

Lene: 'Being a teacher, because I think I learned a lot by helping others in the class. And it was fun. I like writing poems. You can write about your own feelings'.

Martin: 'I liked the newspaper subject because when you finished, there was a result to look at. And with that subject we worked very well together'.

Which activity/activities did you not like? Why?

Jeanette: 'When me and Karin and Gitte decided to do something about English schools even though you have lots of materials you don't know how to start and what to write'.

Soren: 'I didn't like ''Animal farm'' maybe because we didn't really get started'.

Figure 7.3 Fourth year of English – awareness of preferred/dispreferred activities

'*How did you like the way you have worked this year? Positive/negative things?*'

Martin: 'I think it's great that we may decide ourselves what to do. It's more exciting to choose and work with what we want to work with. I think you wouldn't work that concentrated if it's Leni who decide what to do'.

Figure 7.4 End-of-year evaluation after four years of English – awareness of ways of working

module. The three learners are at different levels of their linguistic development: Karsten is an advanced learner, Max is an average learner and Susan is a weak learner (Figure 7.5).

Karsten's goals are skill oriented, very specific and closely related to the official curricular guidelines, which were distributed to the class at the beginning of term. Due to his own previous evaluation of oral and written presentations, he has realised that his pronunciation and spelling ability need improvement, and he has a very concrete idea of how to go about it. Max's goals are also closely related to the official guidelines. In

> Karsten:
> My personal contract for April I will read aloud from my book when I am sharing homework to practice my articulation. I will write some stories as homework, to practice my spelling and written language.
>
> Max:
> My personal contract for April: 1. I will translate a page from my book every time when I read in my book. 2. I will read a whole book a month. 3. I will write one story a month.
>
> Susan:
> My Personal contract. how am I going to Be Better: talk to my tape recorter read aloud. read more I am going to get Better at spelling write more English.

Figure 7.5 Fourth year of English – example of individual goal setting (cf. also Dam, 2006: 270)

contrast to Karsten, though, Max takes a more global approach to improving his English. For example, he does not specify what aims he pursues with his translation practice. On the other hand, he sets himself a very ambitious and rigorous time framework for his activities. Entries in Max's logbook the following weeks show that he was capable of working to his self-set tight schedule – a phenomenon often experienced with linguistically average learners in the autonomous classroom. Susan's plans – apart from intending to read and write more – include the use of a tape recorder (something she has 'learned' from Karsten).

Capacity for forming groups
   Another important issue when being in control of your own learning is the capacity for forming groups. The students were asked to consider the following question: *Who would you like to work with and why?*
   The examples in Figure 7.6 show the learners' well-developed awareness of reasons for forming groups. Furthermore, they illustrate an important working principle of the autonomous classroom: group work is expected to support individual goals while working on a common product.
   After realising that previous group work results did not live up to their expectation, the learners might deliberately choose the same classmates for a new project, in order, as it were, to make up for the disappointing earlier outcome. In Karsten's case, the relative failure seems to have stirred his ambition. He obviously hopes to make up for

Karsten writes:
I would like to work with: Lasse, Michael, Lars. Because we had a blast last time, but the product wasn't as good as it could be, and we would like to improve that. I would like to make a videoprogramme. Because I have never done it before.

Max writes:
Who do I want to work with? Jan. Why? Because I can help him spelling correctly and he can help me talking fluently. What do I want to work with? Stories and some games. Why? Because then I can learn some more and better English language and I can get a nicer handwriting.

Susan writes:
I would love to continue working together with Nanna, because I think we both need to talk a lot, and Nanna can help me spelling.

Figure 7.6 Fourth year of English – learners' capacity for forming groups (cf. also Dam, 2006: 272)

this by challenging himself as well as the others by choosing a new and unknown type of activity.

The formation of groups might also be triggered by the insight that other learners share the same needs as in Max's case, or that a classmate might be especially suited for a particular support. This, of course, presupposes some knowledge or familiarity with the partner's weaknesses, which is obviously the case here – an awareness gained from the constant reflection on the social aspect of learning in the autonomous classroom. Susan's wish to work together with Nanna has no doubt been provoked by the insight that they, on the one hand, share a common need, and that they would thus both profit from cooperation. On the other hand, she feels Nanna is the right person to support her with her spelling difficulties.

Capacity for evaluating the effectiveness of working procedures

After having formed groups (cf. Figure 7.6), the learners were asked to comment on their new groups at the end of the lesson (Figure 7.7).

Karsten makes the point that a working relationship can profit from good interpersonal relations. For Max, the pair work forms the 'give-and-take' situation he had expected when choosing Jan for his partner.

Karsten enters:
Comments: Michael and I is good friends and therefore good partners in English. I think its a way to work, by choosing a couple of things you wish to be better at, and practise them with your sidekick.

Max expresses it like this:
Comments on todays work: It was good to have a new partner like Jan because then I have the chance to learn some more English and the chance to help Jan with his spelling.

Figure 7.7 Fourth year of English – capacity for evaluating the effectiveness of group work

Capacity for peer-evaluation

   After having been away from school for a week in grade 8, in order to try out different jobs as part of work experience, the learners in class 2 were asked to present their experience orally to the rest of the class. Their peers were asked to evaluate the presentation in their logbooks and at the same time give advice for improvements. As can be seen from the examples in Figure 7.8, the teacher couldn't have done it any better.

   Like all the other capacities mentioned and exemplified, developing the capacity for peer-evaluation is a long process that has to be nurtured and systematically supported. Furthermore, peer-evaluation presupposes a learning environment where the learners feel secure.

   The example in Figure 7.9 is taken from the second year of English in class 2. At the end of a lesson, a group of boys, including Dennis (a menace) – a very weak and difficult learner (cf. Dam, 1999) – presented their play. It was the first time that Dennis tried to be 'on stage'. The peer (Nanna) nicely captures this development in her evaluation.

Capacity for assessing personal linguistic outcomes and needs

   The final example in Figure 7.10 is taken from class 3 and is based on questionnaire data collected towards the end of the learners' fifth year of English.

   After five years of English, there is now more focus on grammar and form, and the average and above average students are fairly specific as

Evaluation of Lasse – a very weak learner:
Karsten:
he swoped his words, and talked a little Danish in between, but it was a long interview.

Max:
Lasse had worked in a place with some newspapers. I could hear that. I couldn't hear it all. and then he started to talk Danish. He has to practice his English language, and English pronunciation.

Evaluation of Michael – a slightly below average learner:
Karsten says:
Some good words, almost fluent, could be a little longer.

Max says:
It was good because I could understand what he said. It wasn't all good, because he didn't speak English all the time. And he has to practice to talk fluent English.

Figure 7.8 Fourth year of English (second half) – peer-evaluation of oral presentations

'It was a very good play because they have practised very much. Dennis was very good to play Dennis menase. I have never heard him say so much. They played very well'.

Figure 7.9 Second year of English – peer-evaluation of a play called Dennis the Menace

regards suggestions for improvements. This development might be due to several causes. On the one hand, there are the curricular guidelines that put more emphasis on language awareness and metalinguistic knowledge. On the other hand, the teacher's growing awareness of how to ask more precise questions might also have had an impact. A third reason might have to do with the fact that the use of portfolios was introduced in this class in grade 7, i.e. one year earlier. Among the products in the portfolio, learners keep self-selected samples of work, e.g. examples of essays and tape-recordings of oral productions according to their own goals for the period (cf. the learners' capacity

for goal setting). When selecting products for their portfolios, the learners' awareness as to whether or not they have reached their goals is sharpened.

Please consider the following:

1. Things I've improved
2. Things I want to improve further

Andre (new student in class)

1. I've become a lot better since last year at my old school because I didn't say anything. I was afraid of my teacher.
2. That I can do a good job if I just want to or if it is of interest to me.

Michelle (average student)

1. I've become better at spelling, but I still need to work a lot with it; I now write with a better language because I use my good expressions, and I have improved my behaviour, but I still need to work with 'becoming down to earth'.
2. Spelling, grammar, a better language, use my good expressions, my behaviour. I will give myself more time to study at home, keep my things tidy at school, I will try to use good expressions and the good starters and read aloud for my mother once a week.

Karina (average)

1. I think I have improved writing essays, because when I looked at the old essay I found out that there were more good expressions in the new one. I do also think I have improved my grammar, because when I look at some of the old texts I saw that that there were more is- and are-mistakes.
2. I still want to improve my writing and get a bigger vocabulary, and also write more good expressions when I write essays and other texts. I also want to be better at reading faster and reading aloud.

Anette (above average)

1. I've got a larger vocabulary. I read a lot more now. I use good starters a lot more now. Grammar.
2. Pronunciation. Grammar. Reading aloud. I will not write 'so' and 'but' all the time. Make more projects that involves pronunciation. Read aloud from my book to my parents.

Figure 7.10 Fifth year of English – assessment of linguistic competence

Concluding Remarks

The examples in this chapter are intended to show that when it comes to 'testing the untestable' in autonomous language learning, the obvious and only answer for teachers to get a real insight into the learning processes is constant and recurring evaluations where the learners' voices are heard and taken seriously. This insight provides the teacher not only with invaluable clues for supporting the individual learner systematically, but also with a possibility of reflecting on her own teaching approach.

As regards the question of evaluation *versus* testing in autonomous language learning, tests can and should of course be applied to areas that lend themselves to testing, especially in cases where learners find them helpful for assessing their own developmental stage and for ascertaining their feeling of progress (cf. Figure 7.1 and Figure 7.10).

Apart from providing valuable insights into learning/teaching processes for learners as well as teachers, evaluations also have an important influence on the learners' communicative competence. The constant demands on the learners to reflect on whatever they are doing and to evaluate past experiences as well as outcomes – as early and as often as possible in the target language – no doubt also develops a heightened *language* awareness (cf. also Little, 1997, 1999, 2006). These reflections and evaluations furthermore form *authentic* topics for communicative interactions, which are generally held to be the 'crucible' (Breen & Littlejohn, 2000) in which language learning takes place. Linguistic outcomes of an autonomous classroom (cf. class 2 above) have been documented in the LAALE project (Language Acquisition in Autonomous Learning Environment; cf. Dam & Legenhausen, 1996; Legenhausen 1999, 2001, 2003).

What has not been mentioned so far, though, in connection with evaluations is that one of the essential pre-conditions for this undertaking is the learners' self-esteem. It takes an assertive attitude to be able to make evaluative reflections 'public'. However, self-esteem is not only the basis for evaluative interactions. It is also a prerequisite in the process of developing learner autonomy. A cyclical relationship can be observed here, since a heightened self-esteem will result from the learners' independent actions.

The following learner's voice (Nanna), taken from her end-of-year evaluation in her final year at school (fifth year of English), emphasises this interrelationship between self-assertiveness and self-esteem, on the one hand, and autonomy, on the other. It furthermore adds yet another

example to the number of untestable issues in autonomous language learning:

> I have also via English learned to start a conversation with a stranger and ask good questions. And I think that our ''together'' session has helped me to become better at listening to other people and to be interested in them. I feel that I have learned to believe in myself and to be independent.

References

Bachman, L. (1990) *Fundamental Considerations in Language Testing.* Oxford: Oxford University Press.

Breen, M.P. and Littlejohn, A. (eds) (2000) *Classroom-Decision Making. Negotiation and Process Syllabuses in Practice*. Cambridge: Cambridge University Press.

Dam, L. (1995) *Learner Autonomy 3: From Theory to Classroom Practice*. Dublin: Authentik.

Dam, L. (1999) Dennis the Menace – and autonomy. In B. Mißler and U. Multhaup (eds) *The Construction of Knowledge, Learner Autonomy and Related Issues in Foreign Language Learning* (pp. 13–26). Tübingen: Stauffenberg.

Dam, L. (2000) Why focus on learning rather than teaching? From theory to practice. In D. Little, L. Dam and J. Timmer (eds) *Focus on Learning Rather than Teaching: Why and How?* (pp. 18–37). Dublin: CLCS, Trinity College.

Dam, L. (2006) Developing learner autonomy – looking into learners' logbooks. In M. Kötter, O. Traxel and St. Gabel (eds) *Investigating and Facilitating Language Learning* (pp. 265–282). Trier: Wissenschaftlicher Verlag Trier.

Dam, L. and Legenhausen, L. (1996) The acquisition of vocabulary in an autonomous language learning environment – the first months of beginning English. In R. Pemberton, E.S.L. Li, W.W.F. Orr and H.D. Pierson (eds) *Taking Control – Autonomy in Language Learning* (pp. 265–280). Hong Kong: Hong Kong University Press.

Dam, L and Legenhausen, L. (1999) Language acquisition in an autonomous learning environment: External and self-assessment compared. In D. Crabbe and S. Cotterall (eds) *Learner Autonomy in Language Learning: Defining the Field and Effecting Change* (pp. 89–98). Frankfurt/M.: Lang.

Ellis, R. (2003) *Task-based Language Learning and Teaching*. Oxford: Oxford University Press.

Holec, H. (1981) *Autonomy and Foreign Language Learning.* Oxford: Pergamon.

Kohonen, V. (1988) Evaluation in relation to communicative language teaching. In *Evaluation and Testing in the Learning and Teaching of Languages for Communication. Project No. 12* (pp. 11–26). Strasbourg: Council of Europe.

Legenhausen, L. (1999) The emergence and use of grammatical structures in conversational interactions – Comparing traditional and autonomous learners. In B. Mißler and U. Multhaup (eds) *The Construction of Knowledge, Learner Autonomy and Related Issues in Foreign Language Learning* (pp. 27–40). Tübingen: Stauffenberg.

Legenhausen, L. (2001) Discourse behaviour in an autonomous learning environment. *AILA Review* 15, 65–69.

Legenhausen, L. (2003) Second language acquisition in an autonomous learning environment. In D. Little, J. Ridley and E. Ushioda (eds) *Learner Autonomy in the Foreign Language Classroom – Teacher, Learner, Curriculum and Assessment* (pp. 65–77). Dublin: Authentik.

Little, D. (1997) Language awareness and the autonomous language learner. *Language Awareness* 6, 93–104.

Little, D. (1999) Metalinguistic awareness: The cornerstone of learner autonomy. In B. Mißler and U. Multhaup (eds) *The Construction of Knowledge, Learner Autonomy and Related Issues in Foreign Language Learning* (pp. 3–12). Tübingen: Stauffenberg.

Little, D. (2006) Learner autonomy and the language of reflection: The importance of inner speech. In M. Kötter, O. Traxel and St. Gabel (eds) *Investigating and Facilitating Language Learning* (pp. 255–263). Trier: Wissenschaftlicher Verlag Trier.

Oller, J.W. Jr. (1976) Evidence of a general language proficiency factor: An expectancy grammar. *Die Neueren Sprachen* 75, 165–174.

Oller, J.W. Jr. (1979) *Language Tests at School*. London: Longman.

Oller, J.W. Jr. (ed.) (1983) *Issues in Language Teaching Research.* Rowley, MA: Newbury House.

Oskarsson, M. (1984) *Self-assessment of Foreign Language Skills: A Survey of Research and Development Work.* Strasbourg: Council of Europe.

Oskarsson, M. (1988) Self-assessment of communicative proficiency. In *Evaluation and Testing in the Learning and Teaching of Languages for Communication. Project No. 12* (pp. 46–58). Strasbourg: Council of Europe.

Ridley, J. (1997) *Developing Learners' Thinking Skills*. Dublin: Authentik.

Weir, C.J. (1990) *Communicative Language Testing*. New York: Prentice Hall.

Weiss, C.H. (1972) *Evaluation Research: Methods for Assessing Program Effectiveness*. Englewood Cliffs, NJ: Prentice-Hall.

Part 3
Literature

Between Scylla and Charybdis: The Dilemmas of Testing Language and Literature[1]

AMOS PARAN

Introduction

In the *Odyssey*, Homer tells of the time when Odysseus had to navigate carefully in order to avoid the monster, Scylla, who had six heads on six long necks, and the whirlpools of Charybdis, which threatened to engulf his ship. Weir and Porter (1994), in a paper entitled 'The multi-divisible or unitary nature of reading: The language tester between Scylla and Charybdis', remind us that testing also forces choices (though possibly not always as life threatening as this). In this chapter, I suggest that teachers who teach literature within second language learning, too, are forced to make uneasy choices, and I present these choices in the form of six dilemmas. After articulating these dilemmas, I use them to draw a number of principles for test construction, and then present a list of possible tasks that can form the basis for choice for test items. In this, I build on previous work in this area, particularly Carter and Long (1990), Spiro (1991) and Parkinson and Reid Thomas (2000), all of whom have offered typologies of items and questions for testing literature in foreign and second language learning.

One issue that raises its head at this point is the issue of context. My main experience is in using and teaching literature within secondary school contexts and that is where my main interests lie. However, much of the discussion of literature in language education within the published research literature refers to university contexts (see Paran, 2008). One important task for the future is to disambiguate between these different contexts; in this chapter, I discuss the various contexts together, attempting to differentiate between them when possible.

Another issue is one of terminology and of approach. As I point out below, there is a split between teachers who wish to teach literature and

literary skills, and teachers whose aim is to use literature in the foreign language (FL) classroom for language development purposes. Thus *teaching* literature and *using* literature are not the same – and there are many positions on the cline between the two. Referring to 'teaching and using literature' throughout would be cumbersome; I therefore use 'teaching literature' as shorthand for activities that encompass both types of approach to some extent. Where a distinction between the two is important and only one of them is referred to, this is clarified in the discussion.

Six Dilemmas for Testing Literature in Foreign Language Teaching

Dilemma 1: To test or not to test?

This dilemma goes back to the core of the issues discussed in this book, namely, the conflict between teaching and testing (see Chapter 1 for an extended discussion of this issue). The conflict is even bigger where literature is concerned, because the two activities, testing on the one hand and teaching literature on the other, have two very different goals. Testing is an external activity with an external goal, often some sort of gate-keeping – for example, entrance to university, placement upon entrance into a language course, progress to the next level of a course. In secondary school settings (which are the settings where literature is most likely to be taught and used), it is not only the school leaving examination that is high stakes (as Alderson, 2004 acknowledges): in many such settings, a test can decide on the examination for which pupils may be allowed to study, or whether they will be allowed to progress to the next level of schooling or be required to repeat a grade.

Teaching literature, on the other hand, has as its aim a cluster of internal goals – the development of the individual in affective and intellectual terms; personal growth; developing private appreciation of literary works (thus leading to growth and development beyond the classroom). As Cook (1996: 152–153) says, 'education (indeed all communication) is an attempt to change others – to interfere with them... The issue is not one of *whether* teachers should inculcate value judgements into their pupils, but of *which* values these should be'. It is therefore clear that it is impossible to dissociate a discussion of literature in FL teaching from its educational context, from values and from the communication of values to learners. Thus, Carter and Long (1990: 217) see the potential of literature to 'promote greater understanding and knowledge of human behaviour'. Brumfit (2001: 91–92) suggests that

'the teaching of literature can thus be seen as a means of introducing learners to... a serious view of our world, of initiating them in the process of defining themselves through contact with others' experience'. Parkinson and Reid Thomas (2000: 143) rightly suggest that 'literature is inherently affective in a way which perhaps applies to no other subject', and speak of its ability to lead to 'an enhanced awareness and understanding of one's own emotional life'.

However, this view of teaching literature in FL education is not always apparent in discussions within applied linguistics contexts. Support for teaching and using literature in FL teaching is often based on linguistic argumentation, and the arguments against also tend to focus on language. For example, Edmondson's (1997) suggestion that literature has no special claim within FL teaching is backed by purely linguistic points. This line of argumentation is at best partial, and at worst invalid, as it does not in any way take into account the vision of what the goals of FL teaching must be in an educational context. Commenting on Edmondson (1997), I have suggested that 'language teaching, and the part which literature has to play in it, have to be seen as part of the whole educational endeavour, and not apart from it; this relationship is thus a major concern of the literature/language interface' (Paran, 2000: 76; see also Paran, 2008).

Once we introduce testing into the equation, it becomes apparent that the values implicit in the act of teaching literature on the one hand, and those values implicitly inculcated into our pupils through the examination system on the other, are in fact at odds with one another. Spiro (1991: 27) discusses what she calls the 'split' between the aims of teaching and the aims of testing, such as the attitude to errors (in testing situations, errors are penalised, whereas in teaching, they are seen as contributing to learning) or the attitude to interaction between students (encouraged in learning, forbidden and penalised in testing). This discrepancy between the goals of testing and the use of literature in FL teaching gives rise to the first dilemma – should we really need to test literature within a FL context?

In an ideal world, the answer to this would be a clear 'no'. Literature would have its place in the language classroom, it would serve its purpose in terms of the learners' personal growth; teachers and students would engage with literature, and it would achieve its role within language learning or language acquisition through continuous engagement and discussion. There would be no need to test it; what would be tested would be only the language learning that has resulted from engagement with the literature.

However, anyone familiar with educational contexts, and in particular secondary school settings, will realise that this is not very likely to happen. Using literature in the classroom has little face validity, especially for the learner, and the power of tests 'to manipulate . . . educational systems, to control curricula and to impose (or promote) new textbooks and new teaching methods' (Cheng & Curtis, 2004: 6) is therefore particularly pertinent here. If the teaching of literature is not linked quite clearly to assessment, it might simply disappear, as Brumfit (2001: 93) suggests: 'if we want literature to be taught, in most societies, we have to put up with it being tested'. Indeed, one might argue that the main advantage of including literature in FL tests is the washback effect, which means that much of what we do in the classroom is the result of what is being tested or our perceptions of what is being tested (see Alderson & Banerjee, 2001; Cheng *et al.*, 2004; Rea-Dickins & Scott, 2007 and the special issue of the journal *Assessment in Education* to which it is an introduction). Spiro (1991: 29) valiantly suggests that the way in which testing and teaching are interconnected should not mean 'that classroom procedures should be modelled by the test; but rather that test procedures should be reshaped by the strategies and goals of the classroom'. Unfortunately, what happens in reality is that the test is an all-powerful influence not just in shaping, but also in dictating the classroom behaviour of teachers. An example of how this happens where literature is concerned is provided by Ferman (2004), whose respondents explained how they chose shorter and easier texts in order to help their learners succeed on the English as a foreign language (EFL) examination (see also Shohamy, 2007, for a variety of other examples).

Dilemma 2: Testing language or testing literature?

The second major aspect that we need to take into account arises from the issue presented in the introduction to this chapter, namely, that the language/literature interface constitutes a complex relationship, within which there are various responses and educational practices. At one end, there are practitioners who use literature only as a vehicle for language, possibly as part of an extensive reading programme, or use extracts to teach language through them. At the other extreme are university courses where the aim is to study literature, either with the aim of building up literary knowledge or building up literary skills. Introducing testing into this situation further complicates the issue, as the question arises which of the two different competences – language competence or literary

competence – we are trying to assess, and how the two fit together (see Spiro, 1991 for an extensive discussion of literary competences).

One attempt to bring the two together is Bachman (1990), who presents an oft-quoted model of language competence, which has been 'increasingly influential' (Alderson, 2004: 13). In this model, Bachman includes what he calls 'imaginative functions'. This is defined as the function that

> enables us to create or extend our own environment for humorous or aesthetic purposes, where the value derives from the way in which the language is used. Examples are telling jokes, constructing and communicating fantasies, creating metaphors or other figurative uses of language, as well as attending plays or films and reading works such as novels, short stories or poetry for enjoyment. (Bachman, 1990: 94)

However, Bachman does not go on to discuss how the imaginative function of language can have a role in testing.

The Common European Framework of Reference for Languages (Council for Cultural Cooperation, 2001; henceforth CEFR), seems to accord some importance to aesthetic uses of language, stating that 'imaginative and artistic uses of language are important both educationally and in their own right' (CEFR, 2001: 56); users of the framework are then invited 'to consider and where appropriate state which ludic and aesthetic uses of language the learners will need/be equipped/be required to make'. In the illustrative scales provided throughout the CEFR, such uses in fact make fairly frequent appearances. They appear in oral production (e.g. relating plot of books at B1; CEFR, 2001: 59); they appear in reading comprehension (e.g. reading 'highly colloquial literary and non-literary writings' at level C2; CEFR, 2001: 69) and, interestingly, they appear in an illustrative scale devoted to creative writing (CEFR, 2001: 62). Having said that, there is a sense in which literary texts are mentioned explicitly in the higher levels only, and are possibly subsumed under more general mentions of texts at the lower levels. In reading comprehension scales, literary texts make an explicit appearance only at C2, in the self-assessment grid (CEFR, 2001: 27) literary texts are mentioned at levels B2 and above.

The question still remains, however, how literary uses of language can be brought into a testing situation, and what happens, in fact, is that in many contexts we end up employing at least two criteria when we are grading or evaluating any sample that we have as the result of our literature test, a literary criterion and a language one. On the face of it,

this is unproblematic, and merely echoes any situation in which performance is measured on more than one set of criteria, with each set receiving a differential weighting (though see Low, this volume, for some of the issues that can arise in such situations). This is common practice for the many language examinations where there is a general marking scheme and a task-specific marking scheme. One could claim that in the case of literature and language, we are doing the same – we are assessing literary knowledge or skills and at the same time assessing language performance. But the situation is in fact more complex than that. When task achievement is included in the criteria for assessment of a letter of complaint, for example, this is because we have a particular view of communication and of the way language is used to achieve a goal within this communicative context. When we are testing literature the situation is different – what is being assessed are two competences that are fairly distinct (even taking into account the role of language awareness and linguistic competence in literary awareness; indeed, this may be one of the reasons why the imaginative function does not feature prominently in Bachman's discussion of testing). It is therefore important to be very clear about which competence we are tapping, and which aspect of performance in the test we are going to mark.

The discussion above clearly refers to cases that involve production of a sample of language. Parkinson and Reid Thomas provide examples of item types that do not require the examinee to do this, but only require reading and recognition skills – the example they provide is: 'Underline sentences in a story which show the type of narrator, and use a letter key (supplied) to show which type' (Parkinson & Reid Thomas, 2000: 151). However, this type of test is in the minority, and all the other types of test they mention do require production. (It is also worth considering the washback effect that this type of test would have.)

Another solution is to decide not to examine one of the two main competences involved. So, for example, a group of tutors who worked on producing a literature test decided that 'linguistic accuracy will not be marked, nor will inaccuracy be penalised unless the answer is incomprehensible' (Spiro, 1991: 76). An opposite response to this dilemma is the approach taken by Cambridge ESOL, where literary knowledge is not assessed; what is assessed is language. At First Certificate in English (FCE) level, for example, assessment of answers to the questions on the set text 'is based... on control of language in the given context' (University of Cambridge, n.d.: 17). At Certificate of Proficiency in English (CPE) level, 'credit will be given for task achievement and

language competence; candidates are not expected to demonstrate skill in literary analysis' (University of Cambridge, 2008: 22). The CPE task description says that, 'candidates are only assessed on their language output and the relevance of their content. They must have read the book, know it well, and be able to respond appropriately to the prompts in the question' (www.cambridgeesol.org/teach/cpe/writing/aboutthepaper/part2/index.htm). Knowledge of the set text is thus a necessary but not sufficient condition for passing the test.

Testers will therefore need to make a clear decision which of the two competences is being tested, what the relationship between them is, and, in cases where both are being tested, what the weightings should be.

Dilemma No 3: Testing knowledge or testing skills?

The next dilemma concerns what we want to teach – and test – within the literary field. Do we teach and test knowledge *of* and *about* literature, or are we interested in the development of literary competence and what we are testing are the skills that are involved in literary competence? This is very strongly connected to what we think we are teaching when we teach – or use – literature in the classroom.

To frame the discussion, I shall refer to the examples provided by Carter and Long (1990), who define three types of questions commonly used. In the first type, which they call 'paraphrase and context', students are given an extract, and are asked to identify it and comment on it (both in terms of plot and character and in terms of literary devices used). They are provided with a list of very specific guiding questions to do this. The second type is what Carter and Long (1990) call 'describe and discuss' – for example, 'Describe Snowball and explain what happens to him'. The third type is what they call 'evaluate and discuss', for which their example is 'Illustrate from the stories how Lawrence's attitude to his characters is often a mixture of ridicule and compassion' (all examples taken from Carter & Long, 1990: 216).

Clearly, what these questions do is test knowledge, either *of* or *about* the piece that is being examined. Types 1 and 2 mainly test memory and public knowledge of and about the work, and although answers may not be identical, they will be convergent and extremely close to each other. Type 3 does not directly test knowledge, but clearly relies on knowledge and on memory, though the responses might be less convergent than the answers to Types 1 and 2. However, with Types 2 and 3 there is an additional problem, namely, that the student does not need to have actually read the text in order to answer the question. As Carter and

Long (1990) point out, this is what Short and Candlin (1986: 89) describe as the 'flight from the text'. Rather than reading the original piece, students may have read a crib, read a translation or indeed read a crib in translation. Another possible washback effect here is that preparing for the examination can become a matter of preparing a crib, as used to happen when I was teaching in a secondary school in Israel. We would give our students, in preparation for the matriculation examination, a handout with about 50 questions about the play we were teaching, Arthur Miller's *All My Sons.* In a way, this handout 'worked' – the students did get to know the play. But their preparation for the examination, instead of being based on the play, on re-reading it, on building up their own engagement and their own view of the play, was quite clearly based on going through this list of questions and finding answers either in the play or in the crib they had, preparing notes for their answers and so on. They were not engaging with the play – they were engaging with the questions about the play.

It is also important to note that in all three types, we are not looking at literary competence or the ability to apply knowledge to new situations. What we are doing is testing the ability to re-produce other people's meanings (see also Dilemma 4). It is interesting that even the Type 3 questions that Carter and Long (1990) provide as illustration do not require the testee to form an opinion, but to provide evidence and argue for opinions that have been formulated for them.

In contrast to the testing of literary *knowledge*, we can test literary *skills*, which is often done through the use of unseen passages. For example, the students can be given a text with a non-literary paraphrase, and asked to point out the differences. They may be given a passage and asked to predict from it. These skills thus demand quite a close reading, and the washback effect will be to bring about close reading and the development of literary skills. For example, Lin (2006) describes a curriculum model where students learn analytic skills through engaging in guided reading of a poem, followed by the reduction of support from the teacher and working on a different poem on their own – in effect an unseen poem, replicating what will later happen in the examination.

Dilemma 4: Testing private appreciation of literature or testing public knowledge about/of literature?

The discussion of the previous dilemma parallels to some extent what Rosenblatt (1985) has called efferent versus aesthetic reading. Rosenblatt suggests that the process of reading literature is governed by what the

reader does with literature, and she has posited that there is a continuum between two main types of reading. *Efferent reading* (from the Latin *efferre*, meaning carry away) is the reading that results in the type of knowledge that is public. The focus is on shared meanings, on 'what is to be retained after the reading – to be recalled, paraphrased, acted on, analysed' (Rosenblatt, 1985: 101). In *aesthetic reading*, the reader's attention is focused on the experience as it is lived, the pleasure that is derived from the act of reading itself. As Rosenblatt (1985: 102) says, 'the range of ideas, feelings, associations activated in the reservoir of symbolisations is drawn upon. (The reader may retain much afterwards, but that is not the differentiating aspect)'. For example, while working on this paper, I overheard a friend talking about mobile phones and answering machines, and connecting this to a novel she was reading at the time, Theodor Fontane's *Effi Briest*, a novel in which letters play an important part. My friend began talking about how, in the 19th century, people communicated through frequent letter writing, sometimes writing to each other every day. So, what she was doing as a reader was entering the world of the novel, imagining what it was like to live then, and making connections with life in the 21st century and her own life.

A great deal of literature teaching around the world is probably of the efferent kind, which is precisely what makes literature seem like such a chore to students. It is possible to 'teach' aesthetic reading and encourage learners to do so; Tutaş (2006) has shown how it is possible to do this in a university setting in Turkey. Ironically, very often this can be done through *not* talking about literature; in the classroom it is done through consistently encouraging students to talk about what they liked and didn't like, and consistently asking them to connect what they have read with their everyday life.

In testing terms the problem is immediately apparent. Questions like the ones from Carter and Long (1990) discussed above all refer to public knowledge. The washback effect of doing so is not only to make learners flee from the text, but also, even when they are reading the text, make the reading they experience efferent reading: they will approach a novel not with the idea of enjoyment, but with the conscious knowledge that as they read it they need to mentally note issues of character, plot, setting etc. Their reading experience will be coloured by this. For example, the York Notes for *Pride and Prejudice* (Pascoe, 1998) include a section on 'How to study a novel'. The section tells the student, 'start by reading (the novel) quickly for pleasure, then read it slowly and carefully'. This is because 'further readings will generate new ideas and help you to *memorise* the details of the story' (Pascoe, 1998: 5; italics mine). The advice

goes on to say 'make careful notes on theme, plot and characters of the novel. The plot will change some of the characters. Who changes?' (Pascoe, 1998: 5). The rest of the advice is all focused on literary analysis and efferent readings, providing a clear indication of what it is that testers are requiring in this context.

However, one of the reasons that tests concentrate on public knowledge is that it is easy to devise such tests, and, more importantly, it is easy to devise marking schemes for them. How does one devise criteria for personal response? How does one characterise, for example, an 'appropriate' personal response? Having said that, it has been suggested that 'there are signs in assessed coursework that some students can be as apt in providing what they think will be ''approved'' responses (feminist, Marxist, structuralist) as they were at parroting approved interpretations' (Protherough, 1991: 15).

Dilemma 5: Authentic/genuine tasks or pedagogic tasks?

I use the words 'authentic' and 'genuine' in this context in a general sense, to denote what it is that most readers 'do' with literature after they have read it. I would like to suggest that the most common 'real life' activity we perform with literature is talk about it. We tell each other plots, we recommend new books to read for each other, we explain why we like a book to a friend or a partner. This is one of the reasons for the exponential growth of reading groups and reading clubs in the UK, for example. The example above of a reader talking about *Effi Briest* is a good illustration of the way in which the literature we read and our reactions to it become part of our lives, something to talk about and refer to. Another example is provided by Fecho *et al*. (2007: 45), who describe a classroom conversation in which the teacher and the pupil 'used classroom time to act increasingly like two friends having a chat about books at a coffee shop rather than taking the inquisitor/rote responder roles seen all too frequently in too many literature classrooms'. Written responses to literature are rarer among non-specialists, although the emergence of the web has meant that, increasingly, non-specialists are writing reviews of books, both relatively short (e.g. some of the reviews for online bookshops) and much lengthier ones (e.g. on personal blogs). So, most probably, if we wish to include some sort of 'task achievement' in our criteria for marking a task, we should really test literature orally or, if we are testing it in writing, present the non-specialist version of the task – the web review.

Dilemma 6: Should we require metalanguage?

To some extent, this is a false dilemma, because for me the answer is clearly a yes. If we are teaching skills, or if we are encouraging a response, then we need also to teach the vocabulary to talk about that response. The use of metalanguage is a sign of being able to take advantage of the text, of being able to view it from a distance, of the ability to make generalisations about it. In addition, teaching metalanguage both draws upon and reinforces cross-curricular connections with the teaching of literature in the learners' first language (L1), because the metalanguage used will often be identical in the L1 and the second language (L2). Teaching metalanguage then means that it needs to be tested, or at least listed as one of the criteria in marking.

Principles for Testing and Assessment

Having articulated the various issues that arise when testing literature in the language classroom, I now turn to my own response to the dilemmas in the form of a number of principles.

Include both public knowledge and private appreciation

Testing knowledge about literature is testing public knowledge, mainly through display questions to which the tester knows the answer, and to which the answers are convergent, representing conventional readings. Of course, this makes the tester's life easy, mainly in terms of reliability. But this tells us little about what impact the literature has had on the learner, and as we have seen above, this leads very easily to a 'flight from the text' (Short & Candlin, 1986: 29). The only way out of this dilemma is to ensure that the questions we ask not only ascertain that the learners have the public knowledge about the works they have read, but also provide the learners with an opportunity to show their private appreciation of the works. The latter will also entail knowledge of the metalanguage needed for a discussion of basic issues in literature. This will also lead to differentiated marking schemes and criteria.

Use a variety of tasks

The most important implication of the first principle is that we need to use different tasks to test different aspects of literary knowledge. Public knowledge questions about plot, background, character etc., are best tested through multiple choice, true/false, cloze texts etc., which enable the testee to produce the convergent display answers mentioned above

(which, in addition, require little language production, and sometimes none). On the other hand, personal understandings and appreciation of the work can only be tested through extended writing or through creative reactions involving production. We normally think of those types as requiring a linguistic response, but there are ways of doing this without a linguistic element (see the section 'Include choice'). Finally, literary competence can only be tested through using unseen passages, or through asking the learners to perform new operations on texts that they have already seen. A test of literature should thus include a large variety of different tasks, which together ensure that we are tapping the different aspects of the response. (It is interesting to note that the example provided by Carter and Long (1990) includes a list of 17 different questions, falling into three types – general comprehension, text focus, and personal response and impact. The assumption is that there has been some class discussion of the possible meanings of the poem.)

Include choice

Within the variety of tasks I am suggesting that we use, we should also allow the examinees a choice. We are not really accustomed to choice in mainstream language tests, with the exception of writing, where it is acknowledged that there are likely to be individual factors that will make a task stimulus more difficult for one learner than for another, or possibly totally unsuitable. Since the previous principle entails an elaborated linguistic response, this means that there has to be choice in such tests.

Provide texts

If a test taker takes the test with the text provided, i.e. some sort of open book examination, then the test stops becoming a test of memory; it taps into the skills that the learners have acquired, or can tap into their reactions. On such examinations, retrieving information from the text is demoted to a minor role, because the information is in fact provided in the examination paper. (One aspect of 'open book' examinations is that examinees often believe that because they can consult the text, the test is easier. Of course, there is an element of truth in this, but in fact students normally do not have much time to consult the texts, and doing so extensively may take up a great deal of time, which is reduced when the students are well acquainted with the text.) The example from Carter and Long (1990) clearly indicates that texts are to be provided; in another context, Brumfit and Killam (1986: 253) suggest that students should be allowed to 'consult up to six set texts during the examination'. Spiro

(1991), too, suggests that 'test items should require contact with actual text', though the reason that she provides for this is 'eliminating the possibility of dependence on prepared notes'. I would suggest that the reduction in the need to memorise that results from providing a text during the examination, can have a positive washback and lead to engagement with more texts, because less time will need to be spent on any specific text.

Include portfolio assessment

Portfolios, which are becoming increasingly popular in a variety of educational and other contexts (see Klenowski, 2001), incorporate many of the points I have made above through their very nature, and are 'the most frequently identified example of alternative assessment' (Fox, 2008: 99). They provide variety in terms of task; they can provide choice in terms of text and in terms of task; and since they are done at home, or in class, with the texts in front of the student, they do not need to test content and memory. Ferradas Moi (2000) and Zoreda (2002) are two examples of the way in which portfolio assessment can be used in this area – the former in a Shakespeare seminar for future English teachers, and the latter in a non-credit-bearing university language course focusing on science. Ferradas Moi (2000) provides a response to some of the dilemmas I have listed above, mainly in the way in which the educational activity of teaching can be combined with the gate-keeping function of testing and assessment, and how the two activities may also result in disappointment for the learner. Her situation exemplifies how a move from assessment *of* learning to assessment *for* learning cannot be a complete one – there is always going to be, at some stage, assessment *of* learning, and the no-test option, in most cases, is not a real option. Thus, as Klenowski (2001: 11) says, 'for summative or certification purposes, the portfolio is usually considered along with other evidence'.

Ensure that criteria are transparent

In many language examinations, we do not actually tell the students what the criteria are going to be (even if they seem obvious to the tester); we also normally do not reveal any differential weighting of the criteria. In the situation that I have described until now, students need to be told how their work will be marked, as this may strongly influence the way they go about planning it and planning their time (see also Spiro, this volume, regarding the transparency of criteria).

Minimise the weighting of language

This obviously looks an odd principle in a discussion of language teaching. It is, however, one type of response to the first two dilemmas I presented above, and is probably the only practical way of responding to them.

Construct multi-part tests

Finally, one overarching principle that emerges from a consideration of all the principles discussed above is that applying them entails constructing *multi-part tests*, which would include sections that address the following issues:

- A reaction to the whole – a personal reaction that exhibits knowledge of the text.
- An appreciation of literary devices and the use of metalanguage.
- An ability to engage, to a certain level, with unseen texts.

The best way of doing this is to include a large number of short tasks, each of which tests one of the above. These different tasks do not all have to relate to the same text, but can relate to a variety of texts, some seen and some unseen.

The Principles Applied

In the final section of this chapter, I look at the task types that are available for testing literature. I will discuss them under four general headings: language-based tasks; non-linguistic tasks; linguistic tasks involving L1 and L2; and tasks involving unseen texts.

Language-based tasks

This is the most familiar category of the four headings I have mentioned above. It includes all the questions in the two typologies that Carter and Long (1990) offer: the typology that I discussed above (paraphrase and context; describe and discuss; evaluate and criticise) and the typology that they offer in its place (general comprehension; text focus; personal response and impact). This category offers a variety of tasks that require the learner to incorporate in their answer both public knowledge and personal response. (In many cases, these tasks also require creativity in the responses; see Spiro, this volume, for a detailed discussion of the issues involved in the assessment of creativity.)

- *Writing a review of a book for a newspaper or a magazine*
  This is an attractive option, but may prove harder than it seems, since it often requires a sophisticated understanding of the genre, and a sophisticated ability to manoeuvre, in the review, between the two positions. In fact, a third element is being introduced here, namely, adherence to genre conventions, which is probably never seriously considered in language tests. A variant of this task is to write a review for a website. The advantage of this is that reviews on websites are a far less conventionalised and therefore a less demanding genre – to the point that one might even query whether the conventions for this are fixed enough for this to be called a genre. Another possibility is to respond to a review from a website – one or more reviews from a website could be downloaded and used as a stimulus for students to write their own review or a response. This is a variant of the task of responding to a critic, or if possible, two (see, e.g. Paran, 1999; see also Spiro, 1991: 86, where examinees are asked to rank short statements made by critics). This makes a response easier, as the student can argue for or against someone, rather than having to come up with their own ideas. An additional (computer-based) response is to create a website for the work, with links to other sites. The criteria here would need to be quite explicit, because it would be very easy to produce something that has little value. One way of doing this would be by providing learners with a list of, say, 10 or 15 websites, and restrict them to only 5 links, for example. They would then need to produce a rationale for the links that they have chosen.
      This would also have important washback effects – for example, getting students to engage with websites during their learning as well. As part of preparing for an examination, students might either create their own websites (see, e.g. Schaumann, 2001) or simply post a review on a website. In a climate where most educational systems want students to raise their knowledge and awareness of technology, this type of washback would be particularly welcome.
- *Writing a sequel or a prequel*
  Again, there is a long tradition of literary works that have done this – the most famous is probably Jean Rhys's *Wide Sargasso Sea*, the prequel to *Jane Eyre* – thus legitimating the genre.
- *Writing a literary piece on the same theme*
  Odile Guigon (personal communication), who teaches in secondary schools in France, has taught stories from Julian Barnes's *Cross Channel*, a collection that explores the connections between France

and the UK. As a follow up, she has asked her learners to write short stories about the same theme.

- *Rewriting the work for the 21st century*
  Not all works may be suitable for this, but again there is a long tradition of such endeavours: think, for example, of Thomas Mann's *Dr. Faustus* and Klaus Mann's *Mephisto*, two 20th-century literary manifestations of the Faust story. In contemporary literature, one of the most striking examples is Will Self's *Dorian*, an ingenious re-telling of Wilde's *The Picture of Dorian Gray*, in which the picture in the attic becomes a video installation, thus lending a realistic aspect to the deterioration of the work of art. This is a literary adaptation, but of course adaptation is something that happens all the time in the theatre and in film. Watching a scene from Zefirelli's *Romeo and Juliet* followed by a scene from Baz Lurman's version is a way of showing learners what can be done. Thus, a task such as this – rewriting a work for the 21st century – is for me a successful task because it mirrors a real life activity that has often been carried out by many writers and poets.

  Obviously, the students would not be able to flesh out a full adaptation of a novel, a play or even a short story, but the task might ask them to write a proposal for 'modernising' the story line and making it more accessible for our times. The task rubric, though, would need to specify the criteria on which the answer would be judged, and the answer would need, for example, to show a knowledge of the original, as well as an understanding of what the various plot devices and props are doing, and which of them can be easily modernised. This would be slightly akin to a book proposal.

Non-linguistic tasks

- *Draw the poem/story/scene/chapter*
  Figure 8.1 illustrates the responses of pupils of mine, aged 15–16, in 10th grade, to Isaac Rosenberg's poem, *Returning, we hear the larks*. Some of these are extremely powerful and in my view a perfectly valid proof that the pupils who drew this understood the poem and responded to it.
- *Design a cover for a book*
  This is a similar task to the previous one, though it would need to incorporate the title and the author's name, as well as take layout
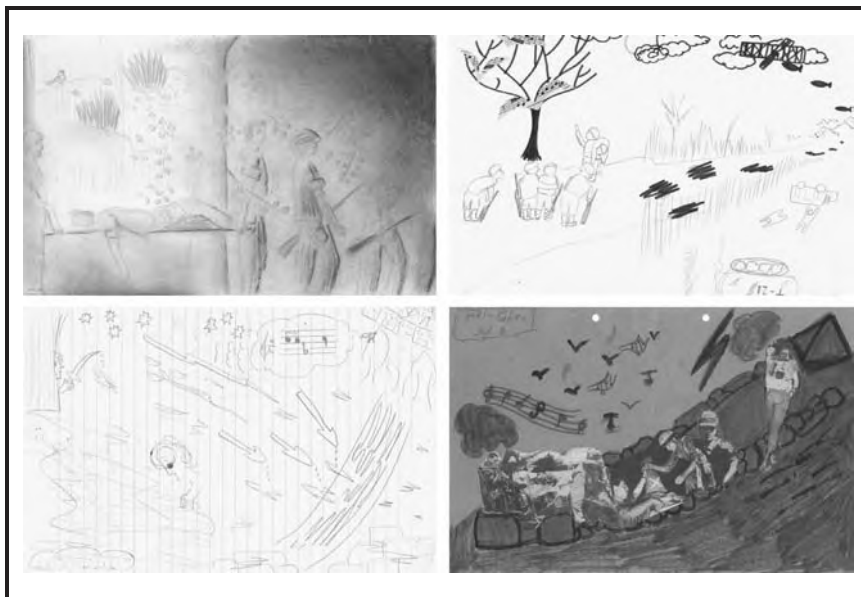
Figure 8.1 Learner responses to Isaac Rosenberg's *Returning, we hear the larks*

issues into account. Both this and the previous task might be accompanied with a rationale for the choices made.

- *Choose an appropriate cover for a book out of a choice of five*
  A detailed discussion of such choices would bring out a great deal of knowledge about the book, and questions of which aspect of the book each cover relates to, what each cover evokes about the book etc. This would entail a discussion of the way covers relate to books in general, how covers function in the social context and an understanding of visual and multi-modal communication.
- *Choose an accompanying piece of music for the work*
  This too should be accompanied by a written justification and explanation.

Linguistic tasks involving both first and second language

- *Choose the best translation from a number of alternatives*
  This task is constrained by the specific translations available, but multiple translations are in fact quite common. In Hebrew, for example, there are at least eight different translations of Dylan

Thomas's *Do not go gentle into that good night* (Reich, 2004); there are five translations of W.H. Auden's *Funeral Blues* (Reich, 1998). A consideration of the different translations provides the students with an opportunity to show their knowledge and understanding of poetry and of metalanguage, as well as incorporating cross-curricular links with L1 study and L1 literature study. The choice of translation means that the learner is considering a variety of poetic forms and meanings, and evaluating the way in which they are expressed linguistically. There are issues here of rhyme and of metre, but also of diction, lexical choice and grammatical issues that impact quite strongly on the verse. In writing about this, the students will need to exhibit quite a high level of literary competence, as well as awareness of the two languages involved (see Widdowson & Seidlhofer, 1996, for a discussion of the way in which the process of translation contributes to language awareness; see Kramsch, 1993, for a specific example of comparing existing translations.)

- *Translate part of the work into your L1 and comment on the product and on the issues that arise.*
  In the same class that the drawings in Figure 8.1 were produced, two students chose to translate the poem rather than draw it. Each found a different solution to the phrases *But hark!* and *lo*, both of which appear in the poem, showing their awareness of register both in English and in Hebrew.

Tasks focusing on unseen texts

Lin (2006) describes how including an unseen passage on the English examination in Singapore seems to have changed teaching quite radically. The examination tests literary competence and skills, and teaching for the unseen means that the students need to be taken through the process of relating to literature, to be taught how to analyse a piece of literature, and to learn how to identify what to pay attention to. What they are taught is a procedure, a combination of skills. The testing then sees whether the students have been able to internalise this way of reading poetry and whether they are able to apply the analytical tools that they have been given.

It is also possible to choose unseen texts that are close or similar to the work that has been used in class – e.g. one that uses similar metaphors or that addresses similar themes. For example, in a 12th grade class in Israel, I taught W.H. Auden's *Museé des Beaux Arts*, which focuses on

three paintings by Brueghel (see Berga, 1999 for a discussion of this poem). In a class test that I gave, I included the following question:

> Read the following poem by W. C. Williams carefully. After you have made sure that you understand all the words, compare it to *Musée des Beaux Arts*. Consider the following points: the language, the imagery, the focus of each poem, the theme, and any other point you wish.

A variation on the unseen text is asking the students to perform new analyses on a piece that they have already read. For example, Spiro includes the following task:

> Read the final paragraph of *The Great Gatsby*. What predictions could you make about the plot, setting and characters of the novel if this were the opening? How would these predictions differ from the actual novel? (Spiro, 1991: 58)

Conclusion

Throughout this chapter, I have stressed the importance of the washback effect of assessment; this is particularly important for literature, precisely because the reasons that we teach literature – or use literature – are very different from our reasons for teaching FL. When we teach FL, we often teach language as a tool for achieving something else – as a tool for academic success; as a tool for succeeding in business etc. Although we may also want people to appreciate language for itself, the bottom line is that if the football coach learning English does not appreciate the language, that does not really matter. But our reasons for teaching literature are different. As the quote from the CEFR presented earlier suggests, we are teaching literature because it is worth teaching in and of itself; because it will enrich our students' lives; because an appreciation of literature is an appreciation of many of the spiritual and intellectual aspects of life, an appreciation that can operate at a large number of levels. So, if the washback effect of the examination is that students cram for it, if its effect is a flight from engagement with literature, if it is a flight from the text (Short & Candlin, 1986), then what the assessment procedure has done is merely to alienate students from literature, and further contribute to what Ellis with Fox and Street (2007: 4) describe as 'the marginalisation of the aesthetic as a uniquely important way of knowing' (see also Tomlinson, 1998, for a discussion of this alienation). It is precisely in order to prevent this type of washback that it is important to collect cumulative evidence of the process of ongoing engagement *with* literature rather than

collect summative evidence of knowledge *about* literature. When I have run into ex-pupils of mine, they never say, 'I remember that wonderful lesson you did on the present perfect' or 'I will never forget that wonderful lesson where you taught us to skim and scan'. But they do remember lessons on literature, they remember specific lessons using songs, they remember the moments when they were engaged as individuals and when they were engrossed in what they were doing. The challenge for us is to ensure that the types of testing and assessment that we engage in enable us to bring into the classroom literature that engages our learners as individuals and contributes to their growth and development.

Note

1. This paper is a written version and development of a plenary with the same title, presented at the 5th International Conference of ETAI (English Teachers Association of Israel), Jerusalem, 11–14 July 2005. I would like to thank Rosemary Wilson and an anonymous reviewer for their comments on an earlier draft of this paper.

References

Alderson, C. (2004) The shape of things to come: Will it be the normal distribution? In M. Milanovic and C. Weir (eds) *European Language Testing in a Global Context. Proceedings of the ALTE Barcelona Conference, July 2001. Studies in Language Testing 18* (pp. 1–26). Cambridge: Cambridge University Press.

Alderson, J.C. and Banerjee, J. (2001) State of the art review: Language testing and assessment (Part 1). *Language Teaching 34* (4), 213–236.

Bachman, L.P. (1990) *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Berga, M. (1999) Fallen angels: On reading landscape and poetry. *Language Awareness 8* (1), 51–61.

Brumfit, C. (ed.) (1991) *Assessment in Literature Teaching*. London and Basingstoke: Macmillan (Modern English Publications in association with the British Council).

Brumfit, C. (2001) *Individual Freedom in Language Teaching*. Oxford: Oxford University Press.

Brumfit, C.J. and Carter, R.A. (eds) (1986) *Literature and Language Teaching*. Oxford: Oxford University Press.

Brumfit, C.J. and Killam, G.D. (1986) Proposed examination paper. In C.J. Brumfit and R.A. Carter (eds) *Literature and Language Teaching* (pp. 253–255). Oxford: Oxford University Press.

Carter, R. and Long, M. (1990) Testing literature in EFL classes: Tradition and innovation. *ELT Journal 44* (3), 215–221.

Carter, R. and McRae, J. (eds) (1996) *Language, Literature and the Learner: Creative Classroom Practice*. Harlow: Longman.

Cheng, L. and Curtis, A. (2004) Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng and Y. Watanabe with A. Curtis

(eds) *Washback in Language Testing: Research Contexts and Methods* (pp. 3–17). Mahwah, NJ and London: Lawrence Erlbaum and Associates.

Cheng, L. and Watanabe, Y. with Curtis, A. (eds) (2004) *Washback in Language Testing: Research Contexts and Methods*. Mahwah, NJ and London: Lawrence Erlbaum Associates.

Cook, G. (1996) Making the subtle difference: Literature and non-literature in the classroom. In R. Carter and J. McRae (eds) *Language, Literature and the Learner: Creative Classroom Practice* (pp. 151–165). Harlow: Longman.

Edmondson, W. (1997) The role of literature in foreign language learning and teaching: Some valid assumptions and invalid arguments. In A. Mauranen and K. Sajavaara (eds) *Applied Linguistics Across Disciplines. AILA Review*, 12 – 1995/6, 42–55.

Ellis, V., Fox, C. and Street, B. (eds) (2007) *Rethinking English in Schools: Towards a New and Constructive Stage*. London and New York: Continuum.

Ellis, V., with Fox, C. and Street, B. (2007) Why English? Rethinking the school subject. In V. Ellis, C. Fox and B. Street (eds) *Rethinking English in Schools: Towards a New and Constructive Stage* (pp. 1–14). London and New York: Continuum.

Fecho, B. with Amatucci, K. and Skinner, S. (2007) Deep winters, invincible summers: Teaching English for existential purposes. In V. Ellis, C. Fox and B. Street (eds) *Rethinking English in Schools: Towards a New and Constructive Stage* (pp. 32–46). London and New York: Continuum.

Ferman, I. (2004) The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng and Y. Watanabe with A. Curtis (eds) *Washback in Language Testing: Research Contexts and Methods* (pp. 191–210). Mahwah, NJ and London: Lawrence Erlbaum and Associates.

Ferradas Moi, C. (2000) Assessing the literary competence of future EFL teachers. *IATEFL Testing, Evaluation, and Assessment SIG Newsletter*, June, 17–28.

Fox, J. (2008) Alternative assessment. In E. Shohamy and N. Hornberger (eds) *Encyclopaedia of Language and Education* (2nd edn). Volume 7: *Language Testing and Assessment* (pp. 97–109). New York: Springer.

Klenowski, V. (2001) *Developing Portfolios for Learning and Assessment. Processes and Principles*. London and New York: Routledge/Falmer.

Kramsch, C. (1993) *Context and Culture in Language Teaching*. Oxford: Oxford University Press.

Lin, B. (2006) Exploring the literary text through grammar and the (re)integration of literature and language teaching. In A. Paran (ed.) *Case Studies in TESOL: Literature in Language Learning and Teaching* (pp. 101–116). Alexandria, VA: TESOL.

Paran, A. (1999) Methodological issues in using poetry in the ELT classroom. *ELT News and Views: Literature in ELT Supplement* 6 (1), 20–22.

Paran, A. (2000) Survey review: Recent books on the teaching of literature. *ELT Journal* 54 (1), 75–88.

Paran, A. (ed.) (2006) *Case Studies in TESOL: Literature in Language Learning and Teaching*. Alexandria, VA: TESOL.

Paran, A. (2008) The role of literature in instructed foreign language learning and teaching: An evidence-based survey. *Language Teaching* 41 (4), 465–496.

Parkinson, B. and Reid Thomas, H. (2000) *Teaching Literature in a Second Language*. Edinburgh: Edinburgh University Press.

Pascoe, P. (1998) *Jane Austen/Pride and Prejudice*. Harlow: Pearson Education/ Longman; London: York Press.

Protherough, R. (1991) Assessing response to literature. In C. Brumfit (ed.) *Assessment in Literature Teaching* (pp. 9–15). London and Basingstoke: Macmillan (Modern English Publications in association with the British Council).

Rea-Dickins, P. and Scott, C. (2007) Washback from language tests on teaching, learning and policy: Evidence from diverse settings. *Assessment in Education: Principles, Policy and Practice* 14 (1), 1–7.

Reich, A. (1998) Blues Halvaya – shiro shel Auden be-chamisha targumim. (*Funeral Blues* – Auden's poem in five translations.) *Ha'aretz*, 16 January, D-4. (In Hebrew.)

Reich, A. (2004) Al Telech el Halaila me'et Dylan Thomas bi-shmona targumim. (*Do not go gentle* by Dylan Thomas in eight translations.) *Ha'aretz*, 20 February, D-4. (In Hebrew.)

Rosenblatt, L. (1985) Viewpoints: Transaction versus interaction – a terminological rescue operation. *Research in the Teaching of English* 19 (1), 96–107.

Schaumann, C. (2001) *Erzählraum im virtuellen Raum*: Rewriting Bernhard Schlink's *Der Vorleser* on the Web. *Der Unterrichtspraxis* 34 (2), 150–157.

Shohamy, E. (2007) Language tests as language policy tools. *Assessment in Education* 14 (1), 117–130.

Short, M.H. and Candlin, C.N. (1986) Teaching study skills for English literature. In C.J. Brumfit and R.A. Carter (eds) *Literature and Language Teaching* (pp. 89–109). Oxford: Oxford University Press.

Spiro, J. (1991) Assessing literature: Four papers: 1. Language teaching and the literature test: Laying the foundations. 2. Defining literary skills. 3. Measuring literary skills. 4. Writing the literature test. In C. Brumfit (ed.) *Assessment in Literature Teaching* (pp. 16–83). London and Basingstoke: Macmillan (Modern English Publications in association with the British Council).

Tomlinson, B. (1998) And now for something not completely different: An approach to language through literature. *Reading in a Foreign Language* 11 (2), 177–189.

Tutaş, N. (2006) Theory into practice; teaching and responding to literature aesthetically. In A. Paran (ed.) *Case Studies in TESOL: Literature in Language Learning and Teaching* (pp. 133–145). Alexandria, VA: TESOL.

University of Cambridge, ESOL Examinations (n.d.) *First Certificate in English, Handbook for Teachers*. On WWW at www.cambridgeesol.org/assets/pdf/ resources/teacher/fce_handbook.pdf. Accessed 23.6.08.

University of Cambridge, ESOL Examinations (2008) *Certificate of Proficiency in English, Handbook for Teachers*. On WWW at www.cambridgeesol.org/assets/ pdf/resources/teacher/cpe_handbook.pdf. Accessed 23.6.08.

Weir, C.J and Porter, D. (1994) The multi-divisible or unitary nature of reading: The language tester between Scylla and Charybdis. *Reading in a Foreign Language* 10 (2), 1–19.

Widdowson, H.G. and Seidlhofer, B. (1996) Poetic translation and language pedagogy. In L. Bredella and W. Delanoy (eds) *Challenges of Literary Texts in the Foreign Language Classroom* (pp. 125–143). Tübingen: Gunter Narr Verlag.

Zoreda, M.L. (2002) Teaching short science fiction stories in English as a Foreign Language in Mexico. Paper presented at the Annual Meeting of the Popular Culture Association and American Culture Association, Toronto, Ontario, Canada, 13–16 March. On WWW at http://eric.ed.gov/ERICDocs/data/ ericdocs2/content_storage_o1/00000000v/80/0d/e7/20.p. Accessed 13.5.07.

Chapter 9

Crossing the Bridge from Appreciative Reader to Reflective Writer: The Assessment of Creative Process

JANE SPIRO

Introduction

This chapter explores the process of arriving at an assessment of creativity. It describes the transition from critical appreciation to creative production among a group of native and non-native English-speaking students at Oxford Brookes University, and explores how such a process might be assessed. During the process, the students are guided to appreciate literary texts and the language systems and poetic strategies that make them work, and to transfer this appreciation to creative texts of their own that go beyond modeling, imitation or dependency on others. Here, 'creative texts' are defined as those in which the student is encouraged to 'make something new' that represents their own voice and their own story. How can so personalized and elusive a process be assessed so that learning is enhanced and not inhibited? A number of questions are confronted and answered in the course of arriving at solutions: Are creativity and assessment not contradictory to one another? What notion of 'creativity' is being used here, and how is this skill or term anatomized so it makes sense to both learner and assessor? How can this assessment be assured of transparency, objectivity and meaningfulness to the learner? What can we learn about how native and non-native speakers respond to this process of transition, from critical to creative? These questions will be answered by tracing the 'story' of an assessment process, and including in this story the voices of its key characters, the students, the teachers, the assessors.

The overarching aim of the chapter is to explore the idea of 'making something new' as an objective that is both teachable and measurable. The

chapter, centering on a case study with Oxford Brookes undergraduates team-taught by Rob Pope and myself, illustrates the proposition that every learner has the capacity for creativity and that assessment, far from being a constraint, can in fact be a trigger for creativity.

Creativity and Language Learning

The term 'creativity' is subject to so many definitions, it ceases to have any shared meaning. Pope (2005) explores the full spectrum of definitions, from the ancient notion of *ex nihilo* and god-like '*poesis*', to contemporary notions of 'creativity' as the non-exclusive capacity of all healthy human beings, and several gradations in between. In order to navigate through this complexity, and for the purposes of this chapter, 'creativity' here is located at the latter end of this spectrum, and Pope's (2005: xvi) working definition is the one invoked: 'the capacity to make, do or become something fresh and valuable with respect to others as well as ourselves'. As a shorthand, this working definition will be described as 'making something new', or the capacity for the learner to *transform* knowledge and skills into something of their own.

Given its elusive, non-measurable and contentious nature, why should 'creativity' have a role at all in the language learners' curriculum? Second language learning research has recognized the role of 'expressive writing' in first language learning, but has been skeptical of its validity for the second language learner: 'Although many L2 writers have learned successfully through (expressive writing), others may experience difficulties, as it tends to neglect the cultural backgrounds of learners, the social consequences of writing, and the purposes of communication in the real world, where writing matters' (Hyland, 2003: 10). These are concerns that, as second language educators introducing a creative writing assessment into the curriculum, we needed to take seriously. Specifically, we needed to unpack what assumptions and expectations were embedded in the term 'creative'; and justify why and how this might be of value to learners.

Language learning theory makes clear that the capacity to 'make something new' is essential to the learning process; we assimilate 'rules' in order to subvert them and generate unique and specific messages with them. The speaker/listener has the capacity 'to produce or understand a potentially infinite number of sentences they have not previously encountered' (Maybin & Swann, 2006: 12). To explain this capacity without the notion of linguistic creativity is problematic. We also know that 'ordinary'/everyday use of language involves 'wordplay', invention and creativity. Carter explores the ways in which 'creative' use of

language has a social, personal and pragmatic function and is part of our competence as language users: 'creativity is basic to a wide variety of different language uses, from everyday advertising language and slogans to the most elaborated of literary texts' (Carter, 2004: 18). This includes 'wordplay, puns... verbal ambiguities... sexual innuendos, word inventions', to name just a few.

If we look at synonyms of the word 'to create', we are offered many models of what learners could do with the knowledge they receive: *transform*, *formulate*, *generate*, *adapt*, *change*, *give birth to*, *develop*, *evolve*, *spring from*, *make*, *piece together.* If we take 'language' as the knowledge in question, this set of verbs looks a great deal more promising. Take this as an example:

> To *transform* rules of language into meaningful utterances
> To *generate* utterances that are unique to the learner's experience

In this paradigm of learning, the learner is taking the knowledge and transforming it into something new: using the ingredients to *make something new happen*.

However, this is a broad definition, and does not offer detailed guidance for a learner attempting to engage with this as an educational objective. To be more explicit, our 'making something new' is contextualized within various understandings as to the nature of the 'literary/ poetic'. How is this to be defined and anatomized so it makes sense to learners? Carter identifies six features that represent 'literariness', and which offer criteria by which to measure literary merit. Rather than 'an absolute division into literary/non-literary' texts, he posits a 'cline of literariness along which texts can be arranged', and which form a useful framework for the assessor (Carter, 2006: 85). These six features are: *medium dependence* (literary texts depend less on other media); *genre mixing* **(**recognizing that 'the full, unrestricted resources of the language are open to exploitation for literary ends' (Carter, 2006: 82)); *semantic density* (the capacity for a text to work at several levels); *polysemy* (words resonate with multiple meanings); *displaced interaction* ('meanings... emerge indirectly and obliquely' (Carter, 2006: 82)), and finally, *text patterning*, such as the use of repetition, echoes and recurrence of motifs throughout a text.

In focusing on the 'anatomy' of literariness, Carter helps us with closer definitions of how we might evaluate literary creativity. But how 'fixed' is this cline of literary value? And how watertight will it be as an instrument for assessment? The storm of response to prizes awarded in the fields of fine art (such as the UK art award, the Turner Prize) or literature (such as the UK Man Booker Prize), suggests that the judgment

of creative output is far from transparent or easily received and understood. The values attached to creativity are time-sensitive, and judgments at each period in history are mirrors of contemporary values. Thus at different times, aesthetic judgment has, for example, honored and discredited representation, seen memory on a cline from solid to unreliable, and seen imitation on a cline from the practice of high art to indictable dishonesty (Hunt & Sampson, 2006; Pope, 2005). Assessing creativity, therefore, cannot be socially and contextually disengaged. Assessors need to be explicit about which paradigms are in place, and to unpack assumptions about what is valued and why. The next sections trace the way we responded to these questions.

Creativity and the Scaffolding of Learning

We have seen in the previous section that creativity is integral to how a language is learnt and used; it is thus in the interests of both learner and teacher to make sure that this capacity is allowed to flower in the classroom. But how can this be done? My suggestion is a 'scaffolded creativity', in which the learner is guided through the levels of language, acquiring the skills and ingredients they need to be inventive with language. The earliest definition of 'scaffolding' described a process in which the learner is enabled to 'solve a problem, carry out a task, or achieve a goal which would be beyond (their) unassisted efforts' (Bruner cited in Weissberg, 2006: 248). Since its first formulation, the term has come to be used in many different ways: 'a framework for learning, an outline, a temporary support, a mental schema, a curriculum progression' (Weissberg, 2006: 248). However, both 'strong' and 'weak' accounts of scaffolding share the idea that learners are guided towards learning through carefully framed tasks that offer appropriate levels of challenge, based on the learners' current starting point. Similarly, there is overall agreement that the aim of 'scaffolding' is to lead the learner towards self-sufficiency and the successful completion of tasks.

In order to achieve this double role – framing tasks and guiding towards self-sufficiency – we need to be clear what levels of challenge are being demanded, and what progression through these levels will really mean. How can creative output be staged in a way that will enable a 'scaffolded' approach? Cropley (2001) cites Urban who identified six developmental stages of creativity. Although Urban is defining progression from child to adult modes of creativity, the different stages also form a helpful account of the process of creative 'maturation'. The stages he describes are: *autonomy*, in which the creation bears no relation to other

stimuli; *imitation*, in which the writer simply copies texts without transformation; *concluding/completing*, in which the response is still closely connected to the original stimuli; *producing thematic relations*, in which theme has been 'owned', absorbed and transformed; and finally, *holistic responses*, in which theme and response have been developed to form a fully coherent, and independently successful text (Cropley, 2001: 92). (A sixth and earlier stage relates to children's production of visual images, and is described by Cropley as *isolated animation/objectivation*.) Note that this progression is not unique to the notion of creativity. Bloom's taxonomy of intellectual/thinking levels also places imitation, reproduction and copying of ideas at the lower end of the spectrum, and the invention/generation of ideas at the highest level. His notion of invention/generation is also supported by the capacity to evaluate and analyze (Bloom, 1956; Vahapassi, 1982; Weigle, 2002). In using the term 'scaffolded creativity', I will be invoking this understanding of the progression from imitation to self-sufficiency, through the medium of tasks that guide and highlight strategies.

The Language through Literature Module

This section will show how the scaffolding discussed above was achieved with reference to a 'Language through Literature' module and two generations of students that have moved through it. The module was team-taught by myself and my colleague Rob Pope, and ran over 10 weeks, for 2.5 hours a week. It was offered in a range of different capacities within the University: as a module integral to an under-graduate degree with either an English Literature or an English language focus; as a free-standing module that can combine with other subjects; or as a stand-alone module forming part of an exchange visit to Oxford. Thus, there was a highly mixed demography in the class, including native-speaking students with a strong traditional background in critical response to literary texts, non-native-speaking students whose goals are predominantly language development and cultural exchange, and students with no particular specialism in either language or literature, combining this module with programs in Tourism, Business, Hotel and Catering (for example). The chapter describes a 2005/2006 cohort of 53 students, and a 2004/2005 cohort of 54. These classes comprised an average of 22% international and non-English-speaking students, and 7% mature students (see Table 9.1). In every sense, then, the learners represent a challenging range of levels, agendas, proximity to literature as a field and proximity to language as a goal.

The module guides students through a series of linguistic building blocks. In 2004–2005, we began with the smallest unit of language, the phoneme, working outwards to larger components as in Table 9.2. In response to student feedback, we reversed the order in 2005/2006, starting with the largest chunks of language, whole texts, and moving back to discrete units. In both cases, the module leads the student from level to level, both from the perspective of an appreciative reader, and from the perspective of a creative writer.

To accompany the journey through language levels, students were issued with two collections of poems: one 'long list' for discussion

Table 9.1 Student groups on the language and literature module

	TOTAL	*Native English speakers*	*Non-native English speakers*	*Mature students (over 25)*
English Literature specialists	58	56	2	6
English language/ Communication specialists	20	18	2	1
Exchange students on 1-year General Study exchanges	25	2	23	0
Other disciplines	4	4	0	0
	107	80	27	7

Table 9.2 Language stepping stones

1. sounds
2. words: morphology
3. connections between words: phrases
4. utterances and sentences
5. connections between sentences: cohesion
6. organization of information: coherence
7. whole texts; genres and text types

and analysis in class (see Appendix 1) and a second shorter collection of 12 poems, for use in the assessment cycle (see Appendix 2). Each 'level' made explicit key concepts and terminology and applied these both to the appreciation of texts, and to the creation of them.

Each discussion and session starts with a key question, which connects language with purpose and meaning; then explores key concepts and terminology that lead to detailed understanding of how language works at that level; then illustrates these ideas with reference to a range of poems that are deconstructed and analyzed; and, interwoven throughout, is the question – 'how could I use these linguistic and poetic strategies to convey my own messages and write my own texts?'

Table 9.3 shows the architecture of the module, in that critical/analytical appreciation is the first part of a process leading to creative response, and linguistic concepts are explored as a stepping stone towards deepened understanding of a text and its processes. In other words, each text and language level is explored from the point of view of how it illuminates the process both of reading, and of writing. The session concerning Word level, for example, considers: affixes and suffixes, how these can influence meaning and word function, how 'new' words can be constructed through word families, morpheme changes and word compounds, the semantic properties of words and how these are influenced by linguistic environment. All these are illustrated through literary texts with particular lexical vitality: for example, John Agard's *Half-caste* and John Updike's *Superman*. Through encounter with writers' lexical experimentation, learners are invited to experiment for themselves, by generating word compounds, word families and collocations of their own. The session on Genre follows the same pattern, exploring the generic experiments of recipe, list, memo as poem, and the ways in which generic features are both deployed and subverted by the writer.

Assessment as a Scaffold for Creativity

The assessment cycle

Thus far, I have shown how students on the Language through Literature module were given a linguistic/poetic scaffold with which to read and appreciate texts. This section explores how the assessment framework evolved to take account of best practice in assessment, the stages of creative maturation, the complexity of defining creativity itself and the need to be transparent and accountable to both learners and to the academy.

Table 9.3 Deconstructing the appreciation of sound as reader and writer

	Sound	Reading poems
Key question	How is sound used to convey meaning?	
Key terminology/ concepts	articulation of sounds, such as plosive, fricative; the capacity of sound to convey meaning (onomatopoeia); phoneme v. grapheme; patterns of sound: alliteration, rhyme, etc.	
Analysis to creative response	**Sound as meaning** What actions or moods are conveyed by the sounds? How might we convey moods such as command, anger, surprise, fear, using sounds alone? Experiment with telling a story or setting a scene using sounds only.	*The Blacksmith* (**Anon**) *Tik, tak! Hic, hac! Tiket, taket!* *Tyk, tak!* *Lus, bus! Lus, das!*
	Sound:Spelling relationship How does the sound:spelling link convey the poet's message about identity and language? What is the difference between the sound: spelling relationship in the two poems? Experiment with spellings in order to suggest dialect and idiolect. What is the effect of your spelling changes?	**John Agard's** *Half Caste* *Explain yuself* *wha yu mean* *when yu say half caste* **Roger McGough's** *Streemin* *Im in the bottom streme* *which means im not brigth*

Table 9.3 (*Continued*)

Sound		Reading poems
	Sound patterns: What is the effect of alliteration in the medieval alliterative poem *The Blacksmith?* What feelings, movements are conveyed? How are these influenced by the actual mode of articulation (plosives: voiced/unvoiced sounds)? Identify alliteration in today's advertising jingles and names of products; write one of your own for your own product.	*The Blacksmith* (**Anon**) *They spitten and spraulen and spellyn many spelles: They gnawen and gnashen, they groan together.*

The assessment involved two phases, a first formative task analyzing and responding creatively to a single poem, and a second final task involving a comparison between two texts and the creation of a third that connected in some way. A collection of 12 poems was provided for the purpose (see Appendix 2), none of which had been discussed in class. A further collection of 12 poems was provided for experimentation, discussion and modeling of the assessment process in class. Both sets of poems could be paired in a number of ways, on the basis of theme, genre, linguistic patterns and style, form, and technical qualities such as the use of metaphor. Creative response could spring out from any aspect of the stimulus text: for example, theme, pattern and form, a subversion or mixing of the genres, an extending or reworking of the central metaphor. Finally, students were asked to analyze their writing process and decisions, explaining in what ways the chosen texts had formed a springboard for their own work, and analyzing their own linguistic and thematic choices. The tasks were weighted 30% for the first task, and 70% for the second one.

We chose comparison/contrast as a starting point, since this can often focus analysis in a way that single-text discussion cannot. Contrast provides a context and incentive for detailed analysis. It is often easier to describe by negatives: to say what a poem is not, in contrast with what another is. Another reason for asking students to choose two starting points was to illustrate the point that creation does not spring *ex nihilo*, that writers themselves draw on one another and that this is a natural part of the development process; and to focus on the writer's own understanding of the difference between imitation/plagiarism, and legitimate inspiration. Two texts also provide a rich environment from which readers might identify patterns, strategies, topics and poetic ideas, and 'own' them from their own perspective.

We also believe that a reflective process of writing may lead to more generative possibilities: 'liking' or 'copying' a good idea will not be the same as understanding its origin and its components, and thus being able to own, transfer and apply them to other contexts. This combination of texts and tasks thus provides the richest possible mix of opportunities, from which something might 'trigger' identification and ownership.

Assessment for learning

The assessment cycle here exemplifies the notion of assessment for learning described in Chapter 1: the practice of embedding the assessment process into learning objectives and generating assessment

activities that emphasize process and facilitate development. The assessment is underpinned by the notion of assessment as an opportunity for learning, self-appraisal and focused feedback, rather than – or as well as – grading and measurement. This principle is well-founded in recent research and policy that suggest that 'assessment *for* learning (as compared to assessment *of* learning) is one of the most powerful ways of improving learning and raising standards' (ARG, 2002, my italics and parenthesis).

The Assessment Reform Group found that five factors helped to achieve learning through assessment. These five factors informed the structure of our assessment and our teaching on the module, as detailed below.

Effective feedback: the mid-module formative task was designed to be a learning opportunity and preparation for the final assignment. Feedback was matched to criteria, which were honed to be increasingly explicit about the stages of creative maturity (Cropley, 2001) and the perceived levels of intellectual engagement (Bloom, 1956; Vahapassi, 1986).

Active involvement of learners in learning: students were invited to take ownership of the themes and issues raised by the 12 classroom texts. The texts were chosen to reflect different Englishes and the multicultural voices of writers. They also showed poets experimenting, inventing and genre mixing. Students were invited to engage with the meaning and form of the texts, to compare responses and issues with their own contexts and cultures, to identify themes and characters, and to develop creative responses of their own.

Adjusting teaching to take account of assessment: as a result of the formative task, a more detailed set of criteria was developed and offered as a learning/teaching tool in the second half of the semester.

Recognition of the connection between assessment and motivation: the first task led to a reading of work in progress in which students benefited from hearing how peers responded to the same initial stimuli.

Need for self-assessment and recognizing ways of improving: the more detailed criteria, group readings and teacher feedback formed a framework, or 'scaffold', for the second final assignment. The formative task represented 30% of the whole grade, as an incentive to engage with this seriously.

In establishing principles of good practice for 'learning through assessment', we also needed to take account of creative 'maturation': what are the different stages of 'learning' in this specific 'creative' domain, and how can these be reflected and prepared by the criteria? It was important, therefore, to take account of the creative stages discussed

above (Cropley, 2001; Vygotsky, 1962). Imitation and close proximity to the initial stimulus are flagged as a first and less advanced response to the task; adopting and taking ownership of themes and issues as a higher level; and the capacity to generate independent free-standing texts that are complete and coherent, as the highest level. A response that bears no reference to the stimulus text and does not acknowledge influences is deemed a less-valued response than one that fully engages with the stimulus text and has an informed and critical recognition of its influences.

Establishing transparency

Finally, the assessment developed to be transparent and accountable to both learners and the academy over three years of the module, in three ways: clarity of the rubric, clarity of criteria, congruence and appropriacy of the feedback.

Clarity of rubric

Over three years, the rubric became increasingly explicit, taking account of underlying principles, student feedback and core beliefs about good practice.

- Any reference to the term 'creative' was deleted from the rubric and the criteria, given the assumptions and ambiguities embedded in the word.
- Students were explicitly informed that all texts were complete, as several students believed they were reading extracts from longer works: *Select two texts from the Poems for Assessment booklet of 12 short and complete texts*.
- Students were offered a list of ways in which texts might be paired in the final assignment, and encouraged to invent their own pairings. The 'thematic' pairing is first in the list, as a suggestion that this can lead to a higher level of response: *The two texts might invite comparison or contrast from a range of viewpoints: theme/topic, genre, patterning, language experimentation. Students are invited to make their own combinations of texts, and justify these in their analysis*.
- Students were reminded that their analysis might refer to all levels of language discussed in class: *Compare and contrast your chosen two texts, drawing on all levels of language discussed in class to explore the similarities and differences between the message, form and meaning of the texts: sound, word, word-mixing (imagery), syntax and sentence patterns,*

connections between sentences (cohesion), text organization, generic features, discourse features.

- Students were reminded of 'genre mixing' as a high level of 'literariness' (Carter, 2006), and encouraged to try this for themselves: *Write a third text of your own, which responds to one or both of your chosen texts. You may wish to take your cue from the aspects featured above. The result may take the form of, for example, a poem, story, dialogue or a combination of all of these and more. Be adventurous!*
- The components of a reflective commentary were made explicit, so teachers and students had a shared understanding of this; thus encouraging the stages of analysis and evaluation (Bloom, 1956) and the acknowledgement of context and conventions (Vygotsky's objective creativity): *Go on to add a reflective commentary in which you discuss your own writing process: how the initial text(s) influenced you; how you changed and developed it/them; which strategies and features you deployed; and how successful you feel you have been in achieving your aims.*

Clarity of criteria; clarity of feedback

As criteria became fine-tuned and more precise, they formed a framework for collective feedback to the cohort as a whole. Criteria evolved in two phases. At first, we formulated a broad-based set of points, which connected with the maturation stages of creativity described above, and also with the program as a whole.

In terms of content, the criteria made known to the students guide them towards the development of independent, holistic and 'medium independent' texts:

- *Understanding of key principles and strategies for exploring and experimenting with texts*
- *Ability to apply these principles independently and critically*
- *Ability to use linguistic terms and techniques accurately and appropriately*
- *Ability to respond to texts creatively as well as critically, and to reflect upon your own writing processes*

In terms of presentation, the rubric here ensures learners' work is firmly embedded within the conventions of the academic writing community; thus that they are capable of an 'objective' creativity that references itself to the outside world:

- *Clarity and coherence of writing, including clear and effective layout*
- *Accurate and complete referencing of sources*
- *Adherence to the given word limits*

The first formative assignment provided a learning opportunity for both students and teachers as assessors. We, Rob Pope and myself, were able to analyze our own responses, our own values and expectations concerning what we were looking for in a 'critical' response, and what we meant by a 'creative' response. The notion of creativity specifically came to mean: 'ownership' of themes and message, 'saying something new', personalized and authentic and being aware of the strategies and techniques for doing so. Mechanical experimentation, dependency on others, paraphrase or 'borrowing' of the themes or issues read about elsewhere, were all judged by the two assessors as outside good practice. The reflective process of creativity was also made explicit; that the writer be aware of the way he/she is conveying messages, and the way sources have informed and inspired the text. Specifically, these 'ways' might be interpreted as linguistic, thematic, poetic, artistic. This could also involve reversing, echoing, personalizing, transforming what has been read; direct 'borrowing' would lie outside the scope of creative good practice, unless this 'borrowing' were to comment, reverse or embed into a different setting.

As a result of this evolved understanding, the following guidelines were generated, with which to interpret and unpack the core criteria:

Creative response

These were some of the ways you responded creatively:

- *mechanical playing with the language of the original*
- *replying to the original*
- *changing the genre of the original: eg. poem to email, letter, monologue, personal ad, story*
- *staying close to the original but manipulating some of the language*
- *reversing the theme/message of the original: eg. positive to negative (eg. love to hate, praise to blame)*

All of these were excellent, but those which were ONLY about mechanical manipulation tended to do the following:

- *not fully engage with the meaning and message of the poem*
- *not involve your own thoughts, beliefs, ideas*

So next time, try and write something that is saying something important to you. Try and include **yourself** *in your creative response. The best assignments did do this.*

Critical response

Take care to use terminology and vocabulary precisely and only when you fully understand their meaning. Some words have an 'everyday'/secular

meaning, and a very specific one when used in a linguistic context. Take care that you use these words precisely, as appropriate to the genre of academic discourse and linguistic analysis. Here are some examples. Check their meanings in a linguistic context.

- *simple*
- *complex*
- *genre (connect this with the notion of **text type**)*
- *Appropriate (use this instead of **correct/incorrect** language)*
- *Colloquial (use this instead of **casual** language)*
- *Formal/informal (check your understanding of the term **register**)*
- *present, past and future tense: these are incomplete descriptions. Remember that we also need ASPECT in order to describe a tense fully*

Take care to connect linguistic features with meaning.
When you discuss a linguistic feature, remember to show:

- *what effect this might have on the reader*
- *what meaning this conveys*
- *what you think the writer's intention or message might have been*

Thus far, we have shown how the assessment framework took account of models of creative intelligence, literariness and assessment for learning; and how rubric and criteria were honed to reflect both underlying principles and evolving experiences. Having considered researcher, assessor and teacher perspectives, we now turn to the responses of the learners themselves.

Students Crossing the Bridge: Appreciative Reader to Reflective Writer

The assessment cycle, and the formative opportunities offered before final submission, yielded work with a high quality of engagement and personal investment. As an overview of responses, pairings that were based on theme/topic were the most popular. Pairings on the basis of experimentation with form were chosen by very few across the two cohorts. The most popular 'pairings' included: theme/topic (e.g. Louise Bennett and John Agard – see below); the two praise poems, a Dinka poem to the bull and praise to the Behemoth in the Book of Job; patterning (e.g. Kit Wright and Mirosluv Holub both use repetitive sentence patterns as an echo or refrain); language experimentation (for example e e cummings and Edwin Morgan experiment with punctuation, invented word compounds, onomatopoeic constructed words).

Students in the highest band of success engaged with themes, rather than attempting to stay close to the surface features of the original. This can be seen in the examples of thematic development from a number of different 'pairings' in the assessment collection, discussed below.

Pairing choice 1: John Agard: 'Oxford Don', and Louise Bennett: 'Colonization in Reverse'

This was one of the most popular choices. Students identified several shared themes, some highly politicized, such as colonizer/colonized, insider/outsider, oppressor/oppressed, public language/private language, unemployment. Others responded more personally, with, for example, personal responses to place, work and status.

Jeanne connected the issue of colonization in the poems, with her own story of immigration and her dual nationality.

> I felt concerned by the themes presented in these poems: colonisation: because Ghana where my parents are from was part of the British Empire and immigration because my parents immigrated to France. ... Being the daughter of immigrants, you experience the feeling you are in between. I'm not completely French but I am not completely Ghanian. (So) – I have integrated the first sentence of the Ghanian national anthem – and the first sentence of the French national anthem (into my poem).

> Ghana is my homeland
> But
> I am also a daughter of France.

> – ''but'' is isolated to show the ambiguity I feel when it comes to explaining where I am from.

Camille focused on the idea of speakers absorbing mainstream languages in order to assimilate. She explored the idea of language speakers influencing one another and generating a mutually meaningful code.

> The ideas that interested me most... were that a minority language could usurp and influence a widely accepted form of speech, and that language itself could be infectious. At first I considered writing a poem or story using an invented language or way of speaking. I also thought of languages that have been used as a form of control (George Orwell's *newspeak* and secret, defiant languages, like Nushu: a Chinese language spoken only by women). I decided, however,

that I wanted to write something where the minority in question is just one person, rather than a whole race...

A sketch follows, of a girl who speaks her own ''other-language'', which is gradually adopted by everyone around her.

Iris drew from the two poems the theme of employment/unemployment. She also recognized the contemporary power of the poems, relating them to the Labor government's 2006 restriction on Jamaican entry into the UK.

Tony Blair is a very hard man...
Him say ''inna Englan Jamaican don' belong''...
De only immigrant him want is illegal one

Pairing choice 2: Kit Wright: 'The Magic Box' and Miroslav Holub: 'The Door'

Both these poems had surface parallels: repeated sentence patterns, a refrain-like repetition, a surreal dream-like quality created by the crossover of abstract/concrete meanings. Wright's magic box contains the uncontainable; Holub's door opens onto the intangible. These poems led naturally towards dreams, hopes and memories.

Katrin explains in her commentary that the Wright poem 'allowed me to think that I wish I had this box in my life, or could change my life in some way. I kept the same structure and tried to follow this theme'.

Marie generated a 'dream-like' poem that echoed the patterns and repetitions of the two chosen poems. The main connecting theme was, as with Katrin, the resonance between poem and personal experience. 'My rewrite is based on my personal experience of being abroad, leaving my boyfriend for nine months'.

You put in my hand
A spoon of the softest Saharan sand
A piece of iceberg powdered with snowflakes
And a cloud full of rain, smooth and warm.

Pairing choice 3: Dinka praise poem, 'My Magnificent Bull' and 'O Behemoth' (Book of Job)

Both poems build up praise of the animal through a series of powerful hyperbolic metaphors, and do so with a repetition of sentence patterns that gives them an oral, incantatory quality. The students who chose this

pairing tended to remain with the praise of animals, rather than, for example, subverting the genre into a 'flitting' or hate-poem, or extending praise to people, objects, places or settings.

Elisa responded to the incantatory prayer-like qualities of the poems by evolving a 'praise/prayer' to the Elephant, using words from her childhood language, Welsh.

> Just as ''O Elephant'' uses ''ajanaku'', the Nigerian word for elephant, I thought I would give the Welsh words for God – ''Du'', thank you – ''diolch'': and Jesus – Iesu. ...Also ''sing my song'' refers to the fact that I am a singer but suffer quite a lot with nerves and praying gives me confidence that solos will go well.

Marking the Journey: Tutor Feedback

How easy was it to place these personally engaged responses within a framework of measurement and assessment? Clearly, the assessment framework was strongly criterion-referenced, and the clarification of language levels, and objectives within these, made the knowledge base explicit. The shared values as to the nature of the 'good' creative task had been made explicit, and these were used as a yardstick against which to measure our responses, and moderate one another's, during the feedback and assessment phase.

Scorer reliability

During the marking, we ensured that the following criteria were met:

- The two pieces of work written by each student were each marked by a different assessor.
- The two assessors standardized their marking by sorting the batch alphabetically, at both formative and summative stages, and marking the first and last three in each batch together.
- Wide differences in grading of an individual student across the two assignments were second marked (i.e. a difference of one band up or down).
- Distinctive cases were second marked, i.e. borderlines, distinctions and fails, as well as a sampling of students at the top, middle and bottom of the range.

With a shared understanding of objectives at each language level, and how these were to be applied to personal 'voice', we found there was rarely a discrepancy of more than 5% in a response to an individual piece of work.

Critical feedback

The complete cohort feedback was examined after the marking process of both assignments was complete, and a checklist identified of comments that appeared repeatedly or seemed characteristic of a lower assessment band. These comments fell into two clear groups: comments about language analysis, and comments about the creative process and clarity of reflection. Typical criticisms about the language analysis, found repeatedly in the assignments marked 55% or below, included:

- Lack of precision.
- Inaccuracy when describing linguistic features.
- Inaccurate use of terminology.
- Omission of core features in the published text, such as tone.
- Value-laden, judgemental responses.
- Inaccuracy in writing conventions.
- Gaps in the discussion (e.g. unsubstantiated generalizations).
- Misreading of the published text.

Typical criticisms about the student's creative response included:

(1) Mechanical response to the published text, such as manipulation of language without addressing meaning.
(2) Remaining too close to the original.
(3) Lack of authenticity and not exhibiting ownership of the theme.

Similarly, comments that emerged as characteristic of higher assessment bands offer insights into what was valued by the assessors. Positive feedback about language analysis found repeatedly with assignments marked 65% and above included:

- Using dialogue/varied voices to excellent effect to explore meanings.
- Providing good explanations of terminology and core values, and applying these to the texts.
- Revealing full awareness of the relationship between the original and the new text and of the influences of the original.

Positive feedback about the student's creative response included:

- Transforming the original text into something new, lively and interesting.
- Authenticity and engagement with content.
- Using strategies meaningfully to convey message.

In engaging with the criteria and applying this to actual student responses, several new questions emerged. Our initial assumption was that a choice of texts would offer a stimulus and support for creativity. However, students in the lower categories used these texts to imitate and manipulate, and did not spring out into any production of their own. Why? Are we wrong, perhaps, in believing that appreciation of other texts is an important starting point and springboard for creativity? What is the 'threshold level' needed for this process to work?

At the opposite end of the spectrum were students who responded zealously to the list of 'literary' and poetic features and reconstructed them fulsomely in their response. Yet the texts with the highest number of examples of 'literariness' were not necessarily also the most successful. How to explain and define this? Do other judgments override notions of successful 'literariness'? If so, what are these? Authenticity, integrity, engagement might be starting points for answering this. However, this leads to a third question: powerful personal engagement with the topic was not enough to be valued highly. In fact, in some cases, this powerful engagement actually led to weaker, *less* highly valued outcomes. Why? What are we learning here about the balance between an ego-centered creativity versus one that is 'adult' and publicly accountable? What is emerging about how these are valued, and how they interface with 'literariness'?

Analysing Results

Table 9.4 illustrates the following quantitative dimensions of the students' results:

- The relationship between formative and final assessments.
- The number of students whose results were in the top band (70% and over).
- The number of students who were in the bottom band (52% or under: borderline failure or fail).

This enables us to answer several interesting questions. Firstly, did our criteria for critical:creative response become more transparent to the students? 33% of students did improve between one assignment and the next, and among these, were just under one third of the international students and one third of the native speaker students. This suggests some progress that could be attributed to increased understanding of the task. A smaller percentage (19%) actually dropped a grade, with little significant difference between native and non-native English speakers. Again,

Table 9.4 Overview of student achievement in the language through literature module

	TOTAL	*Native speakers **	*Non-native * speakers*
Students who improved by one grade/band between the two assignments	35 (33%)	28 (34%)	7 (29%)
Students who dropped one grade/band between the two assignments	20 (19%)	15 (18%)	5 (20%)
Students whose final grade was over 70%	18 (17%)	10 (12%)	8 (33%)
Students whose final grade was 52% or under	11 (10%)	8 (10%)	3 (12%)
Total Student numbers 2005/2006	107	83 (78%)	24 (22%)

Note: * Percentages in these boxes are based on the native speaker/non-native speaker sub-totals.

reasons for this are merely speculative, but could suggest a difficulty in responding to the critical:creative challenge as its demands grew more explicit. In measurable terms, only 1 in 10 were assessed as failing in the process of transition from formative to final assignment: nearly one fifth were thought by both assessors to have been highly successful; another third made significant progress between the first and second assessment. These results do indicate that the formative stage had indeed provided a structure for development and that our concept of 'creative response' became more fully transparent to the learners so that we shared an understanding by the time we reached the final assessment.

Secondly, were non-native speakers advantaged or disadvantaged in any way by the nature of the assignment? Table 9.4 shows that the non-native speakers were significantly represented in the top band of achievement: 8 out of 18 first class results were non-native speakers. Among the native-speakers, two were in fact bilingual. Even more interesting is the fact that nearly one third of the non-native speaker group achieved at this highest level (compared with 12% of the native speaker group). Thus, far from arriving at the view that non-native speakers might be disadvantaged in comparison with native-speaking

literature students, it seems that the experience of language learning and knowledge of a second language were significant factors in success. Those learners who had accessed English through the metalanguage of grammatical and linguistic descriptions, were distinctly advantaged. Similarly, those students with experience of a second culture were more able to extract broader public and political themes from the chosen texts, and to engage with them personally.

Thirdly, was there any pattern among the students who struggled most in 'crossing the bridge' from appreciative reader to reflective writer? The figures suggest that the challenge of transition was similar for the native speaker and the non-native speaker (10% and 12%, respectively). From analysis of the feedback, it is apparent that the causes for failure were identical in both cases: either inaccuracy at the appreciation or analysis stage, or lack of engagement and authenticity at the creative response stage. 'Non-nativeness' was not significant at this level, compared to the highest achieving group, where a bicultural/bilingual experience actually appeared to contribute to success.

Conclusion

We now return to the questions posed in the opening section of this paper, and respond to them in the light of what has been discussed so far.

Are creativity and assessment not contradictory to one another?

This case study shows that an assessment exercise can also be an opportunity and an incentive for creativity. Students specifically praised the assessment cycle as the reason they enjoyed and benefited from the module. The early, formative stage offered the opportunity to understand better the difference between dependence and ownership; it allowed an opportunity for tutors to hone their definition of a 'creative' task, and students the opportunity to experiment with the relationship between critical and creative response. Thus, in the end, assessment provided an incentive and a framework from which creativity could emerge.

The cycle described here offers an exemplar of assessment processes as an opportunity to be 'creative'. It shows how, and with what implications, creative opportunity can be designed into the assessment task. The activity, in our example, offered both a firm grounding in shared knowledge, and a genuine invitation to 'make something new'

from this. Such an invitation requires particular vigilance, clarity and reflectivity on the part of the assessors. Specifically, it requires assessors: to make criteria transparent so that the assessment process is as accountable as any other; to make the activity transparent so that procedures are explained and achievable on the basis of given knowledge; to assess only what has been made explicit and not hidden or unarticulated values. While these are generic principles of effective assessment, they are specifically so in relation to creative activities. Making criteria transparent, for example, requires assessors to be clear about their definitions of creative process, the values they attach to creative outcomes, and the perceived weighting of received and shared knowledge, and the 'new'. Making activities transparent requires them to generate a 'scaffold' in which outcomes are achievable for all students and do not depend on individual imaginative leaps to interpret. In other words, assessment *for* creativity and assessment *through* creativity are real options, but they are ones to be developed responsibly, collaboratively and rigorously.

What notion of 'creativity' is being used here, and how is this skill or term anatomized so it makes sense to both learner and assessor?

We have seen that models of creative stages, of 'literariness' and of everyday linguistic creativity helped to shape the notion of creativity used in this assessment profile. What emerged from the cycle of teaching, feedback, analysis, assessment, is that there are indeed shared values and shared definitions of the term 'creativity', although even these are not without their complexities. 'Making something new' works well as a broad definition; however, the 'new' clearly needs to take account of the 'old' as well as the 'current'. Specifically, the 'something new' most highly valued is informed by peers and precedents; a generative 'something new' that involves the learner in appreciating texts more fully and being more confident in responding to them in their own voice.

How can this assessment be assured of transparency, objectivity and meaningfulness to the learner?

The transparency was established by 'thinking aloud' in the formative stage: what values about the creative:critical relationship are emerging from our formative comments? The linguistic/poetic scaffold upon which the course was framed became the framework, too,

for evaluation: has this 'scaffold' been understood by the appreciative reader, and applied by the reflective writer? Results suggest that, in 90% of cases, this progression from appreciation to writing was achieved, and in 33% of cases, the more detailed criteria helped students to improve between first and second assignments. Some aspects of the framework are still open to question: how do we account for the 'accumulation' of literary features that do not necessarily add up to literary quality? While advocating personal engagement in the topic, how do we account for the personally engaged texts that are in fact weaker because of this? The values by which we judge need to be continually re-examined, especially so where our declared values appear to be subverted by later judgments.

What can we learn about how native and non-native speakers respond to this process of transition, from critical to creative?

Student and marker feedback suggests that an accurate understanding of language systems and a receptiveness towards language creativity assists the process of transition towards reflective writing. Learners with experience of more than one language and culture seemed to be advantaged in having a higher likelihood of this openness. Providing a reflective framework on which to build, seemed to be helpful for the majority of the students, who were able to use this to 'cross the bridge' into creative production. Establishing a shared explanation for 'creative production' also assisted the process.

To summarize, the module did justify our belief that the creative process is accessible to learners across a wide spectrum of backgrounds, language levels and interests, and that providing linguistic and reflective tools allows for greater creative risks and opportunities. Making these 'tools' explicit and placing them within an assessment framework appears to have enhanced the process, at least for the majority.

We have seen the skepticism attached to the notion of creative writing in a second language, and the complexity of issues that need to be addressed in order to do this responsibly. However, to be focused on this only is to lose sight of what is different and important about 'expressive' as opposed to functional and transactional writing. 'The making of art enables individuals to ratchet up their ephemeral lives to the level of high symbolic adventure and philosophical questing' (Abbs, 2003: 7) This opportunity to 'make art' generates deep learning and active engagement. In a broader sense, this assessment cycle and its outcomes provide

further justification that 'making something new' should be recognized as achievable, measurable and central to our notion of meaningful learning.

Note

1. Students cited in this paper have given permission for their words to be used. All names are pseudonyms.

References

Abbs, P. (2003) *Against the Flow: Education, the Arts and Postmodern Culture*. London: RoutledgeFalmer.

ARG (Assessment Reform Group) (2002) Assessment for learning. On WWW at www.aaia.org.uk/pdf/Assessment_for_Learning.pdf.

Bloom, B.S. (1956) *Taxonomy of Educational Objectives: Classification of Educational Goals. Handbook I: Cognitive Domain*. New York: McKay.

Carter, R. (2004) *Language and Creativity: The Art of Common Talk*. London: Routledge.

Carter, R. (2006) Is there a literary language? In S. Goodman and K. O'Halloran (eds) *The Art of English: Literary Creativity* (pp. 84–88). Milton Keynes: Open University.

Cropley, A.J. (2001) *Creativity in Education and Learning*. London and Sterling: Kogan Page.

Goodman, S. and O'Halloran, K. (eds) (2006) *The Art of English: Literary Creativity*. Milton Keynes: Open University.

Hunt, C. and Sampson, F. (2006) *Writing: Self and Reflexivity*. Basingstoke: PalgraveMacmillan.

Hyland, K. (2003) *Second Language Writing*. Cambridge: Cambridge University Press.

Maybin, J. and Swann, J. (2006) *The Art of English: Everyday Creativity*. Milton Keynes: Open University.

Pope, R. (2005) *Creativity: Theory, History and Practice*. Abingdon: Routledge.

Vahapassi, A. (1982) On the specification of the domain of school writing. In A.C. Purves and S. Takala (eds) *An International Perspective on the Evaluation of Written Composition* (pp. 265–289). Oxford: Pergamon.

Vygotksy L.S. (1962) *Thought and Language*. Cambridge, MA: MIT Press.

Weigle, S.C. (2002) *Assessing Writing*. Cambridge: Cambridge University Press.

Weissberg, R. (2006) Scaffolded feedback: Tutorial conversations with advanced L2 writers. In K. Hyland and F. Hyland (eds) *Feedback in Second Language Writing* (pp. 246–265). Cambridge: Cambridge University Press.

Appendix 1: Texts in the Class Poems Handbook

Agard, J. (1982) Pan Recipe. In J. Agard *Man and Pan.* Cuba: Casa de las Americas.

Agard, J. (2005) Half-caste. In J. Agard *Half Caste.* London: Hodder and Stoughton.

Anon (1970) The Blacksmith. In K. Sisam (ed.) *Fourteenth Century Verse and Prose* (p. 169). Oxford: Oxford University Press.

Bateman, D. (1997) The Figs and Biscuits. In A. Gray (ed.) *Eating your Cake and Having It* (p. 88). Beaworthy, Devon: Fatchance Press.

Brown, A. (1998) Quote it's a man's world unquote. In A. Brown *West of Yesterday* (p. 48). Exeter, Devon: Stride.

Cope, W. (1992) The Uncertainty of the Poet. In W. Cope *Serious Concerns.* London: Faber and Faber.

Hopkins, G.M. (1953) Dun Scotus's Oxford. In W.H. Gardner *Poems and Prose of Gerard Manley Hopkins* (p. 40)*.* Melbourne and London: Penguin Books.

McGough, R. (1999) Streemin. In D. Goodwin *101 Poems that Could Save your Life.* London: HarperCollins.

Petrucci, M. (2004) Heavy Water. London: Enitharmon Press.

Updike, J. (2007) Superman. On WWW at http://www.dabush.org/Poems/Superman.html.

Appendix 2: Texts in the Assessment Handbook

Agard, J. (1985) Listen Mr. Oxford Don. In J. Agard, *Mangoes and Bullets: Selected and New Poems 1972–84.* London: Pluto Press.

Anon (1611) O Behemoth. *Book of Job, Chapter XL.* King James Bible.

Anon (1989) The Magnificent Bull. In R.B. Heath (ed.) *Tradewinds* (p. 104). Burnt Mill, Harlow: Longman.

Bennett, L. (1996) Colonization in Reverse. In J. Thieme (ed.) *Post-colonial Literatures* (pp. 844–845). London and New York: Arnold.

Cummings e e (2002) In Just. In *Norton Anthology of American Literature* (Vol. 2) (pp. 1449–1450). New York and London: W.W. Norton and Company.

France, L. (1998) If Love was Jazz. In N. Astley (ed.) *Poetry with an Edge* (p. 316). Newcastle upon Tyne: Bloodaxe.

Gorges, A. (1999) Her Face. In J. McRae and M. Vethamani (eds) *Now Read On: A Course in Multicultural Reading*. London: Routledge.

Harvey, M. (1999) What Colour's Your Inner Tube? In M. Harvey *Standing up To Be Counted Out* (p. 1). Dart Mills, Buckfastleigh, Devon: Dart Valley Press.

Holub, M. (1998) The Door. In N. Astley (ed.) *Poetry with an Edge* (p. 44). Newcastle upon Tyne: Bloodaxe.

Ibrahim, A.G. (1999) Hello. In J. McRae and M. Vethamani (eds) *Now Read On: A Course in Multicultural Reading*. London: Routledge.

Morgan, E. (2007) The First Men on Mercury. In *Learning Resources for Scottish Schools.* On WWW at http://www.ltscotland.org.uk/literacy/findresources/edwinmorgan/poems/index.asp.

Wright, K. (1987) The Magic Box. In K. Wright *Cat among the Pigeons.* London: Viking Kestrel.

Chapter 10

The Taming of the Immeasurable: An Empirical Assessment of Language Awareness

HUI-WEI LIN

Introduction

Assessment is an indispensable part of any teaching process in a classroom setting. It is important not only because the teacher needs to ascertain whether or not students have achieved a certain level of capability in the target language, but also because the learners need to know how they are progressing. These demands, however, pose special problems when it comes to the practice of a literary education (Alderson, 2000; Brumfit, 1991; Hall, 2005; Purves, 1991; Strelka, 1969). Part of the problem of literary evaluation is the difficulty associated with investigating personal reactions and responses to texts. Since one cannot measure various kinds of responses to literature in absolute terms, there is no easy or convenient way to evaluate learners' capacities for dealing with literary texts. This brings us to the question of how the subjective element of literary understanding can be reconciled with the objective requirements of language testing. This chapter seeks to address this dilemma and suggest likely solutions to these questions.

My own interest in the evaluation of students' progress as readers of literature grew out of my experience of teaching Shakespeare. My interest has been the assessment of the progress made within the course, where a stylistic exploration of Shakespeare's language was intended to enhance students' language awareness. This pedagogical stance sub-scribes to 'the *use* of literature as a resource for language learning' propounded by Maley (1989: 10) and other scholars who advocate the study of language and literature as naturally intertwined (e.g. Carter & Long, 1991; Carter & McRae, 1996; Montgomery *et al.*, 2000; Simpson, 1997, 2004; Short, 1996 among others). In effect, I was also seeking to

answer a series of research questions raised by McCarthy and Carter (1994: 169), which has not yet been answered by the testing or related community: 'Should learning *about* language be assessed? Can it be effectively measured? If so, how is this best done? What are the advantages and disadvantages to such an approach?' My point of departure will be a brief description of the study as a whole and the purposes for which I developed a language awareness test following a growing body of studies (e.g. Alderson *et al.*, 1995; Baume *et al.*, 2004; Birenbaum, 1996; Douglas, 2000; Turner & Upshur, 1995, 2002; Weigle, 2002) on educational measurement and test development.

Context of the Present Study

The study reported here took place in a Taiwanese context where students are often tested on their knowledge of facts about literature, but very little on their knowledge about language of the texts. Questions found in examinations of literary understanding are often detached from the language of the literary works, and can readily be answered by reading the texts in translation or by reading lecture notes without direct reference to the texts (Liao, 2004; Lin, 2005). The study of literature is often characterised by a content-based, desk-bound and teacher-centred methodology. The classroom dynamic is a transmission-reception one, where the teacher functions as the authoritative expert and transmitter of knowledge conveyed to students in the form of 'background' to be memorised and reproduced when the examinations require it (Carter & Walker, 1989; Erbaugh, 1990). It is no wonder that after several years of training and studying of literature, a gap still exists between the level of linguistic/literary competence and the level or standard required of students to study literature.

To test my own judgement that EFL learners could derive greater benefits from literary education through stylistics, I used sonnets and play extracts by Shakespeare as a medium to sensitise students to various linguistic devices and their literary functions in a second-year under-graduate class for English majors at National Kaohsiung First University of Science and Technology.

An overview of the instructional programme

Twenty-two Taiwanese students, who enrolled in this second-year undergraduate class, participated in a 10-week period of intervention. Students ranged from 21 to 28 years of age, with 3 males and 19 females

in the group. This ratio reflected the preponderance of females in the English Department at the time.

The pedagogy implemented in the intervention is grounded in Scholes' (1985: 24) view that the teacher's job is 'not to produce readings for our students but to give them the tools for producing their own'. This viewpoint echoes McRae and Clark's (2004: 336) proposition that 'the linguistic tools of stylistics are precisely what EFL/ESL learners need in order to develop their approaches to reading any text, be it literature with a small "l" or institutionally-defined literary study'. Out of such considerations, the pedagogical framework encompassed seven self-contained but interrelated units, whose primary functions were to help students learn, through Shakespeare, how it is possible to explore and exploit the resource of language in original ways, displaying its range and variety in the service of the poetic imagination. The stylistic features were taught in the sequence illustrated in Table 10.1.

My reason for choosing these stylistic devices was that they are recurring linguistic patterns common to literary texts. Illustrated extensively and substantially in many stylistic textbooks and in discussions related to stylistic analysis, these style features may be claimed justifiably to be the most practical and effective tools in putting students on the inside of 'representational texts' (McRae, 1991). From a methodological perspective, language awareness is operationally defined as an awareness of this definite set of stylistic devices. It is therefore construed here as the ability to *identify* the stylistic features in Table 10.1 and account for their representational significance (or effects) in (literary) texts.

Table 10.1 A summary of the instructional programme

Level of language	*Schedule*	*Focus of unit*
Sound	Week 1	Rhythm & metre
	Week 2	Alliteration, assonance & rhyme
Semantics	Week 3	Binary oppositions (e.g. bright vs. dark)
	Week 4	Repetition
	Week 5	Metaphor & imagery
Syntax	Week 6	Antithesis
Context	Week 7	Voice

Developing the Language Awareness Test

The purpose for which the test was developed

To go hand in hand with what I was teaching the works of Shakespeare for, a test for the evaluation of language awareness was necessary. However, the growth of Language Awareness over the past several decades has not been accompanied by a strong and productive interest in issues of its assessment. Surveys of several databases in related fields reveal virtually no papers concerning the assessment of language awareness or of literary competence. Consequently, an innovative test for this specific purpose was called for. Such a test needed to fulfil the dual functions of assessing students' level of language awareness *before* the class and of verifying whether or not their language awareness improved *after* the class.

Test design and specifications

Bachman (1990) and Bachman and Palmer (1996) provide a useful framework for test development. The framework sets out to describe five aspects of test tasks: setting, test rubric, input, expected response and the relationship between input and response. I will use them as a basis for the discussion of the Language Awareness test constructed here.

(1) Characteristics of the setting

Physical setting can be an important variable in a test. So, to have the setting controlled, the test was taken in a well-lit classroom containing 40 movable student desks and a teacher's table with a lectern. As the classroom is where students regularly attend various courses, it is a familiar environment for each test taker. In general, the test was conducted in favourable physical conditions for a testing situation. However, the time of the test was 1 pm, a time when some test takers might feel drowsy after lunch, as some of them reported afterwards.

(2) Characteristics of the test rubrics

Students' knowledge *about* language was measured by means of a pre-test, which served as a preliminary diagnosis of the students' starting-points of language awareness. The pre-test was administered 1 week before the intervention, and the post-test, 3 weeks after the intervention, with a gap of 10 weeks between the two tests. The two tests contained exactly the same text and test tasks, but students were not informed that they would be doing a test after the instructional programme, nor did they know the test would be the same.

(3) Characteristics of the input

The 'input' in a test refers to the text material that test takers are expected to process and respond to. In the present context, rather than being an extract from Shakespeare, the input was a car advertisement titled 'More Pulling Power – The New Terrano II' (Figure 10.1).

The text was chosen as the input because it contains many stylistic features obviously worthy of comment. Many of these style features – alliteration, antithesis and metaphor – are what students learned in the course. The text was therefore suitable for the purpose of verifying whether or not students can apply 'the linguistic tools of stylistics' to a new text or to the world of language and literature around them. Although a picture accompanies the Terrano advertisement, it was not presented to the test takers because such combinations in text may lead to different reading experiences (Alderson, 2000).

Prior to the pre-test, participants were offered necessary information as to what the test would look like: what the test was aimed at (in general

---

More Pulling Power－The New Terrano II

Power, so they say, is an aphrodisiac. And you can't get much more powerful than the new Terrano II. Beneath that raunchy new exterior lies a 2.7 litre turbo diesel intercooler engine that delivers extra horsepower at low revs, just when you need it. Switch to 4 wheel drive and you'll be down that rocky gorge (a doddle with the limited slip differential), through that stream (up to 450mm deep) and up that mountain (up to 39 degrees) with something close to wanton abandon.

But perhaps you should begin with something more gentle. Like tarmac. For it's here that you'll appreciate the Terrano's more sensitive qualities. Its responsive handling. Its smooth as a satin sheet ride. (Thanks to anti-roll bars). The ease with which it will tow a trailer, pull a caravan, and more or less anything else you fancy. Add ABS and driver's airbag and you have a 4×4 that's too desirable for words. Which is exactly why it comes with engine immobiliser and an ultrasonic and perimetric alarm.

YOU CAN WITH A NISSAN

(Text taken from Thornborrow & Wareing, 1998: 204)

Figure 10.1 The input (text used in the language awareness test)

terms), what content was to be covered, what the methods would be, how long it would take, and other neutral information that ensured familiarisation with the test prior to taking it. In so doing, students were provided with enough information and psychological readiness for the test to enable them to perform the test to the best of their capacity (Alderson *et al.*, 1995). The instructions were conveyed in the students' first language (i.e. Chinese) in order to ensure that they understood how the tasks were to be completed. Access to bilingual dictionaries was allowed, as this would normally be available to readers under natural study conditions. Students were given about 80 minutes to complete the tasks.

The first task aims to probe students' affective responses to the Terrano advertisement. A semantic differential, as now used, is a type of rating scale designed to measure the connotative meaning of objects, events or concepts. A scale like this measures directionality of a reaction (e.g. good versus bad) and also intensity (slight through extreme). Hence, it was useful to use this method to draw up a 'map' of students' connotations for the given text. In this part of the test, the students were asked to choose, as intuitively as possible, where their position lies on a number of seven-interval scales defined with bipolar adjectives at each end (see Figure 10.2).

The students' ratings – a number from 1 to 7 – produce a measurement of their responses to the Terrano advertisement. Ratings can be combined to quantify personal feelings, making it possible to study changes of meaning in both individuals and groups. In the present study, scales were constructed by consulting Snider and Osgood's (1969) sourcebook for adjectives that could tap informants' affective reactions to the text. Caution was taken to avoid guiding students' attitudes towards the text. The instructions given were brief and neutral: 'Read the text, see what impression it makes on you, and then rate the scale'. Neither the purpose nor the rationale for doing this were given. This allowed students to decide for themselves about the nature of the Terrano advertisement and their evaluation of literary or non-literary reading.
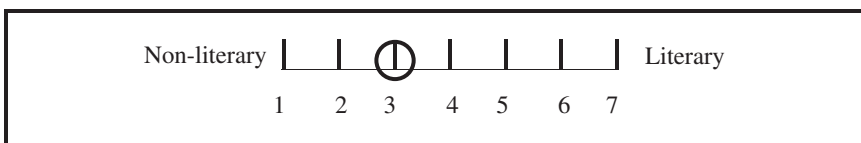
Figure 10.2 A sample response to a semantic differential scale

Upon completion of the semantic scales, students were given the awareness tasks to accomplish. Three consecutive tasks were set. The first was a detection task, in which students were asked to observe whether there are stylistic features embedded in the Terrano text. To do so, students had to tick the boxes provided, with each one representing one particular textual feature. Next, for each of the features they ticked, students were asked to support their judgement by indicating where in the text this feature occurs. This response type requires the student to provide the actual location of a stylistic feature in the given text. This procedure is intended to confirm students' ability to *notice* and *identify* a particular feature in the text, and reduce the effect of guessing. This identification task lays the groundwork for further interpretations to be made by the students.

Finally, students were invited to think (creatively) about the significance of each stylistic feature they had identified. In this part, they were responsible for the wording of the answer by providing a concise and cogent interpretation as to *how* each stylistic feature contributes to the meaning between the lines in the Terrano text. This higher order task has the advantage of not constraining the candidate to the same degree as the previous tasks. Another advantage, in McNamara's account (2000: 30), is that the candidate 'assumes greater responsibility for the response, and this may be perceived as in some ways more demanding and more authentic'.

(4) Characteristics of the expected responses

As this test was aimed at assessing language awareness, not testing grammatical knowledge or language proficiency, students were free to respond to the test items in either Chinese or in English as they preferred. This was in order to ensure that the respondents' writing ability should not inhibit the demonstration of their language awareness. For this reason, candidates were not penalised in this test for weakness in writing skills. This decision is also related to an important issue of validity in scoring. If the scoring of the written responses takes into account spelling and grammar, it makes the measurement of language awareness less accurate, because the task then measures more than one ability. Writing ability should not be considered as a variable in a test like the case in point (see Hughes, 2003; Spiro, 1991, for a detailed account of this issue).

(5) Relationship between input and response

How input and response are related to each other can be discussed in terms of the reactivity, scope, and directness of the relationship. First, test tasks in the present test are 'non-reciprocal', where there is neither

feedback nor interaction between test takers, as is usually the case in a reading comprehension test. Second, the amount of input needed to be processed by the test takers is characterised as 'narrow scope', since they were required to read only one passage (i.e. the Terrano advertisement) in order to do the tasks. Third, the expected response is based primarily on information in the input, and the tasks involve a 'direct' relationship between input and response.

By way of a summary, characteristics of tasks in the language awareness test are presented in Table 10.2.

Designing a rating scale

Because of the open-ended questions in Part III – questions to which there are no fixed or absolute answers – the assessment procedure is bound to be more complicated than a mere calculation of a tally of the 'right-wrong' answers. As with any performance-based language assessment, the test constructed here faces a challenging conundrum – the establishment of procedures that will quantify a more objective observation of testers' performance. For this obvious reason, most performance-based language assessment, such as the Speaking and Writing modules in the International English Language Testing System (IELTS) test, typically requires examiners to judge the quality of examinees' written/spoken language in relation to a rating scale. According to Gannon and Ostrom (1996: 337), the use of a rating scale to define the construct of language proficiency can 'often minimise difficulties associated with the coding and quantifying of a respondent's answers'. For similar reasons, the test here needs a structured scoring guideline for making the criteria for success more explicit, more systematic and more objective. By dint of this procedure, the examiners can identify and mark the important characteristics of the response, thereby making more valid measurements of testees' performance.

The rating scale was devised according to the EBB scale development procedure proposed by Turner and Upshur (1995, 2002). EBB refers to an *Empirically* derived scale which requires *Binary* choices by raters and defines the *Boundaries* between score levels. This resulted in an eight-point rating scale representing eight levels of language awareness (see Figure 10.3). This scale ensured that different raters would follow the same criteria to locate each candidate's performance and arrive at a more objective rating. Here, criteria for 'correctness' were based on the number of stylistic features identified and on how those identified features were interpreted for their representational effect(s). Candidates'

Table 10.2 Task characteristics (based on the framework of Bachman and Palmer, 1996)

Task characteristics	*Test tasks*
The setting	
Physical characteristics	Typical classroom (a familiar place for test takers)
Participants	22 Taiwanese university students of EFL
Time of testing	13:10–14:30
Test rubrics	
Instructions	
Language	Chinese & English
Channel	The instructions are presented both aurally and visually
Specification of procedures	Tasks are clearly and explicitly specified in the test
Structure	
Number of parts/tasks	3
Salience of parts/tasks	Different parts of the test are clearly distinguished from one another
Sequence of parts	Part I. Semantic Differential
	Part II. Detection & identification of stylistic features
	Part III. Interpretation of the significance of features
Number of tasks/items per Part	Part I (7 items)
	Part II (7 items)
	Part III (number of items varies according to the previous items identified)
Time allotment	80 minutes for the entire test
Scoring methods	
Criteria for correctness	The correctness of response is determined by the number of stylistic features being identified
Scoring procedure	Blind double marking; scorers are told to ignore inaccuracy (e.g. spelling, grammar) in responses
Explicitness of criteria and procedures	Test takers are not informed about the evaluation criteria
The input	
Format (e.g. form, length, type, etc.)	Presented as a written text in English, the input is an extended discourse of 183 words; the type of input is a combination of 'item' that elicits a selected response and 'prompt' that elicits a specific, extended response

Table 10.2 (*Continued*)

Task characteristics	Test tasks
Language of input	Advertisement, literary discourse
The expected responses	
Format (e.g. language, type of response)	Test takers must select one response from among several given choices; based on the selected response(s), test takers then provide an extended production response.
Relationship between input/response	
Reactivity	Non-reciprocal
Scope of relationship	Narrow
Directness of relationship	Direct
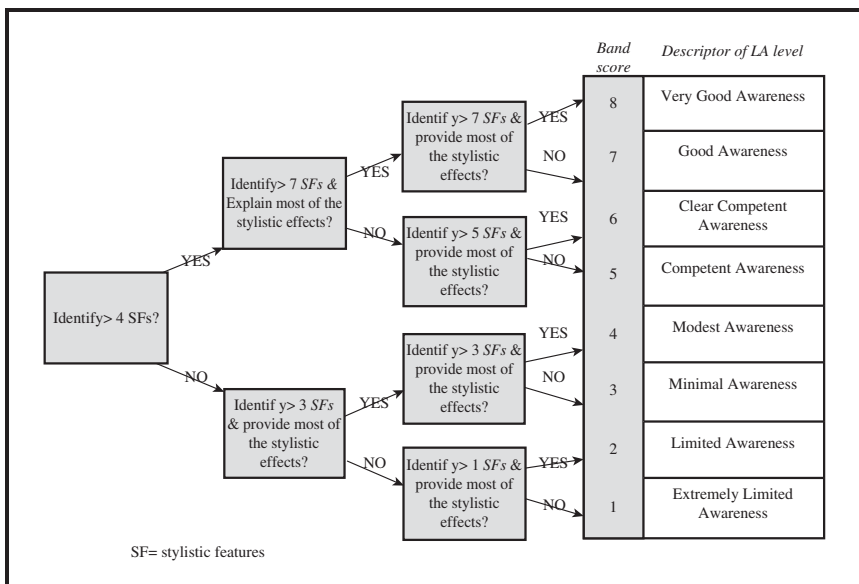
Figure 10.3 Empirically-based rating scales

performance could be rated on a scale of 1–8 (from 1 = Very limited Awareness to 8 = Very Good Awareness).

Accompanying the scale is a descriptive benchmark giving a summary of a candidate's awareness classified at that level. The levels of awareness are described in Table 10.3.

Table 10.3 Interpretations of benchmarks

Levels of achievement	Band descriptors
8 = Very Good Awareness	Candidates in this category: – Show strong sensitivity to and understanding of the given text. – Fully identify stylistic features, along with provision of convincing and creative responses to how each feature is employed for representational effect(s). – Use clearly appropriate details to support arguments.
7 = Good Awareness	Candidates in this category: – Show good sensitivity to and understanding of the given text. – Identify nearly all of the stylistic features, along with provision of convincing responses as to how each feature is employed for representational effect(s). – Use appropriate details to support arguments
6 = Clear Competent Awareness	Candidates in this category: – Locate and identify at least 6 features in the given text. – Explain how stylistic features are used to convey a non-literal message in a sensible way.
5 = Competent Awareness	Candidates in this category: – Locate and identify at least 5 stylistic features in the given text. – Explain how stylistic features are used to convey a message, while providing a few general comments.
4 = Modest Awareness	Candidates in this category: – Are able to recognise and to identify at least 4 stylistic features. – Attempt some justification for the features being identified.
3 = Minimal Awareness	Candidates in this category: – Are able to recognise at least 3 stylistic features. – Attempt some descriptions of the features being identified, though candidate is likely to make some general remarks.
2 = Limited Awareness	Candidates in this category: – Correctly identify at least two formal features. – Attempt responses to the feature identified, though only to provide a general commentary.

Table 10.3 (*Continued*)

Levels of achievement	Band descriptors
1 = Extremely Limited Awareness	Candidates in this category: – Identify less than two stylistic features. – Provide few comments on the stylistic effects. – Essentially have no ability to justify the use of any language feature.

Note: These categories are best viewed as tendencies rather than as absolutes.

The rating procedure

As a degree of judgement is called for on the part of the scorer in marking open-ended responses in this test, a perfectly consistent scoring pattern could not be expected. Therefore, some procedures were taken to finalise the marking scheme and to ensure greater scorer reliability. First, all the tests were scored independently and anonymously by two examiners using consultation with the benchmark. Although research literature (e.g. Hughes, 2003; McNamara, 2000; Weigle, 2002) has quite consistently shown higher scorer reliability when writing is scored four times, it was decided for practical reasons to adopt an independent double scoring method. The two raters are experienced teachers who hold masters degrees in English language teaching. The two ratings given by the two teachers were averaged. If these two raters scored more than 2 points apart, the test would, in turn, be scored by a third examiner (myself), and the two closest total scores averaged to arrive at the final test score. In addition to this procedure, a detailed scoring key was provided that specifies possible answers and assigns points for valid responses given. If the scorers faced difficulties in the assignment of points, they were asked to bring these to the attention of the researcher (myself) for a consistent decision to be made.

One important issue related to scoring is that linguistic forms may be pragmatically as well as semantically ambiguous, so there is not an invariable relationship between form and function. Carter (2003: 65) highlights this point when he notes, 'Appropriate assessment of language awareness is less likely to involve correct production than to elicit the learner's ability to explain how particular forms function'. Because of the noted lack of co-variation of linguistic forms and functions, a reader cannot just recognise a particular stylistic feature and assign a significance to it without referring to the unified whole of text and context. This lack of consistency also accords with van Peer's (1990: 262)

postulations that any stylistic devices 'simultaneously fulfil different functions' and therefore must be 'understood as multi-dimensional'. The understanding that a stylistic device may carry different representations in different contexts requires that we take into account these variations as an important determinant of students' testing performance. On the basis of these considerations, difference in test performance is analysed and assigned to different levels on the EBB scale.

Test Results

Affective attainments

Interpretation of semantic differential scale

With respect to the first task, which aims to tap into students' perceptions of the content of the Terrano text, I used Osgood's (1952) measurement technique in an attempt to assess attitude change as a result of the treatment. Twelve bipolar adjective pairs were provided for the respondents to rate each scale. Then the ratings were averaged on each dimension. In this manner, the profile of ratings is obtained by summating the ratings on the 12 scales used. Figure 10.4 graphically represents students' affective response to the stimulus (i.e. the Terrano text). Here, the scales have been rearranged directionally, so that adjectives representing negative values are situated systematically on the left hand side, whereas those representing positive values are on the right hand side.

As can be seen, the scale differentiated attitudinal intensity based on students' subjective responses to the text. The two response profiles from a test and re-test spaced 10 weeks apart are somewhat different in terms of both shape and elevation. This change indicates that the Terrano advertisement is evaluated more positively by students on all scales. It was judged as more *literary*, *comprehensive*, *exciting*, etc., than it was originally perceived in the pre-test. This is presumably because that attention to textual features, owing to the students' enhanced awareness of their existence, opened up the students to different interpretative possibilities and enriched the re-reading process. Awareness of the existence of these features not only advanced students' insight into textual comprehension, i.e. scale 2 ('comprehensible'), but also enhanced the aesthetic pleasure experienced in reading the text, e.g. scale 3 ('pleasant') and scale 8 ('attractive').

One of the crucial findings revealed by the scale is therefore that interpretation of a text is both (or at least) an interplay of text characteristics
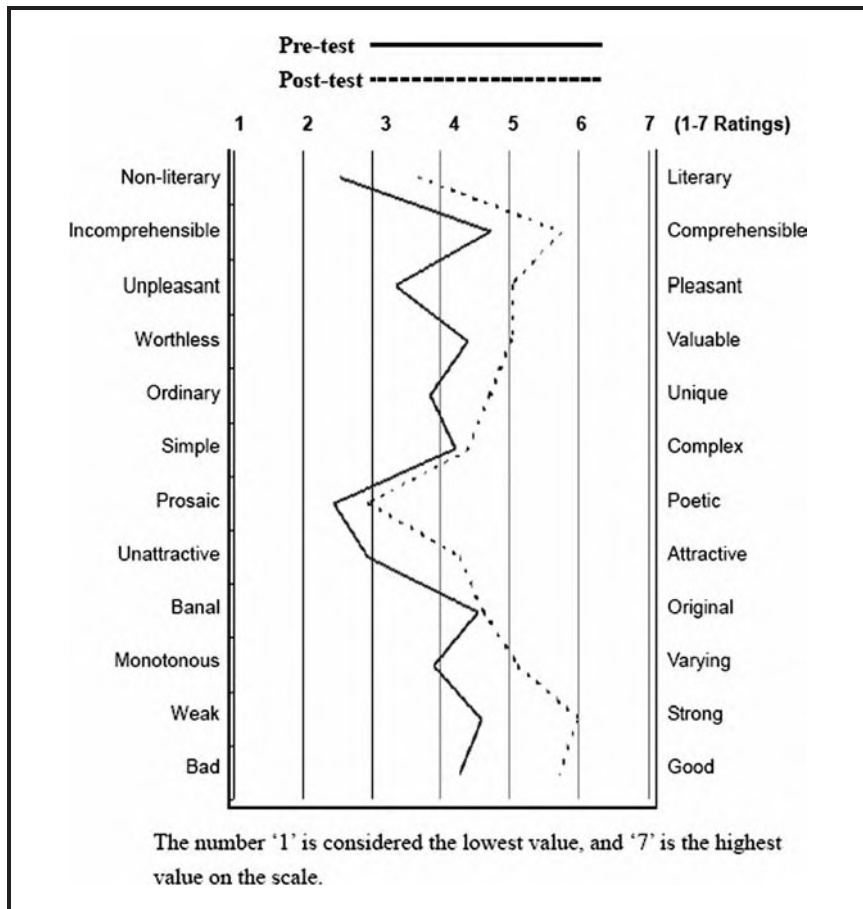
Figure 10.4 A comparison of affective reactions to the Terrano text

and the reader's psychological/cognitive/affective construction, a point well discussed by van Peer (1990) and Miall and Kuiken (1998).

Cognitive attainments

This test was administered as a pre- and post-test so that comparisons could be made to see whether there was any difference between the two tests given. Overall, the results indicated improvements in all students at the end of instruction. While most of them (86%) typically fell below Level 3 (Minimal Awareness) in the pre-test, post-test results showed that

their level of awareness was markedly enhanced. The results for the 22 students are shown in Table 10.4.

As can be seen, all participants were re-allocated within Level 4–7 of awareness (Modest to Good), with four students (i.e. 18%) achieving Level 7 and the five weakest students (i.e. 23%) achieving Level 4. This can be compared with the original distribution of awareness level from 1 to 4 (Extremely Limited to Modest) in the pre-test. By and large, the overall test performance advanced from an average of Level 2 in the pre-test to Level 5 in the post-test. A substantial difference between test and re-test reveals a solid progression – from Limited Awareness to Competent Awareness. What has to be emphasised here is that a greater proportion of students, who in the pre-test noticed no stylistic features at all, was able to identify some and provide quite a few insightful comments on how these features work to express or symbolise the content or propositions of the text. Some of the students' sample responses are provided in Table 10.5 to show the qualitative changes in their language awareness.

Table 10.5 presents the extended verbal responses based on five different stylistic features, comparing the pre- and post-test comments written by eight different students. For example, Zen initially failed to identify the metaphor – 'Power is aphrodisiac' – in the pre-test. As a corollary of his inability to notice this feature, no explanation as to its significance could be provided. After all, only linguistic information that is detected can undergo further processing (Schmidt, 2001). Reading the

Table 10.4 Pre-test and post-test levels of awareness

Level of awareness	Pre-test (n = 22)	Post-test (n = 22)
8 Very good Awareness	0	0
7 Good Awareness	0	4 (18%)
6 Clear Competent Awareness	0	3 (14%)
5 Competent Awareness	0	10 (45%)
4 Modest Awareness	3 (14%)	5 (23%)
3 Minimal Awareness	2 (9%)	0
2 Limited Awareness	11 (50%)	0
1 Extremely Limited Awareness	6 (3%)	0

Table 10.5 Samples of extended responses provided in pre- and post-tests

Stylistic Features (Student)	Limited production responses provided	
	Pre-test	*Post-test*
Metaphor		
(Zen)	(Not identified)	*'Power is aphrodisiac'. Cars are like women. Driving the Terrano is like having the power to control women, have the 'extraordinary' ability to enjoy a sexual relationship with a woman.*
(Webb)	*I think the purpose of it is to imply buyers to buy the car.*	*It implies readers if you want more power you may consider buying the new Terrano II.*
Sound Patterning		
(Sally)	(Not identified)	*The sentence 'Its smooth as a satin sheet ride' abounds the sound of 's' feel very smooth, which reflects actually the ride of the car.*
Binary Oppositions		
(Pamela)	(Not identified)	*Stream and mountain are totally different. It gives the readers that the car can run on any terrains.*
(Veronica)	*Compare two opposite aspects, make the text stronger*	*'. . .that rocky gorge; something more gentle. Like tarmac'. 'up . . .and down': emphasize the car can go anywhere.*
Antithesis		
(Grace Chen)	(Identified, but no comment provided)	*'. . .more or less', like binary's function, shows the differences and flexibility the car has.*
Voice		
(Shirley Wu)	*It can persuade the reader.*	*There are many 'you' in the article. It is like there is a person talking to you like a speech in front of the audience. That can shorten the distance between the author and the reader.*

Table 10.5 (*Continued*)

Stylistic Features (Student)	Limited production responses provided	
	Pre-test	*Post-test*
(Phoebe)	*Let readers get into the article easily and it seems like in the same situation with the speaker.*	*Promote the product and feel there is a sales person telling us the wonderful product.*

Note: All names are pseudonyms; the student 'voices' are cited in their original forms.

same text in the post-test, Zen, along with other students, not only noticed the metaphor but provided an interpretation of *how* it works. He saw that there was more to the meaning of the 'extra horsepower' than simply the literary meaning. The metaphor invoked a sexual interpretation, making the sensation of driving the Terrano ever more sensational. This sensation, according to another student, Sally, is also implicated in the sound patterning of the text. Sally noticed the sibilant sounds in the easy flow of the line: 'It*s s*moo*th* a*s s*atin *sh*eet ride', which is directly associated with the car's 'responsive handling' and 'sensitive qualities'. The ability to discuss how specific linguistic structures are used to produce particular effects is indicative of enhanced language awareness. As Hanauer (1999: 21) notes, development of language awareness can be manifested in 'an increase in a literary student's ability to selectively focus on, use and explicitly discuss specific aspects of a literary text for interpretation purposes'. The responses in the post-test, while hardly mature literary analysis, reveal the development of an ability to find a consistent pattern of meaning and to communicate that pattern to others. From my viewpoint as a teacher of literature, I was delighted to see students apply the know-how in analysing a text under scrutiny.

Conclusions and Implications

Traditional literature assessments often focus on literary comprehension and a set of dead facts that can be easily scored and measured (Brody *et al.*, 1989; Carter & Long, 1990; Purves, 1973, 1992, 1993). As such, an essential element in most literature examinations is the retrieval of memorised information. There is relatively little attention paid to such complex constructs as *sensitivity* to the linguistic components of literary achievement and *awareness* of how language works in literary works. Purves (1990) rightly points to the alarming fact that there has been altogether too little research effort devoted to valid measures of the

readers' awareness stimulated by the power of literature. If this kind of awareness is argued to be critical to learning literature, then it must become one of the goals of literature education. Continuing this line of thought, we need appropriate measures that better match what we want students to gain from their literature experiences and how we assess those gains. By constructing a language awareness test, I have tried to experiment with a more authentic assessment of student learning, giving students the opportunity to show what has been learned, how that learning has been applied and how that learning can be expressed. In developing this type of measure, however, challenges have begun to emerge: to move away from modes of assessing literary reading through comprehension questions or literary criticism, the teacher must give a more appropriate assessment of the reader's response to a literary text; to maintain authenticity, the teacher must design test tasks that can capture or recreate the essence of the real literary experiences. To stand scrutiny, the assessment must provide both criterion-based standards and judgement reliability that may not be required by commonly found literature testing in the classroom. As a natural corollary, designing and implementing such assessment approaches is extremely time consuming.

Due to the scope of the study, the test designed for this project did not involve the assessment of global ability. Since the target and focus of language awareness is broad, encompassing not only linguistic domains but also sociolinguistic and cultural domains, some interesting dimensions are left unexplored and could have called for a new study. Some of the limitations of this study lead to various suggestions for future research. With regard to test development and refinement, a few suggestions can be made.

First, a growing body of recent research (e.g. Alderson, 2000; Bachman, 1990) has demonstrated that the test methods we use to measure language ability influence performance on language tests. From the standpoint of content validity, an ideal test would be one that selects a representative set of tasks. One reason for this, as Hughes (2003: 86) has pointed out, is that 'the more tasks that we set, the more representative of an examinee's ability will be the totality of the samples we obtain'. By doing this, we can avoid the chance of choosing the task(s) that a candidate is particularly good (or bad) at. Clearly, a number of test methods have gone unexplored in this project. The implication for testing of learning in literature is that we need to include more distinct elicitation procedures (e.g. filling-in-the-blanks, matching, sentence completion, multiple-choice, short essays) that can tap, from various

perspectives, the learners' capacity to handle responses – linguistic, emotional, intellectual – to literary texts.

Second, we may avoid practice effects and enhance test reliability by using a parallel text as the input in the language awareness test. Ideally, it is more desirable to use two different texts of equal complexity and difficulty in the pre- and post-test, respectively. One method of doing this is to administer to a control group a test that is essentially two tests, where each question on the one has a parallel question on the other. When the test is scored, the results of both halves can be compared. If the results attained by each student on each half of the test are acceptably similar, the test can be said to be reliable. For this reason, the teachers, along with 'instructor judgment' (Schulz, 1981: 43), need a procedure that allows them to make reasonably objective predictions about the difficulty of the assigned texts. To obtain a rough predictor of readability of a text, the use of a statistical device such as a readability formula may be helpful.

Furthermore, given that one of the fundamental variables in authentic assessments is the rater, it is important to focus not only on the unobservable construct being assessed, but also on the judgements and decision making of the raters (Wigglesworth, 2008). In order to obtain reliable ratings or consistent test scores, we may develop an EBB scale to ameliorate reliability and validity problems. In its emphasis on explicit binary choice relating to student performance, an EBB scale can guide raters to follow precisely the same standards of grading and thus achieve higher reliability. Also, since the criteria in an EBB scale are empirically derived from actual student test performance, it can be seen as a way of achieving a close correlation between the test performance and the criterion performance, thus ensuring the validity of the assessment. It should be noted, though, that the teacher must define the goals of measurement, consider the construct that the scale measures, and use the scale within its defined capabilities.

From a pedagogical perspective, literature should be presented in a way that is directed towards a development of knowledge *of* literature rather than knowledge *about* literature (Carter & McRae, 1996). The important point is that, as traditional literature assessment format causes teachers and learners to emphasise isolated and discrete facts, teachers of literature must now seek alternative assessments that free learners to develop higher level concepts and a deeper level of understanding. The very nature of these new assessment approaches is the potential catalyst for the needed change, capable of shifting the focus of instruction from factual knowledge *about* literature to advanced

knowledge *of* literature. As such pedagogical approaches are widely adopted and the basic principles of testing (i.e. we should test what has been taught) followed, there will be a growing tendency for literature assessment to open up newer modes of assessing student achievement in the study of literature.

References

Alderson, J.C. (2000) *Assessing Reading*. Cambridge: Cambridge University Press.

Alderson, J.C., Clapham, C. and Wall, D. (1995) *Language Test Construction and Evaluation.* Cambridge: Cambridge University Press.

Bachman, L.F. (1990) *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, L. and Palmer, A.S. (1996) *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

Baume, D., Yorke, M. and Coffey, M. (2004) What is happening when we assess, and how can we use our understanding of this to improve assessment? *Assessment and Evaluation in Higher Education* 29, 451–477.

Birenbaum, M. (1996) Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum and F.J.R.C. Dochy (eds) *Alternatives in Assessment of Achievement, Learning Process and Prior Knowledge* (pp. 3–29). Boston, MA: Kluwer.

Brody, P., DeMilo, C. and Purves, A.C. (1989) The current state of assessment. *Report Series 3.1*. Albany, NY: National Research Centre on English Learning & Achievement.

Brumfit, C. (ed.) (1991) *Assessment in Literature Teaching.* London: Macmillan Modern English Publications.

Carter, R. (2003) Language awareness. *ELT Journal* 57 (1), 64–65.

Carter, R. and Long, M. (1990) Testing literature in EFL classes: Tradition and innovation. *ELT Journal* 44 (3), 215–221.

Carter, R. and Long, M. (1991) *Teaching Literature.* Harlow: Longman.

Carter, R. and McRae, J. (eds) (1996) *Language, Literature and the Learner: Creative Classroom Practice.* London: Longman.

Carter, R. and Walker, R. (1989) Literature and the language learner: Introduction. In R. Carter, R. Walker and C. Brumfit (eds) *Literature and the Learner: Methodological Approaches.* ELT Documents 130. London: Modern English Publications.

Douglas, D. (2000) *Assessing Language for Specific Purposes.* Cambridge: Cambridge University Press.

Erbaugh, M.S. (1990) Taking advantage of China's literary tradition in teaching Chinese students. *The Modern Language Journal* 74 (1), 15–27.

Gannon, K.M. and Ostrom, T.M. (1996) How meaning is given to rating scales: The effects of response language on category activation. *Journal of Experimental Social Psychology* 32, 337–360.

Hall, G. (2005) *Literature in Language Education.* New York: Palgrave Macmillan.

Hanauer, D. (1999) Attention and literary education: A model of literary knowledge development. *Language Awareness* 8 (1), 15–29.

Hughes, A. (2003) *Testing for Language Teachers* (2nd edn). Cambridge: Cambridge University Press.

Liao, P.J. (2004) Barriers to literature study: A pedagogical analysis of the problems in the teaching of English literature to Taiwanese students, based on Jane Austen's Pride and Prejudice. PhD thesis, University of Nottingham.

Lin, H. (2005) The application of literary texts within ESP/EAP methodologies: Shakespeare in the university composition courses. Conference paper presented at the 2nd Conference on Teaching Reading/Writing Chinese/ English as a Second/Foreign Language. National Central University, Taiwan, 1 May 2005.

Maley, A. (1989) Down from the pedestal: Literature as resource. In R. Carter, R. Walker and C. Brumfit (eds) *Literature and the Learner: Methodological Approaches* (pp. 10–23). ELT Documents 130. Modern English Publications in association with The British Council.

McCarthy, M. and Carter, R. (1994) *Language as Discourse: Perspectives for Language Teaching*. London: Longman.

McNamara, T.F. (2000) *Language Testing.* Oxford: Oxford University Press.

McRae, J. (1991) *Literature with a Small 'l'.* London: Macmillan.

McRae, J. and Clark, U. (2004) Stylistics and teaching literature in a foreign language. In A. Davies and C. Elder (eds) *The Handbook of Applied Linguistics* (pp. 328–346). Oxford: Blackwell.

Miall, D.S. and Kuiken, D. (1998) The form of reading: Empirical studies of literariness. *Poetics* 25, 327–341.

Montgomery, M., Durant, A., Fabb, N., Furniss, T. and Mills, S. (2000) *Ways of Reading: Advanced Reading Skills for Students of English Literature* (2nd edn). London: Routledge.

Osgood, C.E. (1952) The nature and measurement of meaning. *Psychological Bulletin* 49, 197–237.

Purves, A.C. (1973) *Literature Education in Ten Countries: An Empirical Study.* New York: A Halsted Press Book.

Purves, A.C. (1990) *Testing Literature: The Current State of Affairs.* Albany, NY: National Research Center on Literature Teaching and Learning. (ERIC Document Reproduction Service No. ED321261.)

Purves, A.C. (1991) *The Idea of Difficulty in Literature*. Albany, NY: State University of New York Press.

Purves, A.C. (1992) Testing literature. In J.A. Langer (ed.) *Literature Instruction: A Focus on Student Response* (pp. 19–34). Urbana, IL: National Council of Teachers of Education.

Purves, A.C. (1993) Toward a reevaluation of reader response and school literature. *Language Arts* 70, 348–361.

Schmidt, R. (2001) Attention. In P. Robinson (ed.) *Cognition and Second Language Instruction* (pp. 3–32). Cambridge: Cambridge University Press.

Scholes, R. (1985) *Textual Power: Literary Theory and the Teaching of English.* New Haven, CT: Yale University Press.

Schulz, R.A. (1981) Literature and readability: Bridging the gap in foreign language teaching. *Modern Language Journal* 65, 43–53.

Short, M. (1996) *Exploring the Language of Poetry, Plays and Prose.* Harlow: Longman.

Simpson, P. (1997) *Language through Literature.* London: Routledge.

Simpson, P. (2004) *Stylistics: A Resource Book for Students.* London: Routledge.
Snider, J.G. and Osgood, C.E. (1969) *Semantic Differential Technique: A Sourcebook.* Chicago, IL: Aldine.
Spiro, J. (1991) Assessing literature: Four papers. In C. Brumfit (ed.) *Assessment in Literature Teaching* (pp. 16–83). London: Macmillan Modern English Publications.
Strelka, J. (1969) *Problems of Literary Evaluation*. University Park, PA: the Pennsylvania State University Press.
Thornborrow, J. and Wareing, S. (1998) *Patterns in Language: An Introduction to Language and Literary Style.* London: Routledge.
Turner, C.E. and Upshur, J.A. (1995) Constructing rating scales for second language tests. *ELT Journal* 49 (1), 3–12.
Turner, C.E. and Upshur, J.A. (2002) Rating scales derived from student samples: Effects of the scale maker and the student sample on scale content and student score. *TESOL Quarterly* 36 (1), 49–70.
Van Peer, W. (1990) The measurement of metre: Its cognitive and affective functions. *Poetics* 19, 259–275.
Weigle, S.C. (2002) *Assessing Writing.* Cambridge: Cambridge University Press.
Wigglesworth, G. (2008) Task and performance based assessment. In E. Shohamy and N.H. Hornberger (eds) *Encyclopaedia of Language and Education* (2nd edn) (pp. 111–122). Volume 7: *Language Testing and Assessment*. Berlin, London, Philadelphia: Springer Science + Business Media LLC.

Appendix

Language Awareness Test

Read the following text and answer the questions

More Pulling Power–The New Terrano II

Power, so they say, is an aphrodisiac. And you can't get much more powerful than the new Terrano II. Beneath that raunchy new exterior lies a 2.7 litre turbo diesel intercooler engine that delivers extra horsepower at low revs, just when you need it. Switch to 4 wheel drive and you'll be down that rocky gorge (a doddle with the limited slip differential), through that stream (up to 450 mm deep) and up that mountain (up to 39 degrees) with something close to wanton abandon.

But perhaps you should begin with something more gentle. Like tarmac. For it's here that you'll appreciate the Terrano's more sensitive qualities. Its responsive handling. Its smooth as a satin sheet ride. (Thanks to anti-roll bars). The ease with which it will tow a trailer, pull a caravan, and more or less anything else you fancy. Add ABS and driver's airbag and you have a 4 × 4 that's too desirable for words.

Which is exactly why it comes with engine immobiliser and an ultrasonic and perimetric alarm.

<div align="right">YOU CAN WITH A NISSAN</div>

I. How do feel about this text after you read it?
Circle one of the seven-point scales. Do this as intuitive as you can.

Non-literary	Literary
Incomprehensible	Comprehensible
Boring	Exciting
Worthless	Valuable
Ordinary	Unique
Simple	Complex
Prosaic	Poetic
Unattractive	Sexy
Banal	Original
Monotonous	Varied
Weak	Strong
Bad	Good

1   2   3   4   5   6   7

II. Detection and identification
1. Which of the following style features did you find in the above text?

– Tick (√) the box if you think it is **present** in the text
– Cross (×) the box if you think it is **absent** in the text
  ☐ Rhythm and metre
  ☐ Sound patterning (i.e. alliteration, assonance and rhyme)
  ☐ Binary oppositions
  ☐ Repetition

□   Metaphor (including simile)
□   Antithesis
□   Voice

2. Based on what you have just ticked, indicate <u>where</u> in the text the features occur. (Underline the text; provide those that you have ticked only)

>   *For example:*
>   <u>The winter street is a salt cave</u>. The snow has stopped falling and it's very cold.
>   *metaphor (please write out).*
>   The cold is spectacular, penetrating. The street has been silenced, a theatre of whiteness...

III. What, in your opinion, might be the stylistic function(s) of the linguistic features you have found in the text? (First, list those linguistic features you have ticked in Q1, and then comment on its function or effect in the right column; you may provide more than one possible function or effect a linguistic feature brings about; your answer can be in Chinese.) Remember, you need to relate all your observations about how the language is used to its context.

Linguistic features	How does this feature contribute its meaning, or significance to the text?

Part 4

Language and Content

Chapter 11

Assessing Language and Content: A Functional Perspective

BERNARD MOHAN, CONSTANT LEUNG and TAMMY SLATER

This chapter will discuss the integrated assessment of language and content (IALC), with particular reference to second language learning and use. We will address the central question of IALC: what does it mean to assess language and content in an integrated way? To put the question more specifically: what does it mean to assess how wording constructs meaning (and particularly content meaning) in text in context?

Three recent trends at all levels of education in different world locations have made this chapter's discussion of functional second language assessment particularly relevant and necessary. First, for reasons of equality of access and entitlement, linguistic minority students in Australia, Europe, North America and many other places are placed in the mainstream curriculum where they study a range of subjects and learn the language of education (e.g. English) at the same time (Leung, 2007). Second, the content and language integrated learning (CLIL) approach has been gaining popularity in second/foreign language education in many places, including Europe. Third, increasing numbers of students in India, Pakistan, China and other countries study science and other subjects through English, which is not their first language (L1). We believe the functional approach we discuss here provides the best fit for language assessment among curriculum demands, student tasks and pedagogic uses.

In these situations, second language learners are expected to learn subject content and the language associated with it at the same time. Accordingly, in an increasing number of education systems, an integrated approach to language and content *instruction* for second language learners is mandated policy. However, in a striking inconsistency, policy for integrated language and content *assessment* is essentially absent. For example, NCATE TESOL Standards (2005) promotes integrated instruction, but its guidelines for assessment do not discuss how to assess

language and content in an integrated way. They do not change the standard practice of assessing language and content *separately.* Yet, when a learner writes an essay in social studies or science for example, language and content are integrated. The wording of the essay constructs the content of the essay. A teacher reads and assesses the content using the evidence of the wording. Indeed, the same is true in the language class, though there one might talk of the meaning of the essay rather than the content. The question of assessing how wording constructs meaning in text is fundamentally important, not just for IALC, but also for a great deal of assessment in general; however, it has attracted very little research attention. Why? We will argue that the question requires a view of text as making meaning with language resources rather than the traditional view of text as a display of linguistic forms.

What are our criteria for examining the quality of assessment? We are talking about assessment in a broad sense, which includes situated classroom assessment processes; we are not talking about 'tests', so we are not assessing with reference to the CEFR or the International English Language Testing System (IELTS) or any other foreign language performance criteria. We aim to examine the quality of the validity of IALC. We follow the view that validity is appropriately conceived as a validity argument (Chapelle *et al.*, 2008). Part of the validity argument is a domain definition, which is based on a process of researching 'the nature of knowledge in [the relevant] arena, how people acquire it and how they use it' (Mislevy *et al.*, 2003: 18). Central to our domain definition is the concept of meaning in text, particularly 'field' and ideational meaning, through which learners build knowledge ('content') and we rely upon systemic functional linguistics (SFL) for a theory and analysis of how wording constructs meaning in text. In addition, this theory provides insights into learner development and the demands of academic discourse, which are key to the judgment of individual language performance. Moreover, the theory provides tools to analyze situated processes of formative classroom assessment and teacher judgment, which would otherwise go unrecognized.

How do our criteria relate to previous work in language testing and assessment? An example of a validity argument approach is provided by Chapelle *et al.* (2008), who discuss how to build a case for validity using an 'interpretive argument approach' for the Test of English as a Foreign Language (TOEFL). Some of the inferential reasoning adopted in that approach is relevant to this discussion, particularly in terms of the way in which we construe student meaning making in context (leading to evaluating) and curriculum knowledge (language and meaning) as

understood in terms of field and ideational meaning. The concepts of field and ideational meaning are discussed in great depth in Halliday and Matthiessen (1999) and our particular focus on causal explanation is detailed in Slater and Mohan (Chapter 13, this volume). Situated processes of formative classroom assessment are discussed in Leung and Mohan (2004) and situated cases of teacher assessment are examined by Low (Chapter 12, this volume).

In what follows, we will discuss recent thinking in second language assessment, summarize the state of research on IALC assessment and illustrate classroom dilemmas of IALC. Then we will describe a systematic approach to IALC, identifying relevant theory and providing detailed examples.

Second Language Assessment Research

Up to the late 1970s, second language assessment regarded language ability as a body of discrete knowledge (e.g. vocabulary and grammar) and skills (e.g. reading and writing), the measurement of which was context-independent. From this viewpoint, student writing was seen as a display of grammatical forms and lexical items. Meaning and content were not valued. This traditional 'language as rule' approach thus eliminates IALC.

Three recent developments in second language assessment bear on IALC (see Alderson & Banerjee, 2001, 2002; Bachman, 2000). First, concerned that large-scale standardized formal testing may penalize linguistic minority students, proponents have argued for classroom-based teacher assessment because, *inter alia*, it allows use of teacher knowledge and insight (e.g. Huerta-Macías, 1995). From the IALC viewpoint, classroom-based formative assessment is important, and inextricably tied to learning content, but the actual assessment criteria operated by teachers are not necessarily based on a content-language integrated view.

Second, researchers have promoted models of communicative competence assessment, which aimed to broaden assessment goals beyond the production and comprehension of grammatically correct sentences and the language code to knowledge of how to use the code appropriately in social contexts (Bachman, 1990; Bachman & Palmer, 1996; Canale & Swain, 1980). However, Widdowson (2001) points to a known fatal defect in all of these models: they do not say how the competences relate to each other in actual communication. To remedy this defect, he recommends that knowing a language be conceived in terms of

Halliday's concept of meaning potential rather than the idea of competences. From the IALC viewpoint, then, these models do not say how meaning in text is constructed and they lack the concept of meaning potential. We will discuss these issues further below.

Third is testing languages for specific purposes, where researchers have developed assessments and tests that relate to specific fields or domains of knowledge and skills (e.g. testing oral proficiency for non-native teachers of English or health professionals). Reviewing extensive research in the area, Douglas (2005: 860) recommends that assessors view specific purpose language ability as including 'both specific purpose language knowledge and field specific content knowledge'. From the IALC viewpoint, Douglas' recommendation for relating language and content is a very significant development, and one that underlines the importance of providing a linguistic analysis of how language in context constructs meaning or content.

There is little research on IALC assessment because appropriate theory, analysis and practice are not widely known. In a recent comprehensive and penetrating review of research on IALC, Byrnes (2008: 46–47) notes that 'the assessment of content requires a language-based theory of knowing and learning that addresses characteristics of literate language use in all modalities', but a major difficulty 'lies in the fact that the L2 community cannot as yet readily draw on a theory of language that places meaning and content in the center of its interests'. Consequently 'to date only sporadic work exists that explicitly targets the implications of that reorientation for assessment'. This has made it problematic to describe the link between language form and content. That said, we suggest that the integrated content-language arguments adopted in this section offer a communicatively more adequate view of language in use. At the same time, our treatment of IALC is nomologically consistent, albeit from a different epistemological position, with current debates on the centrality of 'content' in the conceptualization of validity (Lissitz & Samuelson, 2007).

While standard second language assessment research hardly addresses IALC, teachers struggle with IALC problems. Our first case study (Low, this volume) presents a detailed picture of teachers facing IALC dilemmas between language and content and having difficulty relating wording to meaning. Low studied teachers as they mark their students' writing about topics of the content curriculum, articulate the decisions they make and struggle with the dilemmas they feel as they reflect on their students' work. They mark wording and meaning separately ('five marks for language and five marks for content'), and

there is little relation between their judgments of wording and their judgments of meaning: they do not address systematically how the student uses wording to construct the meaning of the text. However, the teachers are also deeply uncomfortable and wish to give credit to their students' achievements in discussing complex matters of meaning. The voices of these teachers thus provide invaluable insights into IALC issues in the classroom.

A Functional Approach to the Integrated Assessment of Language and Content

We will now explore a functional approach to language form and content in IALC. SFL provides a language-based theory of knowing and learning. It sees language as a resource for making meaning. It aims to describe 'meaning potential', the linguistic options or choices that are available to construct meanings in particular contexts. It studies the whole text as a unit of meaning, not decontextualized sentences. SFL provides tools to investigate and critique how wording constructs meaning in text and context: register theory relates social context to text through three meaning components of the language system, ideational, interpersonal and textual, which are described in detail in a semantic grammar (see Halliday & Matthiessen, 2004).

SFL sees language as a means for learning about the world. It models learning as a process of making meaning, and language learning as building one's meaning potential to make meaning in particular contexts. Knowledge is viewed as meaning, a resource for understanding and acting on the world. All knowledge is constituted in semiotic systems with language as the most central (Halliday & Matthiessen, 1999: 1–3).

Halliday considers language as the primary evidence for assessing what a person has learned. If language is the primary evidence for learning, then assessment is primarily assessment of text or discourse, and of how wording constructs meaning in text. Thus, Halliday's theory of learning opens the way towards a linguistic theory of the assessment of learning. Drawing on this SFL perspective, we model assessment as a language process.

The SFL framework offers two complementary entry points to IALC: the assessment of genre and the assessment of register. The assessment of genre can draw upon SFL work on the types of genres that are prominent in education and their typical progression through the curriculum inspired by genre theory (Christie & Martin, 1997). For example, Veel (1997) provided an analysis of the main genres in secondary science

textbooks. In addition, he showed that texts work to construct certain kinds of meaning and argued that these texts construct an 'idealized knowledge path' (Veel, 1997:189) that apprentices students into the social practices of science. According to Veel (1997: 167), this knowledge path progresses from the genres related to 'doing science' (procedure, procedural recount) to 'organizing scientific information' (descriptive and taxonomic reports) and 'explaining science' (sequential, causal, theoretical, factorial, consequential explanation and exploration), to 'challenging science' (exposition and discussion, which try to persuade a reader by presenting arguments for or against an issue). This progression shifts from the grammar of speaking to the grammar of writing, and an increasing use of grammatical metaphor. In part, it moves from specific sequences of events in specific places at specific times, to general sequences of events in a timeless setting to cause-effect sequences involving abstract phenomena. Coffin (1997: 196) mapped a similar pathway that apprentices students into the written text types or genres of school history. The pathway moves from narrative genres to argument genres. There is a move from the past as story (with particular concrete events) through the genres of explanation to constructing 'history as argument' (Coffin, 1997: 198). The pathway moves towards abstraction: from mainly human participants to participants that are generic, from specific to general, and from concrete to abstract. It moves from temporal links to causal links and the resources of appraisal for evaluation.

Taking such L1 work on subject-specific literacies (see Unsworth, 2000) into collegiate foreign language education, Byrnes *et al.* (2006) discuss a project in the German department at Georgetown University, which designed an integrated genre-based and task-oriented curriculum, identified the genres that instantiated the content areas it addressed and developed elaborated statements about their language features. Byrnes (2002) reports on three assessment criteria for writing development: breadth of obligatory and optional genre moves, depth of content information provided in each of these moves, and the quality of language use at the discourse, sentence and lexicogrammatical level in line with genre expectations.

The shift in the knowledge path from the grammar of speaking to the grammar of writing is expanded by the concept of a 'mode continuum' from language as action to language as reflection, and from casual conversation to planned written monologue (Martin, 1992). Taking up this concept, Gibbons (1998, 2002), in a series of classroom studies, has researched how elementary teachers scaffold second language learners'

oral statements about subject matter into more literate and less context-dependent discourse, in a process that can be seen as an example of classroom formative assessment.

A Register Approach to Assessment

We now turn to a register approach to IALC, which is the approach we will focus on in this chapter. A field of educational knowledge such as science education is a semiotic system. The register of that field is a system of meanings that realizes or encodes the field in language. The register is a resource for creating meanings, a 'meaning potential', which can interpret and produce the texts of the field in context (see Halliday, 1999).

A register approach enables us to target directly the vital meaning-wording relation, and to trace the role of content by means of language analysis of ideational meaning. Halliday (1985: 101) has long asserted that ideational meaning in everyday terms is 'meaning in the sense of content'. *Ideational meaning therefore offers essential tools to analyse the integration of language and content*. Ideational meaning constructs our knowledge of the world from our experience, and so is vital to education. The register of a knowledge discipline, for example, includes complex systems of ideational meaning. A register is associated with an 'ideation base' (Halliday & Matthiessen, 1999), which includes ideational semantic resources for construing our experience of the world relevant to the register. Underlying Veel's 'idealised knowledge path' is a claim about the development of ideational meaning, particularly causal meaning (see Slater and Mohan, this volume).

Ideational meaning provides language resources to make sense of three main realms of experience: the identification and classification of things, qualities or processes, the representation of events and activity sequences, and human consciousness, including mental and verbal processes. Mohan (1986; Mohan & Lee, 2006) argues that a human activity or social practice has a coherent 'frame of meaning', which includes all three main realms of ideational meaning in a theory-practice dynamic and summarizes this claim in a 'knowledge framework' heuristic.

We will focus on causal relations because they link with all three of these realms and illustrate the semantic process of reasoning. As Painter (1999: 245) says, 'the ability to infer cause-effect relations is fundamental to notions of ''logical'' and ''scientific'' thinking, and the fostering of the abilities to reason and hypothesize are prominent educational goals throughout the Western world'. She notes that SFL analysis of cause

includes reason, purpose and condition and distinguishes between the 'external' sense of cause as in 'I love him because he gives me flowers', and the 'internal' sense of cause as in 'He loves me because he gives me flowers', meaning 'Because he gives me flowers, I know he loves me'. This internal sense of 'reasons for belief', of proof, evidence or reason for a knowledge claim occurs throughout academic disciplines and classrooms.

The nature of assessment provides a further motive for tracing cause-effect relations: asking for student reasoning should be an intrinsic part of a 'meaning assessment strategy'. To guard against rote memorization, a wise assessor checks that learners actually understand the 'wordings' they are saying and the meanings they appear to be constructing, and therefore needs to ask for relevant semantic inferences and reasoning to provide inductive evidence of these meanings (see Mohan, 1972).

In the remainder of this chapter, we will examine how register theory provides a basis to assess the relation between meaning and wording in text and context. We will show how this has major implications for standards of validity in assessment and for IALC. Using the example of causal discourse, we will show how a register approach applies to the assessment of written discourse, and then how it applies to formative assessment interactions between teachers and learners.

Functional Assessment of Text: A Standard of Validity

In this section, we will discuss the functional assessment of meaning and wording in text and context. We will begin with the assessment of written text. Later, we will concentrate on the strategically important case of formative assessment in classrooms and examine spoken interaction. We will discuss functional assessment with particular reference to (second) language learning and use.

Macken and Slade (1993) provide a general perspective on functional assessment, stating that assessment should be a linguistically principled procedure; it should be explicit about the language resources learners need to perform tasks in different disciplines; and it should provide specific criteria that recognize the difference between different tasks.

As we noted earlier, both content teachers and language teachers assess the meaning of texts in context on the basis of their wording. This common ground underlies and is presupposed by the different evaluations they make. It is therefore a fundamental responsibility of IALC assessment research to provide a linguistically principled account of this common ground.

There are many reasons why it should do this. We would expect an evaluator who was responsible and not arbitrary to be able to explain or justify his or her judgment of the meaning of the discourse by pointing to wording in the discourse that expresses that meaning. Responsible assessment is judgment based on evidence. Furthermore, an evaluator who formatively assesses 'for learning' and aims to help the writer learn to write better also needs to explain how the meaning of the discourse is created by its wording and to be able to suggest alternative ways to convey meaning by wording. As Macken and Slade (1993) suggest, assessment should communicate explicitly about the language resources needed to perform tasks. All these considerations argue towards a central standard of validity for an assessment of text: that an assessment should assess how wording constructs the meaning of the text as a whole in its context on a linguistically principled basis. This standard applies to the productive work of writing or speaking a text, as well as to the receptive work of reading or listening to a text. The standard is a main basis of responsible assessment.

In what follows, we will argue that it is both feasible and essential to assess meaning and wording in discourse, showing how a functional approach to language can provide the theory and analysis needed to relate meaning and wording systematically, and how this approach can inform language assessment practices. We will use two instances of causal explanations, discourse that is found across academic subject areas, to illustrate our argument.

Relating Meaning and Wording in Causal Explanations

Discourse assessment of causal explanations is important since they are a central part of academic discourse in general. We have chosen to use the water cycle in our argument because it is a widely known topic for causal explanations. We will show how the difference in meaning between two explanations of the water cycle is realized by a difference in wording.

The following two explanations, from Mohan and Slater (2004), were elicited using a diagram of the water cycle. Explanation A was written by a secondary school teacher whose L1 is English, and Explanation B was written by a university student who speaks English as a second language (ESL).

Explanation A:
The water cycle.
What are the processes that 'water' goes through?

(1)   Initially, the water cycle begins as snow melts from the glaciers.
(2)   The water then meanders through various water sheds until it reaches rivers and lakes. Water eventually reaches the oceans.
(3)   Water, then, becomes water vapour (it evaporates into the air) and accumulates in what we call clouds.
(4)   The 'clouds' then distribute water in the form of rain, snow, or sleet back to the mountains where the cycle begins again.

Explanation B:

The water cycle: The sun is the source of our water. The water, or hydrological, cycle begins when the sun heats up the ocean to produce water vapour through evaporation. This water vapour mixes with dust in the atmosphere and forms clouds. Cool air causes condensation of water droplets in the clouds, bringing about precipitation, or rain. This rain then falls into rivers, streams and lakes and eventually returns to the ocean, where the cycle begins again.

These two texts differ in interesting ways in terms of discourse meaning and wording. In terms of discourse meaning, each explanation constructs a line of meaning that runs through the discourse (Longacre, 1996). Explanation A constructs a time line of events in time sequence (Event A is followed in time by Event B). Explanation B constructs not just a time line, but also a line of actions and events in causal sequence, in a cause-effect relation (A causes B).

In terms of wording of the text, Explanation A constructs its time line using time conjunctions (*initially, then, eventually*), dependent clauses of time (*as* snow melts, *until* it reaches), lexical verbs of time (*begin*) and a series of event verbs (*melts, meanders, reaches, becomes*). There is only one explicitly causal feature (clouds *distribute* water). Explanation B constructs its causal line using causal dependent clauses (*to produce water vapor*), cause/means as circumstance (*through* evaporation), lexical verbs of cause (*produces, causes, brings about*), nominalization of causal processes (*evaporation, condensation, precipitation*), action verbs (the sun *heats up* the ocean) and a causal metaphor (the sun is *the source* of our water).

A competent assessor of these two texts should be able to recognize the difference in lines of meaning between the two explanations and how this difference is realized by a difference in wording. If the aim of the assessment is to see which of these explanations is a causal explanation, the evidence clearly points to Explanation B. Thus, the assessor can justify the claim that Explanation B is the better causal explanation by pointing to the evidence of the wording, and can explain to learners the

language aspects of causal explanation by showing the difference in wording between the two explanations. The assessor's claim is based on the way Explanation B has used the resources of the language to create meaning in discourse, in this case a causal explanation. The claim is not based on whether Explanation B is more factually correct than A or whether B violates fewer grammar rules or discourse conventions than A.

In SFL, the more precise compound term 'lexicogrammar' is used to refer to what we have been describing as 'wording'. This term signals that the meaning of the wording is realized both in lexis and grammar and has to be traced through both. In Explanation A, a time line is constructed using both lexical verbs of time (*begin*) and time conjunctions (*initially*). In Explanation B, a causal sequence is constructed using both lexical verbs of cause (*produces*) and cause/means as circumstance (*through evaporation*).

The resources to express causal meanings are an aspect of language development both in the culture and the individual speaker. Halliday and Martin (1993) discuss the historical development of Scientific English and find that causal discourse has taken the following developmental path:

from A happens; so X happens
        because A happens, X happens
                that A happens causes X to happen
                        happening A causes happening X
    to   happening A is the cause of happening X
                        (Halliday & Martin, 1993: 66)

Our second case study (Slater and Mohan, this volume) explores this developmental path in individuals by examining how students who are native speakers of English and ESL students develop their resources to express causal meanings. Extending Veel's 'idealised knowledge path' as a frame for the development of causal meaning and wording, Slater and Mohan apply it to the oral data of interviews about science learning in school with English language learners (ELLs) and native speakers of English from two different grade levels. They show the developmental trajectories of the learner in the construction of causal discourse and the associated use of lexicogrammatical resources. Through a combination of description and analysis, they vividly demonstrate how grade 9 English L1 speakers can draw on causal language resources when needed much more readily than ELLs.

Theory of Language and the Relation Between Meaning and Wording

How does SFL support the standard of validity for an assessment of discourse that addresses the meaning ('content') of the text and relates it to the wording of the text? SFL recognizes the importance of text or discourse as a construction of meaning rather than as a display of features of the language system. This meaning is technically termed 'discourse semantics' (see Table 11.1). In our example above, we analyzed temporal and causal lines of meaning as the discourse semantics. (Another possibility would be to analyze the more complex discourse semantics of a register or of a genre of discourse.) SFL recognizes that the discourse semantics of a text are realized by the lexicogrammar of the text. SFL analyzes grammar as 'semantic grammar', as form related to meaning, a very different analysis than traditional grammar, and organizes grammatical meaning under three 'metafunctions': Ideational (construing experiences), Interpersonal (enacting social relationships) and Textual (creating discourse). These three metafunctions co-occur in all texts. Our analysis of causal meaning here foregrounds the Ideational metafunction.

Our example of the two explanations showed that it was not difficult to explore informally the relation between discourse semantics and lexicogrammar in two contrasting texts. However, pursuing the relation systematically requires knowledge and application of the relevant literature (see Halliday & Matthiessen, 2004).

Relating discourse and wording, as is done in SFL, requires certain assumptions about language that are very different from many traditional beliefs about language. Broadly, there is a contrast between an SFL view of *language as resource* and a traditional view of *language as rule*. Our analysis of explanations A and B depends on a view of language as resource. A view of language as rule would not be capable of producing

Table 11.1 The relation between meaning and wording in text

	Language functions in register		
	Ideational	*Interpersonal*	*Textual*
Discourse Semantics (meaning of a discourse)	Construing experiences	Enacting social relationships	Creating discourse
Lexicogrammar (wording of a discourse)	e.g. Transitivity (verbal processes)	e.g. Mood	e.g. Theme

the same analysis, as shown in Mohan and Slater (2004) and discussed later in this chapter. If we contrast traditional beliefs about language with those of SFL as in Table 11.2 (see Derewianka, 2001), one can see that the traditional view eliminates the meaning-wording relation. We suggest that the general failure of second language assessment to deal with the meaning-wording relation can be traced to such traditional beliefs about language.

Traditional grammar sees language as a set of rules for the form and structure of language, and language form is not related to meaning in context. It sees written or spoken text as a display of sentence grammar forms, as evidence of competence in the language rules. Language learning is acquiring rules that result in correct form, and the role of evaluation is to judge this correctness of form. Assessments of meaning are judged independently of form and lack a basis in theory or specific evidence. In this view, the meaning of the text as a whole is not of interest, and nor is the question of how that meaning is realized in the wording.

From an SFL perspective, how text makes meaning through its wording is a central question for language assessment. SFL sees language form in relation to meaning, and sees language as a resource for making meaning. It does not privilege the language system over the text, but

Table 11.2 Assumptions of SFL and traditional grammar (after Derewianka, 2001)

Systemic functional linguistics	*Traditional grammar*
Language as a resource for making meaning	Language as a set of rules
Language form related to meaning	Form unrelated to meaning
Text makes meaning using language resources in context	Written or spoken text as a display of sentence grammar forms
Relates language system to both text and values	Values language rules (competence) rather than text (performance)
Language learning as extending resources for making meaning in context	Language learning as acquiring correct forms
Evaluate text as making meaning with resources in context	Evaluate correctness of form; judge meaning independently from form

values them both. It does not consider the text as a display of language resources, but sees it as making meaning using the resources of the language system in context. Language learning is seen as extending resources for making meaning in context. Evaluation can judge how the learner has made meaning in a text and how the learner has used the resources of the language system. For example, our analysis noted how Explanation B constructed a causal line of meaning and used many more lexicogrammatical resources for causal meaning than Explanation A did. Notice how the analytic emphasis is on what the learners *can* do, and not simply on what they *cannot* do.

Much of second language assessment research appears to assume a traditional language as rule perspective, with its emphasis on error, and fails to deal with meaning-wording relations. Mohan and Slater (2004) explored this issue in two ways. The first was to examine models for assessing communicative competence (Bachman, 1990; Canale & Swain, 1980). These appear to be simple extensions of the assumptions of traditional grammar with a strong emphasis on competence. Canale and Swain took grammatical competence (the knowledge a speaker has about the rules of grammar) as their base model and added sociolinguistic and strategic competences, independently of meaning making in context (Leung, 2005); Bachman added textual and illocutionary competences, all of which appear to be conceived of as a matter of generalized rules or conventions. The learner's discourse becomes a display of correct or incorrect forms of these rules. Assessing grammatical competence means assessing the language learner's discourse for grammatical errors, and assessing for the other competences appears also to be a matter of checking for relevant errors. There is no evidence of a conception of language as a resource for meaning, of a text as a construction of meaning or of the role of lexicogrammar. There is, therefore, no evidence that these models can recognize meaning in a text as a whole or deal with meaning-wording (i.e. discourse semantics-lexicogrammar) relations.

The second way that Mohan and Slater (2004) explored the issue of meaning-wording relations in causal discourse was by using Explanation A and Explanation B as test cases to see if their differences could be recognized by second language assessment instruments. The first instrument was a locally developed test for assessing the communicative competence of potential second language teachers based on Canale and Swain (1980), and the second was the scoring guide for the Test of Written English (Educational Testing Service, 1990). In both cases, the raters looked for errors across a range of categories and assessed the two explanations as equal, judging them on the basis of error. In both cases,

the raters intuitively judged Explanation B as more advanced, but could not recognize this in their assessment because there was no matter of error. In other words, when presented with explanations A and B, assessors working with these communicative competence models assessed the texts in terms of perceived errors only. While they recognized intuitively that B was a better explanation than A, they felt that their models did not allow them to express that recognition in any articulated way. These assessors thus reinforced the notion that these models do not recognize discourse as a construction of meaning realized in wording; instead, the models extend grammatical competence to a taxonomy of textual, illocutionary, sociolinguistic and strategic competencies, and judge discourse as a display of correctness or error in these competencies.

This is consistent with the traditional view of language, which separates meaning and wording, and consistent with some specialists in second language assessment who believe that meaning and wording (content and language) should be assessed separately; judgments of meaning are thus separated from the evidence of wording that could justify them.

As we have argued thus far using the water cycle explanations, a functional approach to the assessment of discourse should judge the meaning ('content') of a text and justify or explain this judgment by relating it to the wording of the text. This should be a central standard for validity. We have noted, however, that much work in second language assessment operates under assumptions that make judging the meaning of a text in relation to the wording difficult if not impossible to do. Therefore, it is essential to draw on a functional approach like SFL to provide the meaning-wording relation with theory and language analysis that will support this standard of validity.

Functional Assessment in Classroom Interaction: Functional Recasts

The Assessment Reform Group helpfully describe classroom formative assessment:

> tasks and questions prompt learners to demonstrate their knowledge, understanding and skills. What learners say and do is then observed and interpreted, and judgements are made about how learning can be improved. These assessment processes are an essential part of everyday classroom practice. (Assessment Reform Group, 2002: 2)

One type of formative assessment is provided by formal recasts of learner errors of grammatical form:

(1)   NS: When does your father work?
(2)   ELL: My father work at night.
(3)   NS: [RECAST] Your father works at night?
(4)   ELL: Yes, he works at night.

The ELL's grammatical error in (2) is correctively recast by the native speaker (NS) in (3), who thereby assesses (2) as grammatically incorrect and provides detailed feedback to the ELL, who corrects the error in (4), showing evidence of learning. Thus, the participants have interacted to construct a brief formative cycle of utterance, feedback and uptake.

Formal recasts (as in 1–4) are associated with a traditional formal view of language that sees assessment as judgment of the correctness of utterances and learning as movement from error to correct form. *Functional* recasts are associated with a functional view of language that sees assessment as judgment of the functional appropriateness of the expression of meaning, and learning as expanding the learner's resources for making meaning.

An example of the formative assessment of *functional integration of form and meaning* is provided in a *functional* recast by a teacher of ELLs' causal explanations during a project on the human brain in a content-based language learning classroom at university level (see Mohan & Beckett, 2003):

(5) S: We can relax our brain by wave.
(6) T: We can relax our brain by wave? How does that work? [RECAST] How does a wave help us relax our brains?
(7) S: Because... the cerebral wave of the stable type appears when the mind relaxes, and it improve the centering power.

This example shows T using a recast as part of a larger strategy for scaffolding causal explanation by formatively probing for an explanation. S offers a causal explanation in (5). In (6), T assesses (5) as needing causal elaboration, and uses the recast to pose a guiding question, making 'wave' the agent of the explanation, not the means, as it is in 'by wave', and offering 'help' as a causal process. S's uptake in (7) offers a much more elaborated causal explanation, making 'wave' the agent, using 'improve' as a causal process, and adding a causal nominalization 'the centering power'. As a causal explanation, as scientific discourse, and as academic discourse generally, (7) is more elaborate and 'developed' than (5).

As with the formal recast above, the participants in the functional recast situation have interacted to construct a brief formative cycle of

utterance, feedback and uptake. The difference here concerns the focus of the assessment, whether on form independent of meaning as in the formal recast, or on adjusting the form to elaborate on the meaning and thus help expand the student's language resources in context.

Further evidence of causal functional recasting includes Mohan and Luo (2005), who studied online computer-mediated communication in a graduate language education course, where ESL students skillfully functionally recast their peers' discourse as part of the normal practice of online academic discussion. Early (2001) contains examples of formative interactions in elementary social studies and elementary literature classes where teachers can be seen to causally recast student statements. Slater and Mohan (this volume) conveys the pervasiveness of causal discourse in science and ways to make functional formative assessment sustained and systematic.

Given the dominant view of second language assessment, and the IALC dilemmas that teachers experience (Low, this volume), it is remarkable to discover that teachers functionally assess their ELL students' utterances in classroom formative assessment, and therefore show an intuitive understanding of functional assessment. We do not suggest that this is done consciously and systematically. However, we do suggest that intuitive functional formative assessment may be a very widespread phenomenon, and that it offers teachers a major opportunity to reflect on their intuitive practices and build on them systematically.

Functional Assessment of Register in a Unit of Teaching

We now move to a broader level: functional analysis of the assessment phase of a unit of classroom teaching, to illustrate assessment of a very simple register and its meaning potential. The field of knowledge is beginning level magnetism. The teacher is one who knows the field and has already constructed the meaning potential of magnetism. The learners, however, have to build up this meaning potential, or frame of meaning, learning the discourse of magnetism. What register meanings are the learners expected to develop? How can one assess that they understand these meanings, and have not simply memorized register wordings?

To explore these questions, we will discuss a study of a Western Canadian grade one/two ESL science class on magnetism (see Mohan & Slater, 2005). In the teaching and learning phase of the unit, the children learned a simple 'theory' of magnetism in experiments with bar magnets, whose poles were marked. Then, in a formative assessment phase, the teacher aimed to assess the children's understanding of magnetism by

having them extend their 'frame of meaning' to the new case of ring magnets, whose poles were not marked and which looked very different. They were to find out if the ring magnets had north and south poles. Thus, knowledge of the theory was developed in the practical situation of bar magnets and was formatively assessed in the practical situation of ring magnets.

A general functional question is: how is the register 'frame of meaning' realized in the three main realms of ideational meaning? The core of the theory was: a bar magnet has two poles – north and south. North and south attract. North and north repel. South and south repel. In terms of ideational meaning, the theory constructs a taxonomy of 'poles' (north and south) and of two causal relations (attract and repel). The children investigated the theory through simple experiments where they pondered answers to experimental questions and evaluated experimental evidence for those answers.

Thus, the children's frame of meaning should include examples of the three main realms of ideational meaning mentioned earlier: the identification and classification of things, qualities or processes (taxonomy of north and south poles), the representation of events and activity sequences (the causal relations attract and repel), and human consciousness (the children investigated, tried things out, discovering or coming to know answers).

To indicate when the teacher was assessing examples of these three realms of ideation, we have highlighted processes (verbs) in the formative assessment discourse below. The first realm of ideation relates to the processes of being and having, which have been bolded, the second realm relates to processes of doing and happening, which have been italicized, and the third realm relates to processes of human consciousness, which have been underlined.

A second general functional question is: what meaning assessment strategies are used to gather inductive evidence that learners actually understand register meanings, and have not simply memorized register wordings? We argued earlier that, to guard against rote memorization, the wise assessor asks for student reasoning, including causal reasoning. To highlight causal meanings in the formative interaction below, we have capitalized some explicit causal elements.

At the broadest level, the teacher's register assessment strategy was to pose the experimental question and to scaffold the students to gather and evaluate the experimental evidence to answer it. This included getting the students to infer descriptions of the case based on their prior

knowledge, asking for semantic inferences about the case and asking for reasons for these inferences.

First, having posed the experimental question of whether the ring magnets had north and south poles, she demonstrated repelling and attracting to guide the learners to describe the unfamiliar case:

> Teacher: So... what *happened* here?
> Students: It *repelled.*
> Teacher: They're *repelling*. Right. They were *repelling* and I'm going to *turn* this one over. What do we <u>call</u> this? North or south?
> Students: North.
> Teacher: North. It doesn't <u>matter</u>. I'm *turning* it over. What...
> Student: *Attract.*

Having got the students to label one pole hypothetically, the teacher next asked for a semantic inference about the 'attract' situation and then for their reasons for it.

> Teacher: SO IF it's *attracting* what **is** underneath here? North or south?
> Students: South.
> Teacher: South. Right. The bottom **is** probably north and this part **is** south. ...WHY? BECAUSE?
> Student: BECAUSE north and south.
> Teacher: BECAUSE north and south and what do north and south always *do*? What **is** the rule?
> Students: *Attracts.*
> Teacher: That's right. North and south always *attract*. What *repels*?
> Student: North and north or south and south.

By mentioning 'repel', the teacher encouraged the learners to make the corresponding inference about the initial 'repel' situation by themselves and work out the answer to the experimental question. Next, the teacher asked the learners to infer the answer and then asked for their reasons:

> Teacher: Okay. SO <u>tell</u> me about these magnets? Do they **have** a north and south?
> Students: Yeah...
> Teacher: How do we <u>know</u>?
> Jack: BECAUSE we <u>tried it out</u>.
> Teacher: And? What did we <u>discover</u>?...
> Jack: BECAUSE IF you *turn* it around it won't *attract* and IF you *turn* it around it'll *attract*.
> Teacher: SO it **has** a north and south? Yes it does.

The teacher is not assessing language form in isolation, she is assessing language meaning and wording. She is not treating the assessment of language separately from the assessment of science. Rather, she assesses the magnetism register, the frame of meaning, the ideational meaning potential that is central to both.

The teacher assesses examples of the three main realms of ideational meaning: the taxonomy of poles ('do they **have** a north and south?'), the causal relations of attract and repel ('what's *happening* here?'), and human consciousness ('how do we know?'). These different kinds of ideational meaning are a first step in meaning analysis, and they suggest how the register frame of meaning constructs a coherent domain of human consciousness of the things and events of magnetism.

The teacher's 'meaning assessment strategies' are asking for inferred descriptions of new practical contexts, semantic inferences and justification of semantic reasoning. The strategies appear appropriate to provide evidence that learners can understand and use the meaning potential of this simple register. Meaning assessment strategies are likely to be an important aspect of future research on formative assessment.

Many of the inferences and justifications depend on causal meaning. For example, Jack's statement illustrates both the 'external' and the 'internal' meaning of cause. The external sense is shown by 'IF you *turn* it around it won't *attract*'. The internal sense ('causes me to know') is shown by 'How do we know... [We know] BECAUSE...'. Requests to provide reasons for inferences are natural contexts for use of the internal sense of cause.

This example of assessment of a simple register has general implications. All academic disciplines and subject areas are registers. All registers are complex frames of meaning. To understand and appreciate formative assessment interactions in these disciplines and subject areas, we need to trace their frames of meaning and their meaning assessment strategies.

Conclusion

IALC is the linguistically principled assessment of how wording constructs the meaning of a text in its context, which is the common ground presupposed by language assessment and content assessment of text.

IALC is disabled by a language as rule view of language. Standard second language testing and assessment does not provide an IALC assessment of text, and does not draw on a theory of language and meaning/content that would support it. Lacking this, teachers find it

difficult to systematically assess, diagnose and help learners in their construction of meaning (Low, this volume).

IALC is enabled by a language as resource view of language. SFL provides a language-based theory of knowing and learning, and a theory and analysis of how wording constructs meaning, and therefore a foundation for validity in integrated assessment. On this basis, we have shown how texts can be systematically functionally assessed for IALC on the evidence of their meaning-wording relations, using the example of ideational meaning and causal discourse, an important area of functional discourse development (Slater and Mohan, this volume).

On the same basis, we have described cases of IALC formative assessment at the micro-level and the macro-level of classroom interaction where teachers intuitively functionally assess how wording constructs meaning. Meaning-wording analysis illuminates functional assessment processes such as causal recasting and aspects of meaning assessment strategies that would otherwise go largely unrecognized. These are also clear cases where we are dealing not just with the assessment of discourse, but with the discourse processes of assessment – analysis of assessment as discourse. These cases are steps towards a linguistic theory of assessment.

We believe that intuitive functional formative assessment may be a very widespread phenomenon. Formative classroom assessment is a strategic area where teachers can take the initiative when larger scale assessment has been found wanting. We believe that functional formative classroom assessment could become a major force to address learners' needs more adequately. We strongly recommend that where teachers are working formatively on IALC, they be given adequate recognition and research support for that work, and adequate resources to pursue it.

A great deal of assessment of all learners, not just second language learners, evaluates the meaning of written or spoken texts on the evidence of their wording. A functional analysis of meaning and wording such as SFL offers a needed basis for validity, as we have shown. We strongly recommend that it be more widely recognized that systematic and principled IALC is not only possible, but necessary, and on a very wide scale.

References

Alderson, C. and Banerjee, J. (2001) Language testing and assessment (Part I). *Language Teaching* 34, 213–236.

Alderson, C. and Banerjee, J. (2002) Language testing and assessment (Part II). *Language Teaching* 35, 79–113.

ref

Huerta-Macías, A. (1995) Alternative assessment: Responses to commonly asked questions. *TESOL Journal* 5, 8–11.

Leung, C. (2005) Classroom teacher assessment of second language development: Construct as practice. In E. Hinkel (ed.) *Handbook of Research in Second Language Learning and Teaching* (pp. 869–888). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Leung, C. (2007) Dynamic assessment – assessment as teaching? *Language Assessment Quarterly* 4 (3), 257–278.

Leung, C. and Mohan, B. (2004) Teacher formative assessment and talk in classroom contexts: Assessment as discourse and assessment of discourse. *Language Testing* 20 (3), 335–359.

Longacre, R.E. (1996) *The Grammar of Discourse*. New York and London: Plenum Press.

Macken, M. and Slade, D. (1993) Assessment: A foundation for effective learning in the school context. In B. Cope and M. Kalantzis (eds) *The Powers of Literacy: A Genre Approach to Teaching Writing* (pp. 203–230). Pittsburgh, PA: University of Pittsburgh Press.

Martin, J.R. (1992) *English Text: System and Structure*. Amsterdam: Benjamins.

Mislevy, R.J., Steinberg, L.S. and Almond, R.G. (2003) On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives* 1, 3–62.

Mohan, B. (1972) Comprehension as semantics plus induction. *Linguistics* 115, 93-105.

Mohan, B. (1986) *Language and Content.* Reading, MA: Addison-Wesley.

Mohan, B. and Beckett, G.H. (2003) A functional approach to research on content-based language learning: Recasts in causal explanations. *Modern Language Journal* 87 (iii), 421–432.

Mohan, B. and Lee, G. (2006) Linking interpretive research and functional linguistics. In G.H. Beckett and P.C. Miller (eds) *Project-based Second and Foreign Language Education: Past, Present, and Future* (pp. 263–280). Greenwich, CT: Information Age Publishing, Inc.

Mohan, B. and Luo, L. (2005) A systemic functional linguistics perspective on CALL. In J. Egbert and G. Petrie (eds) *Research Perspectives in CALL* (pp. 87–96). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mohan, B. and Slater, T. (2004) The evaluation of causal discourse and language as a resource for meaning. In J.A. Foley (ed.) *Language, Education, and Discourse: Functional Approaches* (pp. 255–269). New York: Continuum.

Mohan, B. and Slater, T. (2005) A functional perspective on the critical 'theory/practice' relation in teaching language and science. *Linguistics and Education* 16, 151–172.

NCATE TESOL Standards (2005) *Program Report for the Preparation of Teachers of English to Speakers of Other Languages – Program Report Template.* TESOL, National Council for Accreditation of Teacher Education.

Painter, C. (1999) *Learning Through Language in Early Childhood.* London: Continuum.

Unsworth, L. (2000) Investigating subject-specific literacies in school learning. In L. Unsworth (ed.) *Researching Language in Schools and Communities: Functional Linguistic Perspectives* (pp. 245–274). London and New York: Cassell.

Veel, R. (1997) Learning how to mean—scientifically speaking: Apprenticeship into scientific discourse in the secondary school. In F. Christie and J.R. Martin (eds) *Genre and Institutions: Social Processes in the Workplace and School* (pp. 161–195). London: Continuum.

Widdowson, H.G. (2001) Communicative language testing: The art of the possible. In C. Elder, A. Brown, N. Iwashita, E. Grove, K. Hill, T. Lumley, T. McNamara and K. O'Loughlin (eds) *Experimenting With Uncertainty: Essays in Honour of Alan Davies* (pp. 12–21). Cambridge: Cambridge University Press.

Chapter 12

Teachers and Texts: Judging What English Language Learners Know From What They Say

MARYLIN LOW

- *Look at this response. I think I know what the student is trying to say. But, I don't know what is language specifically and what is content—I can't tell whether the student's language is interfering with his ideas or his ideas are interfering with his language. How would you mark this?*
- *Well, it all depends . . .*

The study of current classroom assessment practices through teacher talk offers insights that can guide important changes to future assessment practices. The question raised by the above exchange – how to judge what English language learners (ELLs) know from what they say – was part of a conversation between teachers that had been circling the text for an hour, always returning to this same point. This is thus an example of how, in discussions about the complex and contrasting understandings of relations that can and do exist between content and language, there continues to be confusion around the assessment of ELL texts. For a long time, we have seen educational institutions structured such that teaching and assessing language is presumed to be a separate event from the teaching and assessing of content, even though language is a main medium of learning across the curriculum. In some second language (L2) contexts, a shift to integrating language and content through content-based instruction has occurred (Grabe & Stoller, 1997; Mohan, 1986). Yet, questions of judging language and content texts of ELLs remain uncertain for many teachers (Low, 1999; Norton & Starfield, 1997), triggering the comment by one teacher, '*Well, it all depends . . .*'.

Recently, work in functional linguistics, especially in the analysis of classroom discourse and its associated assessment (Leung & Mohan, 2004; Mohan & Slater, 2004), has raised important questions regarding differing perspectives on the relationship between language and content

241

and the dilemmas they pose for teachers and others engaged in assessment practices. In particular, this work draws attention to the impact on ELLs and assessment of their academic learning, a focus that is taken up here.

This chapter, framed by Mohan *et al*. (this volume), delves into the discourses of teachers as they engage in the familiar practice of assessing L2 learning. In my earlier qualitative study of teacher decision making in assessment, a study that was situated at a college in Canada with Japanese international students studying in a four-year academic program (Low, 1999), I sat beside 15 teachers for periods of 1–2 hours over several months, sometimes individually and sometimes in groups, and observed and asked questions about language and content as they evaluated ELL texts. Recently, I have extended this enquiry by following its themes in a very different context – teachers in upper elementary grades in a school in American Samoa, where English is a second language for both teachers and students. Even though the cases are diverse, there are important commonalities when it comes to assessing L2 texts. This is remarkable given situational differences between elementary and college classrooms, Samoan and Japanese learners, and native English-speaking teachers and teachers who themselves are ELLs. Finding similar teacher commentaries in two vastly different cases both affirms and strengthens the original thesis – that teachers appreciate the importance of how grammar constructs meaning in L2 academic writing, but their decisions are dominated by a system that assesses for traditional grammatical correctness (or more accurately, grammatical errors of form – see Bachman and Palmer's (1996) detailed and influential assessment framework that aimed to address language *use* rather than *form*, but continued to draw attention to language without concern for the content knowledge it construes). This creates an incongruent and tensioned gap in their practices. It also frustrates opportunities to guide learners to develop the language resources needed to construct their academic knowledge.

Reform efforts in assessment internationally document shifts from independent, standardized product formats to more complex, problem-solving processes embedded in everyday instructional practices (Assessment Reform Group, 2002; Broadfoot & Black, 2004; Leung & Lewkowicz, 2006). Influenced by the assessment literature, much of the research about how teachers assess L2 texts has concerned itself with the 'how' of assessing writing separately from the content knowledge it constitutes, perpetuating a belief that gets lived in a system handed down to teachers in the form of expectations to assess ELLs' language knowledge

separately from content knowledge (Low, 1999; McNamara, 1996; Rea-Dickens, 2007). Yet, there is a strong call for the field of educational assessment, especially in L2 environments, to draw on a conceptual framework for language-based knowing within forms of literacy. Byrnes offers systemic functional linguistics (SFL) as an approach

> that acknowledges that language performance is not individually owned but fundamentally socially construed and passed on through discourse communities. (Byrnes, 2008: 45)

Byrnes shares how Georgetown University used SFL in a foreign language setting as a conceptual framework for exploring content knowledge as socially constructed in language in their Developing Multiple Literacies project (http://www3.georgetown.edu/departments/german/programs/curriculum/index.html). Within a higher education environment, Shohamy (2006) offers a description of how this relationship might be conceptualized in an integrated fashion. Both call for more research about conceptual understandings of a content and language relationship in assessment within L2 environments. This chapter targets that gap and makes a contribution by unpacking, through discourse analysis of teacher talk, the ideological dilemmas of the content and language relationship teachers struggle with when assessing L2 texts.

Discourse Analysis of Social Practice/Activity

In this study, discourse evidence was used to explore assumptions teachers make prior to making a decision on evaluating an ELL text. The analysis is framed by a functional approach (see Halliday, 1994) and those who use social practice or activity as a unit of study and argue that discourses contain evidence of one's resources (theory) and acts (practice) (Harré, 1993; Low, 1999; Mohan & Lee, 2006). Discourse was analyzed using a three-level model of social practice and discourse as follows:

> Generic reflection: '*I know language and content are integrated*'.
> Specific reflection: '*I can't tell whether the student's language is interfering with his ideas or his ideas are interfering with his language*'.
> Action discourse: '*I'm giving this paper 7 out of 10*'.

Harré (1993) explains how the meanings of social acts are made more explicit through such differing types of discourses. He suggests that discourse *acts* are forms of discourse that imply the norms of actions (practices) of an activity, whereas discourse *accounts* of theory and

Table 12.1 Main types of discourse data

Types of discourse data	Activity
Accounts of theory ------------------------------- Commentary on acts	Reflecting ('in' or 'on' action)
(Discourse) acts	Action

discourse *commentaries* in action make evident norms of social acts (see Table 12.1).

This approach claims that evidence from all three types of discourse data is necessary to interpret the meaning of activity more explicitly; that is, commentaries and accounts are viewed as primary evidence of the implied cultural meaning of an act.

On Decision-making: Undecidabilities and Contrary Themes

> Appreciating that tenuous and delicate situation of judgment which is addressed by the name ''undecidability''... does not detract from the urgency of decision; it simply underlines the difficulty. (Caputo, 1993: 3)

This chapter makes use of two concepts related to decision making: Caputo's notion of 'undecidabilities' and what Billig *et al*. (1988) call 'contrary themes or dilemmatic ideologies'. They argue that, in general, decisions are not pre-determined, but instead are negotiated in practice. Contrary themes, dominant meanings, as well as counter themes of negation give rise to deliberation. In deliberation, the social character of language and, in this case, literacy assessment reflected in the teachers' discourses, becomes more explicit. Billig *et al*. (1988: 3) contend that individuals are 'not to be seen as being fully preprogrammed by neatly systematised plans of action, which are awaiting the appropriate triggering stimulus and which obviate the need for all deliberation'. Instead, deliberation of contrary themes is a necessary pre-condition to most decision-making acts. Contrary themes are not automatically reduced to 'either/or', but include multiple and complex historical and current relations. Figure 12.1 conceptualizes the decision-making process, where undecidabilities and contrary themes provide access to the self-determined negotiations teachers engage in before deciding on a mark.
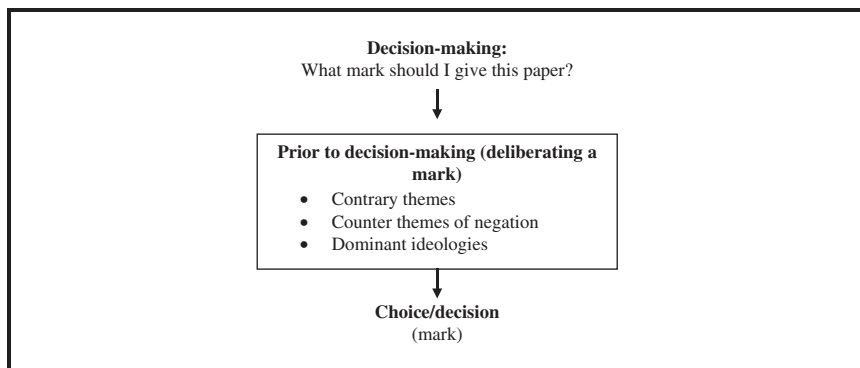
Figure 12.1 The uncertainty of making decisions (judgments)

What do teachers' individual and collective deliberations tell us just prior to marking a text? My previous research suggests that underlying and competing ideologies of how 'language and content' is understood (e.g. traditional grammar versus functional) are enacted every day, resulting in inconsistencies and uncertain practices within the decision-making process (Low, 1999; see also Leung & Mohan, 2004). Analysis reveals the dilemmas teachers face. Yet, within the contrary themes emerging from the discourse, an awareness and appreciation of the ways students make meaning is evident, especially in teachers' discourse prior to decision making.

There are a number of key decision points for teachers when assessing ELL content texts. The following four cases (teachers A, B, C and D) focus on the discourse of teachers prior to deciding on a mark. The ELL texts were written in response to themes from the social sciences. In each case, the teacher has communicated to the student that evaluation will focus on language and content. The teachers do this in different ways through the criteria they set for marking. For example, one teacher commented,

> *I know language and content are integrated but I separate them for marking purposes. See on this one, I give each part a mark – 5 for language and 5 for content – then I add them up. It's easier that way for me and for the students to understand.*

Such a marking scheme separates language from content and sends a message to the students about what is expected of them. Another teacher

shared what he believed to be his responsibility to the ELLs in his classroom:

> *I can't not mark their grammar, even when I want to just focus on meaning, they get mad and tell me I'm not a good teacher.*

In response to this issue, an ELL replied,

> *When he gives me a low mark, I want to know what I did wrong. How to improve my language. My grammar mistakes. I expect him to tell me how to correct not just circle and underline.*

Such expectations are a part of teaching and assessing ELLs and impact how criteria are communicated and then understood and acted upon by students. If language and content are separated, then students' written responses may reflect that understanding – they may pay more attention to grammatical accuracy than to how language communicates meaning. If the criteria address the integration of form and meaning, then that may signal to students to give more attention to how they use language to make meaning. Even with criteria, students are involved in their own decision-making as they craft their essay, some based on what they believe the teachers expect and some based on what they believe to be important in essay writing. For teachers, their assumptions about language and content become part of the deliberation and negotiation process and contribute to inconsistencies when assessing ELL writing.

Instances of Language and Content as/in Contested Spaces

Native English-speaking teachers and second language texts

In the following discourse samples, I sat with each teacher as they rated a text written by an ELL they teach. The first two samples are from my 1999 study – teachers teaching at an international college in Canada, where Japanese nationals study in English in a four-year college program. The curriculum aimed at integrating language and content instruction; there were no language classes – all teachers were language and content teachers and expected to teach and assess the language demands of the content curriculum. An on-going systematic professional development program supported the implementation of a language and content curriculum from a functional linguistic perspective.

Teacher A assigned his students to write an essay about the causes of international debt, a topic they had been studying in class. The prompt

asked, what are the causes of international debt? The sample that follows is the response to the prompt.

Sample L2 text

International debt is crisis many people. Most those live south of equator. North banks borrows money to south countries. They must repay money they borrow. Many don't. Why? World Bank give them borrow to much plus interest they charge. It too much. Also, increases in oil prices makes goverment need borrow more. These countries borrows more than they repay. They have unpayible debt always. As a result, the south country debt grow bigger each year. Resulting in the north-south devide gets bigger to. Sometime debts is forgave but than country borrow more money and debts starts again.

For well-articulated reasons, the teacher did not want to give attention to correcting grammatical form. His focus was on meaning making, evident in the criteria he set.

Teacher A's assessment criteria

- Use key content vocabulary (e.g. international debt, external debt, International Monetary Fund, OPEC, crisis, North/South divide, World Bank, debt, unpayable debt, forgiven debt, interest).
- Provide at least three causes of international debt.

He has prepared his students for the language of this task mainly through vocabulary building.

Prior to decision-making

A: *You see in this essay the student used important words like 'international debt', 'World Bank', 'crisis', 'north and south divide'. This gives substance to the essay. These are words loaded with meaning – they express meaning.*

R (this author): *So, for you, this would be an example of a good paper?*

A: *Well, yes but... no. Look at the grammar errors. You really have to be blind not to see them. Mostly articles and subject-verb agreement. Oh, they'll learn it as they write more. It takes a long time for them. I didn't specify those grammar points for marks anyway so I'll just ignore them. I think about language in a general, overall way. In other words, I ask myself how hard was it for me to understand what they are saying. How many times did I get distracted by their language? What would I give this essay? I don't know. It's not so easy. I guess I should have been more specific for them and for me in what I was looking for. I just wanted the ideas, I thought the language wouldn't matter.*

R: *What is it about the language in this essay that does matter to you?*

A: *Well, they are using the vocabulary words I was looking for – I've ticked seven of them in the introductory paragraph and that's really good – but they don't go together so well and that's distracting. I'm not sure what he knows. Some of the meaning is lost to errors. This is a tough one. My criteria don't work in the way I thought they would.*

Teacher A explicitly states that words are valued when they are 'loaded with meaning', suggesting words gain import when they 'express content'. Yet, he does not value the specific structuring of those words because 'they don't go together so well'. The teacher's account of the valued criteria heavily weights words 'loaded with meaning'. However, he finds the results frustrating, even though the L2 text has met the use of key vocabulary to some degree. Yet, he places little value on those *other* words that link content words and help structure meaning. He indicates a discomfort in the way the 'words loaded with meaning' are structured that is 'distracting', suggesting an intuitive concern for how words and grammar together make meaning. Why is this a 'tough one' for the teacher? It may be that he needs a more systematic way of communicating about content and language as an integration of form and meaning.

The question arises of what was communicated to the student about evaluation criteria, what assumptions of expectations were made by the student (and the teacher) about the prompt, and what decisions the student made in writing based on what was communicated and what was assumed. For example, the student seems to understand the need to use 'key content vocabulary' and the teacher acknowledges this by 'ticking seven of them'. The prompt and the second criterion refer to the construction of causal relations. However, while the student does attempt to make causal meaning about international debt, the teacher does not acknowledge this and instead seems confused and distracted by the grammatical errors. A discourse analysis identifies at least three lines of causal meaning in the L2 response (the cause is underlined in each example):

(1)   They must repay money they borrow. Many don't. Why? <u>World Bank give them borrow to much plus interest they charge.</u>
(2)   <u>…increases in oil prices</u> makes goverment need borrow more.
(3)   <u>They have unpayible debt always.</u> As a result, the south country debt grow bigger each year. Resulting in the north-south devide gets bigger to.

It seems the student understood the prompt and responded as best he could with the language resources he had, even including the causal

words 'as a result' and 'resulting in' to signal two effects of the cause in example three. The teacher did not acknowledge the signal words as a language resource or how they were used in the construction of meaning.

Teacher A's assessment practices have been made more explicit with the inclusion of a sample L2 text and the assessment criteria used. This makes the dilemmas the teacher struggled with obvious. It also shows how it is possible to look at the wording of a text for lines of meaning – in this case, causes of international debt, something Teacher A wasn't able to do. His concern for the errors dominated his evaluation of the text even though formal grammar was not part of the assessment criteria. The case studies that follow do not include an L2 text or criteria, but instead focus on the teacher discourse to substantiate the dilemmas at play for teachers assessing L2 texts.

The following discourse sample comes from another teacher, assessing an L2 text her student wrote. The assessment's main purpose is formative. The teacher is looking for evidence of the student's strengths and needs of language and content in order to tailor future lessons. Her aim is to assess how the student demonstrates her understanding of the causes of climate change. She sets certain expectations through criteria where students are to describe the causal sequences of climate change, combining specific linguistic structures and vocabulary they learned in class.

Prior to decision-making

B: *This one I'm still trying to decide. You see she didn't answer the question in the way I was expecting. She wrote an interesting paper though. Some parts made me think. She didn't just repeat what was learned in class. She has some unique, maybe even original ways with her own words that made me rethink the topic a bit. But it wasn't the answer I was looking for.*

R: *Was it a wrong answer or a different answer?*

B: *Different, I think. I have to decide whether to give this an 'A' or not. If I use the criteria that I marked the others with, she might not even pass because of the grammar. One criterion is that they explain the causes of climatic change. I see only one causal statement and it uses 'because'. She discusses more causes but not the way I taught them – they were to use words like 'because', 'as a result', 'causes', or 'effect' to signal a cause is being expressed. We practiced using these words in class. I wanted to see them. It seems she's found another way to get across her ideas. I want to give her an 'A' but no causal vocabulary and there are so many grammatical errors. If I give her something lower though, I know I'll feel guilty . . . I sure have a lot of work to do with this student. We need to spend a lot more time on grammar.*

While Teacher B appreciates that the student has 'unique maybe even original ways with her own words', she is concerned that the student did not use the key causal vocabulary studied in class and that there is a lack of grammatical accuracy. The teacher holds strongly to the notion that what has been taught must be evaluated and what criteria are established must be applied consistently. The teacher presumed a certain text response. The learner responded differently, using the language resources she had to make meaning to discuss the causes of climate change. Her 'unique way' included listing climatic events one next to the other that implied a causal chain. No causal words were used, yet the relationship between the events was clear. The teacher recognized this but seemed unwilling to value such a structuring of causal meaning. Instead, the necessity of grammatical accuracy and adherence to criteria dominated the decision. While the teacher initially deals with how language form and meaning relate in causal discourse, it is grammatical accuracy that takes precedence, both for marking purposes and for future lessons, exposing a strong belief that grammar as form carries more weight than a semantic grammar of meaning. It raises an important question of what Teacher B communicated to this student regarding expectations for evaluation of the essay. It seems the student was focused on communicating causal relations while the teacher, prior to deciding a mark, shifted from caring about the meanings the student made, to care more about a grammar of form – creating a gap or inconsistency between what is taught and what is assessed. What triggered this shift? It may be that the teacher did not have a way to give feedback on other ways of structuring causal meanings and turned to a grammar that was more familiar to them, missing an opportunity to increase and improve the student's meaning-making linguistic resources.

Non-native English-speaking teachers and second language texts

Two years after the completion of the first study in a Canadian college, I conducted a smaller second study in American Samoa that followed the same themes of the first. The next two discourse samples are of teachers teaching at an elementary school in American Samoa. One teacher teaches grade seven and the other grade eight. At this school, English is the medium of instruction, but 99% of the children entering kindergarten speak Samoan as their first language. They study in Samoan for 45 minutes a day throughout their school experience (K-12). While the emphasis is on English, teachers themselves are English learners, and

often code-switching is the main method of teaching in all classes. Writing in English is a challenging task, as is literacy in Samoan, as Samoan has a strong oral tradition. The teachers work hard to prepare their students for academic success in school and often draw on oral traditions (such as involving the community and studying local issues) to accomplish that goal.

Teacher C prompted her students to write an essay on a certain period of Samoan history, explaining a sequence of events that led up to the current relationship American Samoa has with the US government. She reviewed two of her students' papers, in part to give them a mark, but also to determine how she could help them improve their content writing. As she compared the two papers, she deliberated on meaning making and the grammar of each paper.

Prior to decision-making

C: *The problem is with essays like these two. They were asked to write about the sequence of events in Samoan history at the turn of the century that lead to our relationship with the United States. One has written some good ideas I think. Yet, the language structure is awkward – it sounds Samoan. The other has good structure.*

R: *Is there an order of historical events in each paper?*

C: *Well, in this one the main events are listed one after the other, although two are in the wrong order. She uses 'first' to introduce the events but then uses dates to show the order – she got the dates mixed up on those two. So she is able to show what she knows but the grammar is bad. The other paper has language that is much more controlled and readable but he described an event and it is not very long. He doesn't say much and there is no order of events, he described two together, mixing up the details. Yet, I think I would have to give them both the same mark because the grammar is not good in the first one and his grammar is good. But that doesn't feel right. She has written the history in this paper. He didn't do that. I just can't give her a higher mark when the grammar is like this.*

In her deliberations, the teacher recognized in the first paper the writer's ability to construct a sequence of events amid 'awkward language structures'. In particular, she identifies how the writer used grammar and lexis to construct the sequence. She also recognized, in the second paper, the writer's grammatical ability at the sentence level amid confusion about the sequence of historical events. The teacher intuitively valued the first paper and what the writer was doing – she cared about how the student used language to make meaning and show the history she knew. Yet, in her deliberations, the value of grammar as form

dominated her thinking. Traditional grammar held sway over valuing the meaning-making accomplishments of the first writer. While the teacher is aware of the meaning the second writer is constructing, in the end she recognizes it as faulty because there is no sequence of events. She values that the 'language is more controlled' in the second paper, a commentary that suggests she sees grammar as form, not as meaning making. The potential for valuing how meaning is made in grammar is subjugated by the assumed importance of form. Though she wants to value the meaning constructed in the first paper, she is faced with a dilemma because of the faulty form. She is unable to give credit to the first text for evidence she identifies that demonstrates the student's ability to use grammar and lexis to communicate history.

Teacher D gave his students a writing assignment based on a discussion they were having in their Samoan Studies class about a local issue. While the school and community were located near the top of a mountain and received more rainfall than other villages closer to sea level, their system of rainwater collection (their main source of water) and its dispersion to community members was not effective. The teacher saw the interest of the students and capitalized on this opportunity to create a project where teams of students would research, write and present their solutions to the community. As part of the project, each student had to write an essay describing at least two solutions and the possible outcomes to the community, both positive and negative, for each solution. They were to use the graphic organizers the teacher shared with them in class. The teacher set the criteria on meaning and communicated this to the class.

Prior to decision-making

D: *Although this student still has problems with grammar, the ideas are there. He is working through the choices the community has about their need for a better water supply system. We rely on rainwater here but we need a better system for holding the water and getting it to the people. I gave the students a graphic organizer to use and part of the paper was to explore the community choices and follow that through with possible outcomes – what would happen if the community made this choice or that one. I wanted them to think ahead. They will be presenting their ideas to parents and other elders at a community-school meeting next week.*

R: *What is important to you in this paper?*

D: *I don't like seeing all these errors and I can hear Samoan in the way he uses the English words. We need to work on that some more. But he used the graphic, I can see that in the way he writes, first he states the question the*

community is debating and he writes about one choice and what would happen if the community decided on that. Then he writes about another choice they have and says what is good and bad about it. I like that – oh look, he says this 'will' happen and we don't know that for sure. He should have said 'might' or 'may' happen. We talked about that in class. Hmm, this is difficult. I just wish his grammar errors weren't so bad then I could give him an 'A'.

It seems that the student who wrote the paper understood the criteria and aimed his efforts at using his language resources to discuss a community problem. As the teacher works through the paper, he recognizes the discourse patterns of choices and outcomes. The teacher follows the student's options, valuing how the student structures language to make meaning. He criticizes a statement of outcome as certain ('will happen') rather than possible and identifies needed changes in the grammatical area of modality ('might', 'may') that he had previously discussed with the class. However, the act of decision making begins and ends with comments about grammatical form. While the teacher has helped his students to construct meaning grammatically (integration of language and content), he is hampered by the need for the paper to be correct in form (a separation of language from content). A conscious effort to focus on grammatical meaning making is usurped by an unquestioned priority for grammatical accuracy.

As stated earlier in this chapter, the two contexts of the case studies vary in a number of ways. However, even within two very different contexts, they consistently illustrate contested spaces of undecidabilities and contrary themes where teachers judge what ELLs know from what they say.

Conclusion

Studying current classroom assessment practices, especially the dilemmas involved in marking L2 texts, makes an important contribution to new assessment developments by locating them in the context of ongoing, everyday teacher practices (Byrnes, 2008). While the discourse evidence shows that teachers are still drawn to more traditional, grammar-based approaches to assessing L2 texts, even when it is not their intention, they are also breaking from those traditions and becoming more open to the messiness and struggle of how students explicitly link form and meaning, content and language.

A number of considerations for the field of L2 writing assessment emerge from an analysis of the discourse samples in this chapter. What is clearly evident are the contrary themes or dilemmatic ideologies that

teachers negotiate in the decision-making process of assessing L2 texts. From the opening discourse sample through to the specific cases, teachers describe their struggle with perceived tensions: language form as separate from meaning versus language form as related to meaning. Similarly, the sample discourse from a student suggests a tension between wanting feedback on his language (as meaning) and grammar (as errors). Such contrasting themes seem to be inherent in a system of language assessment where language is mainly talked about in terms of formal grammar and grammar in terms of errors.

These teachers, through their discourse, are highlighting a dilemma that the field of language assessment should heed – teachers want to respond to the way students make meaning with language, yet a perpetuated notion of grammar as form without meaning dominates their thinking. This brings about a discomfort between what teachers value (making meaning with language) and what they believe the system is forcing them to do (focus on grammar as form). While teachers may not fully understand how, they intuitively know that text constructs meaning using language.

In each of the above cases, there is evidence that teachers care about the meanings their students make. They work with causal relations, sequence of events, choices and outcomes. They think about how students are constructing meaning, and value it when their students make good use of the language resources they have to communicate their ideas. If teachers could build on these initiatives to adopt a sustained approach to assessment that systematically relates meaning and language resources, it could become a critical tool for guiding students to construct academic meaning across subject areas and to know how to read and write the world. If we expect teachers to assess what ELLs know (content) based on what they say (language), a focus on the integrated assessment of content and language from a functional perspective holds much promise for teaching and assessment. If we expect ELLs to demonstrate what they know in words and teachers to evaluate it, then we have a responsibility to help teachers reorient their approach to classroom assessment practices to view the learners' use of language as socially and linguistically constructed content knowledge. SFL offers us such an approach that allows teachers to communicate clearly to students integrated assessment of language and content expectations and be willing to stand by them without being persuaded by lingering assumptions that presume to separate language from content.

References

Assessment Reform Group (2002) *Assessment for Learning: 10 Principles.* On WWW at http://www.qca.org.uk/4034.html.

Bachman, L. and Palmer, A. (1996) *Language Testing in Practice: Designing and Developing Useful Language Tests.* Oxford: Oxford University Press.

Billig, M., Condor, S., Edwards, D., Gane, M., Middleton, D. and Radley, A. (1988) *Ideological Dilemmas*. London: Sage Publications.

Byrnes, H. (2008) Assessing content and language. In E. Shohamy (ed.) *Language Testing and Assessment* (pp. 37–52). Vol 7 *Encyclopedia of Language and Education* (2nd rev edn), N. Hornberger (gen ed.). Berlin: Springer-Kluwer Academic.

Broadfoot, P. and Black, P. (2004) Redefining assessment? The first ten years of Assessment in Education. *Assessment in Education* 11 (1), 7–27.

Caputo, J. (1993) *Against Ethics.* Bloomington, IN: Indiana University Press.

Grabe, W. and Stoller, F. (1997) Content-based instruction: Research foundations. In M.A. Snow and D.M. Brinton (eds) *The Content-based Classroom* (pp. 5–21). New York: Longman.

Halliday, M.A.K. (1994) *An Introduction to Functional Grammar* (2nd edn). London: Edward Arnold.

Harré, R. (1993) *Social Being* (2nd edn). Oxford: Blackwell.

Leung, C. and Lewkowicz, J. (2006) Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly* 40 (1), 211–234.

Leung, C. and Mohan, B. (2004) Teacher formative assessment and talk in classroom contexts: Assessment as discourse and assessment. *Language Testing* 21 (3), 335–359.

Low, M. (1999) 'Difficulties' of integrative evaluation practices: Instances of language and content as/in contested space(s). PhD thesis, University of British Columbia.

McNamara, T. (1996) *Measuring Second Language Performance.* New York: Longman.

Mohan, B. (1986) *Language and Content.* Reading, MA: Addison-Wesley.

Mohan, B. and Lee, G. (2006) Linking interpretive research and functional linguistics. In G.H. Beckett and P.C. Miller (eds) *Project-based Second and Foreign Language Education: Past, Present, and Future* (pp. 263–280). Greenwich, CT: Information Age Publishing, Inc.

Mohan, B. and Slater, T. (2004) The evaluation of causal discourse and language as a resource for meaning. In J.A. Foley (ed.) *Language, Education and Discourse: Functional Approaches* (pp. 255–269). New York: Continuum Press.

Norton, B. and Starfield, S. (1997) Covert language assessment in academic writing. *Language Testing* 14 (3), 278–294.

Rea-Dickins, P. (2007) Learning or measuring? Exploring teacher decision-making in planning for classroom-based assessment. In S. Fotos and N. Nassaji (eds) *From Focused Instruction and Teacher Education, Studies in Honour of Rod Ellis* (pp. 193–210). Oxford: Oxford University Press.

Shohamy, E. (2006) Rethinking advanced language proficiency. In H. Byrnes, H.D. Weger-Guntharp and K. Sprang (eds) *Educating for Advanced Foreign Language Capacities: Constructs, Curriculum, Instruction, Assessment* (pp. 188–208). Washington, DC: Georgetown University Press.

Chapter 13

Towards Systematic and Sustained Formative Assessment of Causal Explanations in Oral Interactions

TAMMY SLATER and BERNARD MOHAN

> Student: They aren't stick. They aren't stick.
> Teacher: Why?
> Student: Maybe this made of something with metal. Maybe this made
> with something else.
> Teacher: Why do you think this is attracted to the magnet?
> Student: Because it's both metal. They're both metal. (Slater, 2004)

The questions and answers in the above exchanges are common occurrences in classroom discourse: requests by the teacher for causal explanations and efforts by the students to give them. To succeed in school, students need to be able to explain causally, and teachers need to be able to assess these explanations. Students' causal explanations allow teachers to check understandings of how and why; thus, examining the development of this type of discourse has the potential to provide a framework for formative assessment that can promote learning. Researchers and educators working from a systemic functional linguistic perspective have provided a body of work on causal discourse in science, offering an excellent starting point for examining the development of causal explanations in that subject area. Much of the work that has been undertaken has generally focused on texts written by expert writers (e.g. Mohan *et al.*, 2002; Veel, 1997), such as textbooks and encyclopedias.

Education has historically considered reading and writing skills to be of primary importance, and thus research into written genres has been critical in exploring the language of schooling. An examination of oral discourse development is also important because it is typically through oral interactions in the classroom that the ability to discuss cause and effect is honed. This chapter aims to show that there is a parallel between written and oral forms of causal discourse and presents a model of causal

discourse development based on the findings from studies on written texts. It uses this model to examine the linguistic features occurring in the oral causal discourse of English as a second language (ESL) and non-ESL (native-English) speakers at two grade levels (ages six/seven and fourteen/fifteen). The point of presenting the information in this chapter is to highlight the developmental path of causal language and to suggest that this path can offer a way to support validity arguments in the assessment of these types of explanations and student understanding of causal concepts. The model outlined here is thus a basis for sustained systematic formative assessment that can contribute to our understanding of the development of both oral and written explanation. More generally, this chapter, like Mohan, Leung and Slater (Chapter 11, this volume), contributes to domain definition in the validity argument through domain analysis (Chapelle *et al.*, 2008). However, while Chapter 11 works at the broader level, such as meaning and wording in text, this chapter concentrates specifically on the discourse of causal explanations and its development, researching 'the nature of knowledge in [the relevant] arena, how people acquire it and how they use it' (Mislevy *et al.*, 2003: 18).

What is Meant by Causal Discourse Development?

A concrete example of the type of development this chapter is addressing can be illustrated by using two examples from Gibbons (1998):

> *Text 1:* Our experiment was to find out what a magnet attracted. We discovered that a magnet attracts some kinds of metal. It attracted the iron filings, but not the pin. It also did not attract things that were not metal.

> *Text 2:* A magnet is a piece of metal which is surrounded by an invisible field of force which affects any magnetic material within it. It is able to pick up, or attract, a piece of steel or iron because its magnetic field flows into the magnet, turning it into a temporary magnet. Magnetic attraction occurs only between ferrous materials.

The first explanation, offered as a written text by an English language learner (ELL) after doing experiments with magnets, is very much a recount of what was done and observed. The focal point of this explanation is the generalization 'a magnet attracts some kinds of metals'. The second text, written by a textbook author, contains various linguistic features that characterize it as a much more sophisticated

causal explanation, features such as nominalizations (e.g. attraction) and causal processes (e.g. affects).

A similar developmental progression occurs in oral explanations, as the data from Slater (2004) show. Bob is a seven-year-old native English speaker trying to explain how he knows there is an invisible force acting on magnets.

> Bob: Because um... there even when you can't see it you could somehow you could put it between the magnets and there's a kind of you know it feels kind of real? But another way to prove it is that... you could take another uh thing the magnet will attract to and will be attracted... and then and it would be hard to to like explain... if there wasn't one... like I mean an invisible thing.

Bob used temporal and causal conjunctions to construct his explanation, which became circular and rather confusing in his effort to say what he has understood about magnetism. In contrast, Edward is a 15-year-old student who has demonstrated in his class work a good grasp of science language and content. He offered his recount of the experiment his teacher did and attempted to explain his understanding of a precipitation reaction.

> Edward: When he mixed them... it turned into a yellow substance and he called—he told us that um... when something changes color and produces some sort of powder that's called a precipitation reaction? And it's not gas producing. It's just that... it just produces a solid?

Edward's oral text includes nominalizations (e.g. precipitation) and causal processes (e.g. produce) along with conjunctions, just as the more sophisticated written text did.

The Developmental Path of Cause in Written Texts

In both the written examples from Gibbons (1998) and the oral examples from Slater (2004), a difference with regard to the stages of development can be seen. There are similar issues in language development that can be explored by examining the language features involved.

The proposed model is a schematized developmental path that moves out from the lower left corner, as shown in Figure 13.1. The vertical axis of this model suggests that there is a semantic shift moving from time to external cause to internal cause, or proof. This semantic shift draws upon Veel's idealized knowledge path and his movement from *doing* science to
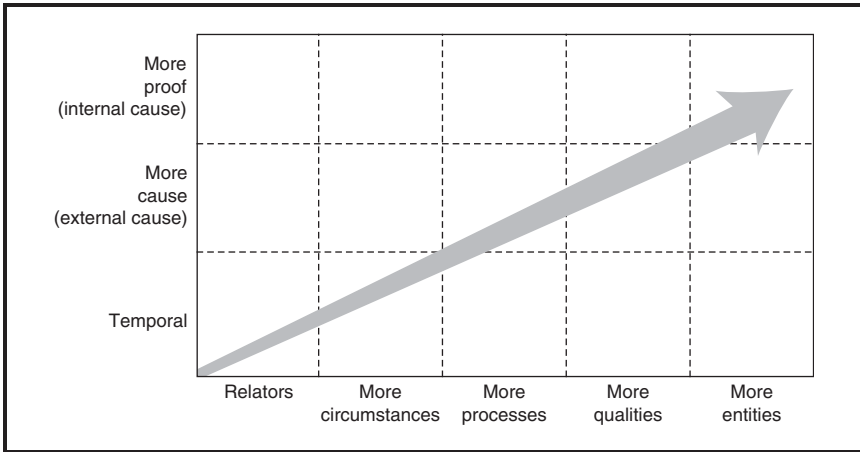
Figure 13.1 The developmental path of cause (Mohan *et al.*, 2002)

challenging science. The horizontal axis of the model suggests that there is also a move away from relators (conjunctions) as the primary marker of causality towards more grammatically metaphoric constructions, such as circumstances, processes, qualities and entities, following Halliday (1998: 211), who described this progression as 'the "general drift" of grammatical metaphor', from the clause complex, through to clause, and finally to nominal group, the most metaphoric construction.

As noted earlier, Veel (1997) argued that explanations for younger students tend to be sequential accounts of observable events, and it is only when the student can deal with more abstract or theoretical concepts that the explanations progress beyond the language of sequence. He proposed that there are four linguistic indicators that mark the development of content and move students from the younger, sequential explanations towards 'the abstract, technical and "transcendental" kinds of meaning we expect of adult, educated discourse' (Veel, 1997: 188). He illustrated the indicators with supporting examples from textbooks, showing that as the curriculum progresses, there are changes in the frequency of key linguistic features. These four indicators were an increase in lexical density, a higher number of nominalizations and abstractions, a shift from temporal to causal conjunctions and a move from external to internal text organization.

Veel provided a clear set of hypotheses surrounding conjunctions, lexical density and nominalizations. His basic hypotheses, based on his analysis of four explanations that progressed from the relatively visible

world of sequential observations (texts 1 and 2) to the more abstract factorial and theoretical explanations (texts 3 and 4), were:

- Lexical density will increase: For example, in text 1, Veel offered data that showed nine lexical items over three clauses (e.g. sugar, cane, comes, farms), suggesting low lexical density. He contrasted this with text 4, which contained twelve lexical items over two clauses (e.g. density, fluid, greater, average), a noticeably greater lexical density.
- The number of nominalizations and abstractions will increase: Using the same two sections of data as above, Veel showed no nominalizations or abstractions in the three clauses of text 1, but four nominalizations over the two clauses of text 4 (e.g. density, weight).
- Temporal conjunctions will decrease: Text 1 contained temporal conjunctions such as 'as' and 'then' in greater numbers than did text 4.
- Consequential conjunctions will increase: Text 4 contained consequential conjunctions such as 'if', 'because' and 'therefore' in greater numbers than did text 1.
- External conjunctions will decrease: Text 1 connected more to the observable world, using conjunctions that reflected visible sequences, such as 'as the sugar cane comes from the farms, it is washed . . .'
- Internal conjunctions will increase: Rather than following a natural sequence of events, the latter examples (texts 3 and 4) made more use of internal conjunctions such as 'firstly', 'secondly' and 'thirdly' to organize the text.

To explore Veel's hypotheses more fully and to elaborate on the role grammatical metaphor may play in the knowledge path, Mohan *et al.* (2002) used a computer concordancing application combined with hand analysis to examine discourse samples of 70,000–75,000 words each from a science encyclopedia for learners aged eight to fourteen and from one targeted for older, university-level students. The features for analysis were taken from lists of causal items provided by previous concordancing studies (e.g. Fang & Kennedy, 1992; Flowerdew, 1998). When Mohan *et al.* tallied their findings from the corpus analysis and held them up against Veel's hypotheses, they found a mixed pattern of results. There was support for the first three hypotheses, but not for the last three.

Mohan *et al.* went on to track the frequencies of various processes, qualities and entities in the two encyclopedias as well. They discovered that whereas the number of external causal processes dipped slightly in the encyclopedia for older students, the number of proof processes

(internal cause) increased. The findings appeared to suggest that causal language develops along two dimensions: a semantic dimension and a lexicogrammatical dimension, as the model in Figure 13.1 captures. Semantically, the numbers suggested that there is a move from time, through cause, to proof. Lexicogrammatically, there is a shift away from the use of conjunctions to more metaphoric ways of constructing meaning. This developmental pattern offers important evidence to support the validity of judgments that rate one performance of causal discourse over another.

The Development of Oral Causal Discourse

To use the information captured in this 'developmental path of cause' in formative assessment, it needs to be seen whether learners' oral explanations follow similar paths to what was found in the texts written by experts. In Slater (2004), native English-speaking students from the primary grades (ages six/seven) and high school (ages fourteen/fifteen) as well as non-native English-speaking students at the same age levels were asked to explain their knowledge of what they had been studying in their science classes. Ten hours of interviews were recorded, transcribed and analyzed with the same concordancing techniques as used in Mohan *et al.*, to see if the results would pattern out in similar ways. In the following paragraphs, trends between the native English speakers in the primary and the high school grades as well as those between ESL and native English speakers will be discussed.

When the interviews from the native English speakers in the primary and high school grades were examined, a similar pattern emerged to that of the Mohan *et al.* data. As Table 13.1 suggests, the developmental move appeared to be both *semantic*, from time to cause to proof, and *grammatical*, from less to more metaphoric. There was a visible shift in the direction of grammatically metaphorical constructions as well as a shift towards causal features as the constructions became more metaphorical. Causal and temporal processes were used more in the older grades, as were participants and metaphoric entities in general. The largest increase in the metaphorical entities occurred with processes and with nominalized qualities, suggesting that the older students have a higher level of ability to manipulate the lexicogrammar than do the younger students. Halliday (1993) stated that this ability to handle grammatical metaphor begins at about grade eight, and although this research cannot verify that claim, it does support the idea that this ability develops at some point between grades two and nine.

Table 13.1 The native English speakers (Slater, 2004)

Linguistic feature	Primary	High school
External temporal conjunctions	25.35	51.11
External causal conjunctions	29.11	12.81
Internal conjunctions	0	.28
Temporal circumstances	15.96	22.56
Causal circumstances	3.76	.56
Temporal processes	0	1.39
Causal processes	1.88	6.41
Proof processes	.94	.7
Temporal entities	0	2.51
Causal entities	.94	4.46
General metaphoric entities	0	16.99

Note: Numbers have been normalized to occurrences in 1000 words

The following discourse examples[1] will attempt to illustrate both the grammatical and semantic differences between the younger students and those in high school. The teacher who was doing the interviews in the primary class asked Bob to offer proof that an invisible force exists, that of magnetism. With regard to the lexicogrammar, Bob mainly used conjunctions to construct the argument, as highlighted in bold. He used one process of proof, in the word 'prove'. The ellipses signify short pauses in his speech, reflecting time needed to think. With regard to semantics, the younger students had difficulty talking about internal cause, specifically with offering proof texts, as this analysis of the example provided above attempts to show:

> Teacher: So how do you know that there is an invisible force? Bob?
> Bob: **Because** um. . . there even **when** you can't see it you could somehow you could put it between the magnets and there's a kind of you know it feels kind of real? But another way to *prove* it is that. . . you could take another uh thing the magnet will attract to and will be attracted. . . **and then** and it would be hard to to like explain. . . **if** there wasn't one. . . like I mean an invisible thing.

Bob's explanation for how to prove the existence of magnetism is unclear. He suggested that it 'feels kind of real' and that it would be hard to explain attraction if there wasn't 'an invisible thing' there. In general, when the native English-speaking primary students were asked about proof, their arguments became quite circular, as the above attempt shows.

By contrast, the high school native English speakers were quite eloquent constructing internal proof, and their ability to use more grammatically metaphorical constructions was much greater, as the following example exemplifies. Sara and Jeanie recounted the experiment they had recently witnessed, using a temporal conjunction in the last sentence of this recount. They used several metaphorical entities (underlined) as well as a process of proof, 'proved' (bold italics), thus making their explanation more grammatically metaphoric than what Bob had offered.

> Sara: **And then** we weighed it again and it was exactly the same to the hundredth of a gram.
> Researcher: And what does that show?
> Sara: That...
> Jeanie: The mass of the <u>reactants</u> is the same as the mass of the <u>products</u>.
> [Which is
> Sara: [That's the law of the <u>conservation</u> [of mass.
> Jeanie: [Of mass.
> Researcher: Mm-hmm?
> Sara: So we ***proved*** it.

These students explained that they were able to prove the scientific law they were studying by carrying out an experiment and applying their findings. In other words, even though both groups of students talked about proof, the older students were able to handle the concept much more logically and convincingly in their explanations, and used more grammatically metaphoric language to do so.

The data from the native English speakers and the ELLs at the primary grades were consistent with the lower end of the model. The students appeared to be dependent on conjunctions for constructing causality, and both temporal and causal conjunctions were used, depending on the question prompts they heard. There were minimal numbers of causal entities or processes, supporting Halliday's suggestion that grammatical metaphor is not a characteristic of young children's discourse. But whereas one might expect that native English speakers would perform

better linguistically than ELLs, no marked differences were found. A possible explanation surfaced from a qualitative study of the ELL classroom. It appeared that the ESL teacher had undertaken a systematic approach to integrating language and content for the students, very carefully building up the meaning and wording of the subject matter and taking considerable care to review orally the material with the students on a regular basis. [For a full description of this research, see Mohan and Slater (2005)].

At the high school level, the data are more revealing, showing aspects of causal discourse development that are of particular importance with regard to formative assessment of ESL students at this level. A path similar to the one Mohan *et al.* suggested appeared when the interviews from the high school ESL speakers and the native English speakers were examined. As Table 13.2 shows, this path was much like the one that was constructed when the interviews from the primary and high school students were explored, with one exception. The high school ESL students used more temporal circumstances than the native English speakers did, but given that one of the tasks in the interview was to explain the changes in the state of water throughout the four seasons, the

Table 13.2 The high school speakers (Slater, 2004)

Linguistic feature	ESL	Non-ESL
External temporal conjunctions	29.68	51.11
External causal conjunctions	30.53	12.81
Internal conjunctions	0	.28
Temporal circumstances	30.31	22.56
Causal circumstances	1.47	.56
Temporal processes	0	1.39
Causal processes	4.21	6.41
Proof processes	0	.7
Temporal entities	0	2.51
Causal entities	0	4.46
General metaphoric entities	11.37	16.99

Note: Numbers have been normalized to occurrences in 1000 words

higher number of temporal circumstances is natural, boosted by phrases such as *in summer* and *in winter*.

As noted earlier, the native English speakers at this level were able to use grammatically metaphoric constructions with relative ease. Edward, for example, was able to move from the more congruent form of a recount using temporal conjunctions (bold) into a generalized statement that included the nominalization of a scientific process (underlined), then move back down through qualities (italics) to the more congruent form, using a causal process (bold italics):

> Edward: **When** he mixed them... it turned into a yellow substance and he called—he told us that um... **when** something changes color and produces some sort of powder that's called a <u>precipitation reaction</u>? And it's not *gas producing*. It's just that... it just ***produces*** a solid?

Sara showed her ability with grammatical metaphor by using it in a very colloquial manner. Her nominalization of a process captured her confusion about the adjustments needed to make a clean flame in a Bunsen burner:

> Sara: And I totally didn't get the whole... <u>gas down here and gas up here thing</u>.

This ability to be creative with the language was not evident in the discourse of the ESL students. Moreover, students such as Edward, Sara and Jeanie were able to make the necessary changes to a word so that it fits logically into the grammatical structure they have chosen to use. Edward shows this ability in his excerpt above. The following three examples show how Sara and Jeanie are able to manipulate the lexicogrammar easily, moving from appropriate processes to entities.

> Sara: So we did this experiment... to observe some substances and how they *reacted* with each other.
> Sara: So the point of our experiment was to judge whether or not... um... a *reaction* occurred between substances.
> Jeanie: When we weighed the product it was less than the *reactants*.

In contrast, not only were the ESL students not using the more grammatically metaphoric features as often as the non-ESL students, as shown in Table 13.2, but when they attempted to use them, they frequently had difficulties. They often used a single term to cover the same concept in a variety of constructions, as the following two examples show.

Belinda: When they join together they will have *reactive*.
Vicki: They... they'll *reactive*. I mean the element is easy to join. It's easy. Of course some elements not join... not *reactive*.

The ESL students had trouble with explanations because they did not appear to have the depth of language resources that the mainstream students had for constructing them. They relied more on conjunctions and circumstances than their native English-speaking peers, they struggled for the correct lexis and they took much longer to respond. Moreover, when they attempted more grammatically metaphoric constructions, they often had difficulty and aborted their efforts in favor of more congruent and familiar language. The following excerpt clearly illustrates this.

Keifer: Then with we nitro so then then we saw a when it got hot water inside... like (xx) element and (xx) water. But go inside and *(makes a whooshing sound)*.
Ken: Explodes.
Keifer: Explode. Yeah. Explode.
Tony: Yeah. That means it's very reactive.
(A few turns later.)
Tony: And sometimes some elements don't react don't act don't react with uh... water or... or air or something else.
Ken: Yeah. But some elements like iron? If you if you have water and put iron into water it will get run rust. It's un that's uh... *(17 seconds pass).* Uh... *(2 seconds pass).* Not reactive. *(6 seconds pass).* Not reactive with... it's like uh iron... iron. Put it into water oh no not iron. Not not reactive with uh element. Put it into water and **it doesn't make**... **isn't**... **it doesn't explode.**
Keifer: Yeah.
Ken: It's not reactive.

What is especially interesting in this excerpt is Ken's attempt to use a highly metaphoric, causal construction. The causal process *make* requires an entity after it ('make an explosion'), and Ken was unable to supply one. He aborted this and instead came down one level in Halliday's drift of grammatical metaphor by attempting to construct a phrase requiring a quality ('isn't explosive'), but was unable to complete that phrase as well. Finally, he finished his thoughts by using the negative of the process he had used earlier with confidence ('doesn't explode'). In other words, Ken appeared to have aimed for the higher 'entity' end of the lexicogrammar, but had to fall back to the process he was more familiar with.

Assessing Causal Explanations in the Teaching and Learning Cycle

The aborted attempts at using more grammatically metaphoric constructions, such as the attempts described above, offer excellent opportunities for using formative assessment in the teaching and learning cycle. In making an unsuccessful effort, the students can perhaps be considered to be in Vygotsky's Zone of Proximal Development (Vygotsky, 1978), where guidance and support of the expert (teacher) is needed to scaffold their development of new concepts and language. Mohan and Beckett (2003: 431) suggested that the scaffolding of ESL students through the teacher's functional recasts forms 'a ''zone of negotiation'' for reworking and reconstructing the text that [the student] is developing, as [the student] practices presenting causal explanations'. Within a classroom situation, a teacher who notices the difficulties that students are having with grammatical metaphor can open up this zone, leading to a potential *assessment-to-teaching/learning cycle*, which is a primary goal of formative assessment.

Students need to be able to explain their understandings of cause and effect in order to further their content knowledge, because their teachers often attempt to build on what they hear their students saying. Students also need to learn how their linguistic choices reflect their developing understanding of the topics they are studying, rather than simply learning correct grammatical forms. Functional recasting to help students learn how to construct more sophisticated oral causal explanations, as described in Mohan and Beckett (2003), is a useful strategy that depends on teachers being able to assess quickly and scaffold the student so that learning occurs within the zone of negotiation. Slater *et al.* (2006), in their paper on assessing projects as second language and content learning, go beyond teacher recasts by offering ways to raise *student* consciousness of the features of sophisticated causal explanations, emphasizing for them the importance of what is happening on this developmental path (see also Beckett & Slater, 2005). If teachers are consistently and reflectively assessing student explanations, focusing on aspects that students are having trouble with, they can provide successful assessment-learning cycles for teaching the forms and meanings of causal explanations. The developmental path of cause suggested in this chapter offers teachers a way to do this assessment and teaching.

This chapter has used data from both native English speakers and ELLs from primary school and high school to explore oral causal language development. The findings suggest that there appears to be a

general path of development for oral causal meanings, just as with written discourse, which moves semantically from time to cause to proof, and lexicogrammatically from conjunctions through circumstances, processes, qualities, to entities. The existence of this path provides a way to support validity arguments, as it offers important evidence to support the judgments that rate one performance of a causal explanation over another. Younger students, who focus more on *doing* science and depend mostly on conjunctions to construct their explanations, appear to be farther behind on the developmental path than older students, who are adept at manipulating the lexicogrammar and the semantics to explain their understandings of cause and effect. But it is the findings from the high school ESL students that have particular importance for the discussion of assessment. Not only do these students appear to be farther behind than their English-speaking peers on the developmental path of cause, there is evidence that they are struggling to construct the more grammatically metaphoric constructions that are typical of advanced causal explanations and a key part of higher-level literacy. Understanding this developmental path of cause allows teachers to create a formative assessment-for-learning cycle, which in turn helps to promote successful language-and-content learning. At a broader level, this model has important implications for judging validity in standardized written tests as well as academic oral proficiency interviews, neither of which has yet approached academic language assessment from such a perspective.

Note

1. All further discourse examples unless noted are from Slater (2004).

References

Beckett, G.H. and Slater, T. (2005) The project framework: A tool for language, content, and skills integration. *ELT Journal* 59 (2), 108–116.

Chapelle, C.A., Jamieson, J. and Enright, M.K. (eds) (2008) *Building a Validity Argument for the Test of English as a Foreign Language.* London: Routledge.

Fang, X. and Kennedy, G. (1992) Expressing causation in written English. *RELC Journal* 23 (1), 62–80.

Flowerdew, L. (1998) Integrating 'expert' and 'interlanguage' computer corpora findings on causality: Discoveries for teachers and students. *English for Specific Purposes* 17 (4), 329–345.

Gibbons, P. (1998) Classroom talk and the learning of new registers in a second language. *Language and Education* 12 (2), 99–118.

Halliday, M.A.K. (1993) On the language of physical science. In M.A.K. Halliday and J. Martin (eds) *Writing Science: Literacy and Discursive Power* (pp. 54–68). Pittsburgh, PA: The University of Pittsburgh Press.

Halliday, M.A.K. (1998) Things and relations: Regrammaticising experience as technical knowledge. In J.R. Martin and R. Veel (eds) *Reading Science: Critical and Functional Perspectives on Discourses of Science* (pp. 185–235). New York: Routledge.

Mislevy, R.J., Steinberg, L.S. and Almond, R.G. (2003) On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives* 1, 3–62.

Mohan, B. and Beckett, G.H. (2003) A functional approach to research on content-based language learning: Recasts in causal explanations. *Modern Language Journal* 87 (iii), 421–432.

Mohan, B. and Slater, T. (2005) A functional perspective on the critical 'theory/practice' relation in teaching language and science. *Linguistics and Education* 16 (2), 151–172.

Mohan, B., Slater, T., Luo, L. and Jaipal, K. (2002) Developmental lexicogrammar of causal explanations in science. Paper presented at the International Systemic Functional Linguistics Congress (ISFC29), Liverpool, UK.

Slater, T.J.A. (2004) The discourse of causal explanations in school science. PhD thesis, University of British Columbia.

Slater, T., Beckett, G.H. and Aufderhaar, C. (2006) Assessing projects as second language and content learning. In G.H. Beckett and P. Chamness Miller (eds) *Project-based Second and Foreign Language Education: Past, Present, and Future* (pp. 241–260). Greenwich, CT: Information Age Publishing, Inc.

Veel, R. (1997) Learning how to mean—scientifically speaking: Apprenticeship into scientific discourse in the secondary school. In F. Christie and J.R. Martin (eds) *Genre and Institutions: Social Processes in the Workplace and School* (pp. 161–195). London: Continuum.

Vygotsky, L.S. (1978) *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.