ELSEVIER
JAI

ADVANCES IN ACCOUNTING

VOLUME 23

PHILIP M. J. RECKERS
Editor

ADVANCES IN ACCOUNTING

ADVANCES IN ACCOUNTING

Series Editor: Philip M. J. Reckers

Volumes 13–22: Edited by Philip M. J. Reckers

ADVANCES IN ACCOUNTING VOLUME 23

ADVANCES IN ACCOUNTING

EDITED BY

PHILIP M. J. RECKERS

PAB Professor of Accountancy
Arizona State University, USA

CO-EDITOR

GOVIND IYER

Arizona State University, USA

ASSOCIATE EDITORS

SALVADOR CARMONA

Instituto de Empresa, Spain

ERIC JOHNSON

Indiana University, USA

LOREN MARGHEIM

University of San Diego, USA

RICHARD MORTON

Florida State University, USA

ELSEVIER
JAI

Amsterdam – Boston – Heidelberg – London – New York – Oxford
Paris – San Diego – San Francisco – Singapore – Sydney – Tokyo

JAI Press is an imprint of Elsevier

JAI Press is an imprint of Elsevier
Linacre House, Jordan Hill, Oxford OX2 8DP, UK
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
525 B Street, Suite 1900, San Diego, CA 92101-4495, USA

First edition 2008

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

ISBN: 978-0-7623-1425-6
ISSN: 0882-6110 (Series)

For information on all JAI Press publications
visit our website at books.elsevier.com

Printed and bound in the United Kingdom

08 09 10 11 12 10 9 8 7 6 5 4 3 2 1

CONTENTS

This page intentionally left blank

LIST OF CONTRIBUTORS

Asokan Anandarajan	School of Management, New Jersey Institute of Technology, Newark, NJ, USA
James Bierstaker	Department of Accountancy, Villanova School of Business, Villanova University, Villanova, PA, USA
Hsin-Yi Chi	Department of Accounting, National Taichung Institute of Technology, Taichung City, Taiwan
Chen-Lung Chin	Department of Accounting, National Chengchi University, Taipei City, Taiwan
Govind Iyer	W. P. Carey School of Business, Arizona State University, Tempe, AZ, USA
Fred A. Jacobs	Luleå University of Technology, Accounting and Control, Department of Business Administration and Social Sciences, Luleå, Sweden
Diane Janvrin	Department of Accounting, College of Business, Iowa State University, Ames, IA, USA
Steven E. Kaplan	School of Accountancy, W.P. Carey School of Business, Arizona State University, Tempe, AZ, USA
Picheng Lee	Department of Accounting, Lubin School of Business, Pace University, New York, USA

D. Jordan Lowe	Accounting Department, School of Global Management and Leadership, Arizona State University, Glendale, AZ, USA
Adam S. Maiga	School of Accounting, College of Business Administration, Florida International University, Miami, FL, USA
Linda M. Parsons	University of Alabama, Culverhouse School of Accountancy, Tuscaloosa, AL, USA
Michael J. Petersen	W.P. Carey School of Business, Arizona State University, Tempe, AZ, USA
Philip M. J. Reckers	W.P. Carey School of Business, School of Accountancy, Arizona State University, Tempe, AZ, USA
Lynn Rees	Department of Accounting, Mays Business School, Texas A&M University, College Station, TX, USA
Dahlia Robinson	College of Business Administration, University of South Florida, Tampa, FL, USA
Janet A. Samuels	Accounting Department, School of Global Management and Leadership, Arizona State University, Glendale, AZ, USA
Yen H. Tong	Division of Accounting, Nanyang Business School, Nanyang Technological University, Singapore
John M. Trussel	The Pennsylvania State University at Harrisburg, School of Business Administration, Middletown, PA, USA
Stacey Whitecotton	W. P. Carey School of Business, Arizona State University, Tempe, AZ, USA

Rebecca Wynalda Department of Accounting, Mays Business School, Texas A&M University, College Station, TX, USA

Ya-wen Yang Department of Accounting, University of Miami, Coral Gables, FL, USA

Philip M. J. Reckers
Series Editor

This page intentionally left blank

EDITORIAL BOARD

STATEMENT OF PURPOSE AND REVIEW PROCEDURES

Advances in Accounting (AIA) is a research series publication providing academics and practitioners a forum to address current and emerging issues in accounting. Manuscripts may embrace any research methodology and examine any accounting-related subject. Manuscripts may range from empirical to analytical; timely replications will be considered. Manuscripts must be readable, relevant, and reliable. To be readable, manuscripts must be understandable and concise. To be relevant, manuscripts must be related to problems facing the accounting and business community. To be reliable, conclusions must follow logically from the evidence and arguments presented. For empirical reports, sound research design and execution are critical. For theoretical treatises, reasonable assumptions and logical developments are essential.

REVIEW PROCEDURES

AIA intends to provide authors with timely reviews clearly indicating the acceptance status of their manuscripts. The results of initial reviews normally will be reported to authors within 90 days from the date the manuscript is received. All manuscripts are blind reviewed by two members of the editorial board and an Associate Editor. Editorial correspondence pertaining to manuscripts should be sent to the editor. A $75 submission fee is required.

Editorial correspondence pertaining to manuscripts should be sent to:

Philip M. J. Reckers
School of Accountancy & Information Management
W. P. Carey School of Business
Arizona State University
Box 873606
Tempe, AZ 85287-3606

Philip M. J. Reckers
Series Editor

EDITORIAL POLICY AND MANUSCRIPT FORM GUIDELINES

1. Manuscripts should be typewritten and double-spaced on 8½ by 11 white paper. Only one side of a page should be used. Margins should be set to facilitate editing and duplication except as noted:
   a. Tables, figures, and exhibits should appear on a separate page. Each should be numbered and have a title.
   b. Footnotes should be presented by citing the author's name and the year of publication in the body of the text; for example, Schwartz (1989); Lowe, Reckers and Whitecotton (2002).
2. Manuscripts should include a cover page which indicates the author's name and affiliation.
3. Manuscripts should include on a separate lead page an abstract not exceeding 200 words. The author's name and affiliation should not appear on the abstract.
4. Topical headings and subheadings should be used. Main headings in the manuscript should be centered, secondary headings should be flushed with the left hand margin. (As a guide to usage and style, refer to William Strunk, Jr., and E.B. White, *The Elements of Style.*)
5. Manuscripts must include a list of references which contains only those works actually cited. (As a helpful guide in preparing a list of references, refer to Kate L. Turabian, *A Manual for Writers of Term Papers, Theses, and Dissertations.*)
6. In order to be assured of an anonymous review, authors should not identify themselves directly or indirectly. Reference to unpublished working papers and dissertations should be avoided. If necessary, authors may indicate that the reference is being withheld for the reasons cited above.
7. Accepted manuscripts ultimately must be submitted on disk or CD.
8. Manuscripts currently under review by other publications should not be submitted. Complete reports of research presented at a national or regional conference of a professional association and "State of the Art" papers are acceptable.

9. Electronic submission of manuscripts is encouraged; send in WORD document form to Philip.reckers@asu.edu. Copies of any and all research instruments should also be included. Alternatively, four paper copies of each manuscript may be submitted to the Editor-In-Chief.
10. The author should send a check by mail for $75.00 made payable to *Advances in Accounting* as a submission fee to:

Philip M. J. Reckers
Series Editor

THE EFFECT OF INNOVATIVE ACTIVITY ON FIRM PERFORMANCE: THE EXPERIENCE OF TAIWAN

Asokan Anandarajan, Chen-Lung Chin, Hsin-Yi Chi and Picheng Lee

ABSTRACT

Firm performance is, among other factors, a function of tangible and intangible assets that the firm possesses and utilizes to maximize value. While research has examined the impact of advertising and R&D on firm performance, in this chapter we, in addition, examine the extent of innovation as measured by patents granted on firm performance. Our findings indicate that overall, innovative activity as measured by number of patents granted, significantly influences firm performance as measured by Tobin's q. Hence patenting activity is value relevant to investors. Further, this relationship is more pronounced when the patents are granted in the United States. We conclude that markets tend to give greater credence to innovative activity when patents are granted to foreign firms by the U.S. Patenting Office. Finally, the stage of the product in the industry chain moderates the influence of patenting activity on firm performance. When patents are granted in the design stage the impact on firm performance is stronger relative to when patents are granted in the

Advances in Accounting, Volume 23, 1–30

ISSN: 0882-6110/doi:10.1016/S0882-6110(07)23001-9

manufacturing and packaging and testing stages. The evidence indicates that patents by companies in the manufacturing end of the chain have a more pronounced impact on firm performance relative to patents granted in the packaging and testing stage.

1. INTRODUCTION

Much research has been conducted on the association of innovative activity and firm performance. Sher and Yang (2005) define innovative activity as any incremental or radical change in technology embodied in product and process. Innovative activity is seen as critical to a firm in achieving strategic competitiveness (Conner, 1991). Hitt, Hoskisson, and Ireland (1994) and Zahra, Ireland, and Hitt (2000) note that innovations enable a firm to offer a greater variety of significantly differentiated products and therefore should lead to a higher level of financial performance. While innovative activity is a difficult construct to measure, in this study it is measured by the number of patents granted to a firm both in Taiwan and the United States. This is because innovation or inventiveness, as represented by intellectual property rights, is protected by national patent offices within national borders meaning that a patent protects an idea in one country and in one market. This idea cannot be copied by other firms in that country; this ensures that the result is not confounded by commercial activities of other firms (e.g. Grupp & Schmoch, 1999). Measures based on patents granted to a firm have been also considered to be a better and more reliable indicator of the technological competitiveness of the firm in the prior literature (see Pegels & Thirumurthy, 1996; Watanabe, Tsuji, & Griffy-Brown, 2001).

The first objective of this chapter is to investigate whether patents granted in Taiwan and in the United States for Taiwanese semiconductor firms have a differential impact on firms' performance as measured by Tobin's q.[1] This study builds on the findings of Shane and Klock (1997) and Chin, Lee, Chi, and Anandarajan (2006) that patenting activity in the U.S. (refer Shane & Klock) and in Taiwan (refer Chin et al.) is viewed by the market as evidence of innovation and is positively associated with increase in firm value. In this respect, our study is a trangulation of the Shane and Klock and Chin et al. studies. Watanabe et al. (2001) state that patent applications to foreign countries, in particular the United States, provide a better demonstration of innovational ability; in particular the frequency or number of patent applications in the United States is, according to them, a strong indicator of inventive ability. In addition, they note that patents granted by the United States Patent and Trademark Office (USPTO) enhance a firm's visibility and exposure. Therefore, in our study, we expect patents granted in the United States to send a stronger signal and have a greater influence on firm performance relative to patents granted in Taiwan.

The second objective of this chapter is to examine whether the stage of a product's life cycle in the semiconductor industry moderates the association between patenting activity and firm performance as measured by Tobin's *q*. We especially examine if the moderating influence is accentuated by location of patent filing (Taiwan versus United States). Semiconductor products include discrete devices, optoelectronics, and integrated circuits, with the last of these accounting for the bulk of all semiconductors. The semiconductor industry was selected for the current study because of the strategic importance of intangible capital such as patents in this segment of the electronics industry.[2] Hall and Ziedonis (2001) indicate that the semiconductor industry provides an excellent setting within which to examine the association of patent activity and a firm's ''rapidly advancing and cumulative technology'' (p. 102). In the United States products of firms in the semiconductor industry go through several stages, namely, design, manufacturing, and packaging and testing. In Taiwan, semiconductor firms specialize in one of the above value-added activities, namely, either design or manufacturing or packaging and testing. Hence, while prior studies focus more on the association between patent and firm's performance, our study provides a unique opportunity to examine how the stage of the industry value chain moderates the association between extent of patenting activity and Tobin's *q*. We examine whether the impact of patenting activity on firm performance is influenced by the different stages in the semiconductor industry value chain.

In this study, firm performance is measured using Tobin's *q*. Tobin (1978) argues that *q* is a measure of profitable investment opportunities. Tobin's *q* was selected in this study as the dependent variable to surrogate for firm performance because it has been used extensively in the literature (e.g. Jaffe, 1986; Shane & Klock, 1997; Bharadwaj, Bharadwaj, & Konsynski, 1999). For the purpose of sensitivity analysis, in addition to frequency of patents granted, we also use another measure, namely, value of patents. We estimated patent values using the Cobb–Douglas production function as used by Seethamraju (2003) and estimated the value of each patent at different value chain stages. We re-conducted our tests using this measure and our results remained unchanged.

Our empirical findings, based on 279 Taiwan semiconductor firm-year observations covering the period of 1990–2002, indicate that patents granted in the United States exert greater influence than those in Taiwan on firm performance as measured by Tobin's q. In addition, patents filed in the design sector have greater influence on performance relative to patents filed in other sectors. We also find that the estimated values of patents are higher for the firms in the design sector relative to the firms in the other sectors, consistent with previous market-based results.

This chapter is different from prior research and contributes to the literature on the association between patenting activity and Tobin's q in two respects. First, we examine the influence of extent of patenting activity (surrogating for innovation) for firms in Taiwan. Most of the research in this area has been conducted in the United States. This research is important because we show whether the association between patent filing and Tobin's q also holds in environments other than that of the United States. If the same relationship holds we can conclude that the general theory is not merely limited to the United States but has external validity. We find that the moderating influence of industry on the association between patenting activity and firm performance is magnified by location of patent filing. The moderating influence is stronger in the United States relative to Taiwan. Second, this moderating influence also applies to the different stages of the value chain in the semiconductor industry.

In the accounting literature this study contributes to the specific niche in financial accounting research that focuses on the value relevance of non-financial information to investors. This particular niche in financial accounting research has gained in importance because Brown and Lys (1999) documented a long-term decline in the value relevance of financial statement information as an important determinant in the market's valuation of a firm. Francis and Schipper (1999) noted the importance of research examining non-financial factors that could potentially have value relevance for investors in the light of "financial statements losing a significant portion of their relevance for investors". In this field, Amir and Lev (1996) found that population size and market penetration were highly value relevant. Hirschey, Richardson, and Scholz (2001) studied 199 U.S. high-tech firms and pointed out that non-financial information was particularly relevant in the high-technology sector where productivity was a vital determinant of long-term success. They concluded a positive association noting that "non-financial information concerning the quality of inventive output appears to sharpen the investors' perception of the on going value created by the firm's inventive and innovation activity". This chapter contributes to the literature on value of non-financial information by building on the Hirshey et al. study and examining the association of patenting activity with firm value in a different environment. Our finding that patenting activity is value relevant enhances the external validity of the Hirschey et al. findings in the global environment. Our findings also show that value relevance of non-financial information (as surrogated by patenting activity) is contingent on the position of the firm in the value chain.

Finally, this chapter contributes to a niche in management accounting literature, namely, research focusing on value chain relationships. This line of study focuses on the added value of each step in a firm's value chain. Shank and Govindarajan (1992) used a case study representing the paper products industry to show how costs can be analyzed along the value chain to determine value added in each stage of the chain. Dekker (2003) conducted a similar analysis with Sainsbury's, a large United Kingdom retailer. Chang and Hwang (2002) examined the value chain of 65 U.S. and 34 Hong Kong companies. They found that U.S. companies invested most of their resources in upstream activities relative to their Hong Kong counterparts. In this study we also contribute to the value chain literature in management accounting by showing specifically that, in the high-tech industry, innovative activity at the upstream end of the value chain magnifies perceptions of added value relative to innovative activities downstream.

Our chapter is presented as follows. In Section 2 we discuss the semiconductor industry. In Section 3 we present the results of prior studies in this area. We develop our hypotheses in Section 4. We discuss our sample in Section 5 and our overall methodology and regression equations in Section 6. Section 7 presents our conclusions.

2. THE SEMICONDUCTOR INDUSTRY

2.1. The Industry in General

Overall, the semiconductor industry is characterized by continuous innovation and has advanced exponentially since its inception in the middle of the twentieth century. The percentage share of semiconductors in global electronics production rose from 2% to 3% in the 1960s to just fewer than 10% in the 1980s and currently is expected to rise to between 25% and 30%.

The electronics industry is composed of "upstream" and "downstream" sectors. Semiconductors, and other parts and components, are in upstream sectors; consumer electronics, telecommunications, and information technology products are in the downstream sectors. The semiconductor industry can also be divided into four core businesses based on production stages (design, manufacturing/fabrication, packaging, and testing). The four core businesses have different characteristics. The design business is knowledge intensive and requires little capital fabrication technology; fabrication is capital intensive, while packaging and testing are both labor intensive in nature. Each sector, however, adds value to the final products that are fed into the downstream sector.

2.2. The Industry in Taiwan

The electronics industry has become the most dynamic sector in East Asia in the last two decades. The region as a whole has exported huge volumes of PCs, disk drives, semiconductors, televisions, and telephones. The commencement of semiconductor production in Taiwan dates back to the 1960s. In 1966, General Instruments of the United States set up the first semiconductor plant in Taiwan. Other firms followed suit. In the 1970s, the government decided to stimulate indigenous production capacity in semiconductors with a view to upgrading the overall technological level. Chen (2002) in an overview noted that the development of the semiconductor industry in Taiwan was the result of a strategy of vertical disintegration which, in turn, facilitated the formation of local and cross-border linkages in pursuit of industrial expansion.

Currently, in Taiwan, the Taiwan Semiconductor Manufacturing Corporation (TSMC)[3] has become the world's leading foundry plant and has grown into the world's fourth largest semiconductor company in terms of market value next only to Intel, Texas Instruments, and Applied Materials in the United States (e.g. Tung, 2001). In 1981, the electronics industry accounted for only 6.8% share of Taiwan's manufacturing production. The industry, however, took off quickly. By 1996, it surpassed textiles and became the largest manufacturing sector with a percentage share of around 11%. In 1998, the ratio rose further to 26%. Within the electronics industry, the dominance of consumer electronics has fallen sharply against the rapid growth of information technology products and semiconductors. The total value of semiconductor production increased at an average annual rate of slightly over 28% in the five years (1993–1998). Currently, Taiwan has become the world's fourth largest semiconductor producer after United States, Japan, and South Korea. Taiwan's semiconductor fabrication industry grew by a total of \$8 billion in 1999 for overall growth of 57.6%. The backend of the business also benefited from strong growth in the fabrication front end. Testing grew by 49.7% and packaging by 63.3%. In 1999 Taiwan's semiconductor industry generated total revenues of $12.5 billion for a growth of 48.1% over the previous year. Due to the prominence and importance of the semiconductor industry in Taiwan, we feel that this is a suitable environment for examining the influence of innovation on firm performance. For a detailed overview of the semiconductor industry in Taiwan please refer Mathews (1997), Chang and Tsai (2000), Tung (2001), and Chen (2002).

2.3. Analysis of Value Added in Each Sector

All processes discussed above distinctly add value to the product. As we move from upstream to downstream, in the design sector, Sher and Yang (2005) note that added value is in the form of continuous research to enhance core technologies. In the middle of the value chain, namely, the manufacturing sector, the focus is on strictly applying the designs set. The manufacturing sector adds value by finding measures to improve controllability and reducing testing time. This type of added value is different in nature to the added value in the design sector. If we consider the downstream end, namely, the packaging and testing sector, the priority is on shrinking package size, reducing weight to ensure more "compactness" and focus on increasing connecting density and lowering heat radiation (refer Sher & Yang, 2005). We postulate that the type of innovation is viewed differently at each stage with the value added at the front end (innovative activity focusing on enhancing core technology) being given greater weight relative to the middle end (innovative activity focusing on improving controls and reducing testing time) and least weight to the downstream end (innovation focusing on increasing compactness, shrinking package size, and lowering heat radiation).

3. LITERATURE REVIEW

Prior research examining the influence of innovation (as measured by patenting activity) on firm performance has broadly used two types of models to measure the dependent variable. In the first category, some researchers measured the dependent variable in terms of the unexpected information (e.g. major change in the independent variable of interest) on the change in market value. In the second category, researchers measured the dependent variable in terms of Tobin's q.

In the first category, most of the studies that examined patenting activity also included R&D as an explanatory variable. For example, Griliches (1981) looked at the impact of unexpected information (i.e. a major change in a company's R&D expenditure) on the change in market value. Griliches (1981) used a panel of 157 U.S. firms and examined the impact of both R&D and patenting activity on market value. He found a positive association, namely, increased R&D and patenting activity influenced investors' perceptions of firm value. In the study by Stoneman and Toivanen (1997) the coefficient on the patent variable was insignificant when both R&D and patent measures were included in the specification. Thus, they concluded that patents did not have incremental value relative to R&D. In general, however, the majority of studies did not come to similar conclusions. Most studies found a positive association and concluded that the extent of patenting activity represented level of innovation, which enhanced the value of the firm to potential investors (Griliches, Pakes, & Hall, 1987; Cockburn & Griliches, 1988; Hall, 2000). Griliches (1990) and Griliches, Hall, and Pakes (1991) also explored whether there was additional information on the rate and output of inventing activity measured in terms of patent numbers above and beyond that already contained in R&D expenditure data. Except in the pharmaceutical industry they found little evidence of this.

In general, most studies in the first category limited their investigations of the association of innovation to market performance with the two proxies cited above, namely, patenting activity and R&D activity. Few studies in this category have moved outside of these two proxies. However, a number of studies incorporated additional variables either for the purpose of study or as control variables. In one of the few studies that moved beyond these two proxies, Hall (1993) included advertising, and Bosworth and Mahdian (1999) included trademarks. Bosworth and Mahdian (1999) found evidence that R&D, patents, and trademarks all played a significant role in explaining market value of United Kingdom pharmaceutical companies. Toivanen, Stoneman, and Bosworth (2002) included four groups of explanatory variables, namely, the firm's debt equity ratio, the change in log of sales to the firm, and cash flow to assets in addition to R&D expenditure. In the first category, while the above studies used change in market value as the dependent variable, other studies used a variant, namely, a dependent variable reflecting the difference between market value and book value of assets (Connolly & Hirschey, 1988, 1990; Greene, Stark, & Thomas, 1996).

The main objective of the second group of studies was to examine how innovation (as measured by patenting activity) influenced investors' perception of the firm as measured by Tobin's *q*. Cockburn and Griliches (1988) examined whether investing in research and development sent a positive signal to the market. Megna and Klock (1993) examined whether investment in research and development and extent of patenting activity influenced investor perceptions of a firm. They found that increased levels of research and development and the number of patents granted (surrogating for innovative activity) had a positive association with Tobin's q implying that investors perceived these activities positively.

4. HYPOTHESIS DEVELOPMENT

Initially we test if, irrespective of the stage of the industry value chain, whether, overall, there is association between added value as measured by patents granted and Tobin's q. Since Taiwanese companies file patents in both Taiwan and the United States, we test whether the association holds irrespective of the country of filing. Prior research involving patenting activity has all concluded that patenting activity should influence perceptions of firm performance (Griliches, 1981; Griliches et al., 1987).

The United States is the biggest export destination of Taiwan semiconductor products.[4] Filing patents in the United States increases legal intellectual property right protection internationally. In addition, patents filed to and granted by the USPTO enhance a firm's visibility and exposure. Hall and Ziedonis (2001) observe that patents filed in the United States should send a strong signal of innovative activity. Based on above discussion, H1a and H1b are developed as follows:

H1a. Innovative capability as measured by number of patents granted in Taiwan has a positive association with Tobin's q.

H1b. Innovative capability as measured by number of patents granted by the USPTO to Taiwanese firms has a positive association with Tobin's q.

Watanabe et al. (2001) state that patent applications to foreign countries, in particular the United States, provide a better demonstration of innovational ability; in particular the frequency or number of patent applications in the United States is, according to them, a strong indicator of inventive ability. In addition, they note that patents granted by the USPTO enhance a firm's visibility and exposure. Further, Grupp and Schmoch (1999) note that international filing of patents other than in the country of origin is a greater indicator of innovative ability.

Hall and Ziedonis (2001) and Grupp and Schmoch (1999) observe that patents filed in the United States should send a strong signal of innovative ability relative to patents filed in other countries. We hence postulate that patents filed in the United States by Taiwanese firms will send a stronger signal of innovative ability relative to patents filed by Taiwanese firms in Taiwan. This is because patents are far more complicated and costly to file in the United States. Thus, only ''real'' innovations are filed. Watanabe et al. note that in Taiwan and Japan, for example, a number of patent applications are ''pseudo'' innovation including decoys for the purpose of establishing defenses against competitors. They note that Japanese and Taiwanese firms apply for patents in the USPTO in a very selective way and do not include applications for decoys but only innovation that is ''really worthwhile''. If this holds true, then patents granted by the USPTO to Taiwanese firms should send a stronger signal of innovative ability.

Therefore, in our study, we expect patents granted in the United States to send a stronger signal and have a greater influence on firm performance relative to patents granted in Taiwan. Hence, hypothesis (H2) is stated as follows:

H2. Patents granted by the USPTO to Taiwanese firms have greater influence on Tobin's q relative to patents granted to Taiwanese firms by the Taiwan patent office.

While the different stages of the manufacturing process have been discussed, design is the only stage to be knowledge intensive (e.g. Tung, 2001). The other stages are relatively capital and labor intensive. While capital and labor are readily available and can be accumulated in the course of business, the skills and creativity required by scarce, highly specialized people to create new knowledge (as is required in the design sector) is a valuable commodity that cannot be easily acquired or accumulated. This, in Taiwan and the United States, is currently accentuated by further specialization in the design sector that has resulted in miniaturization of electric circuitry, which requires even more skilled labor (see Macher, Mowery, & Simcoe, 2002). Sher and Yang (2005) further note that the type of innovation varies distinctly by sector. They observe that the upstream sector in the value chain of semiconductor industry is more innovative than its downstream counterparts. In the design sector, we expect that patenting activity relates to methods to enhance core technologies, including, as noted above, miniaturization of electric circuitry; in the manufacturing sector innovation relates to the ''controllability'' aspects (reducing time required to test the circuitry) while at the packaging and testing stage the focus is on shrinking package size and lowering heat radiation, among others. Therefore, we hypothesize that patenting activity in each is evaluated differently by the market.

Furthermore, a report by the Ministry of Economic Affairs (2001) in Taiwan also indicates that the entry barrier for the design sector market is greater relative to the other sectors. This is because the design sector is stated to be more sophisticated with reduced possibility of imitation. However, the entry barrier to manufacturing, packaging and testing sectors are much lower and characterized by more intense competition. Thus, due to more extreme challenges, patents filed in the design sector should send a stronger signal of inventive activity relative to other sectors. Hence the third and fourth hypotheses are stated as follows:

H3. Patents granted in the design sector have greater influence on Tobin's q relative to patents granted in other sectors in Taiwan.

H4. Patents granted by the USPTO to Taiwanese firms in the design sector have greater influence on Tobin's q relative to patents granted to Taiwanese firms in other sectors.

5. SAMPLE SELECTION

The study is based on the semiconductor firms publicly listed in the Taiwan Stock Exchange. We reviewed a twelve-year period from 1990 to 2002. The Taiwan Economic Journal (TEJ) database was used to identify all semiconductor firms with complete equity returns and accounting data needed for this study. Taiwan's Patent data were collected from Taiwan Patents Database provided by Asia Pacific Intellectual Property Association (APIPA).[5] United States patents data were collected from National Bureau of Economic Research (NBER) website and USPTO.[6]

Table 1 summarizes our data. In Table 1, the stages of the value chain shown were obtained from the database of the Ministry of Economic Affairs (2003) semiconductor industry year book (Taipei). As shown in panel A we had a total of 279 firm years of which 82 were in the design end, 82 in the manufacturing end, and 115 in the packing and testing end of the value chain of the semiconductor industry. In panel B, we identified a total of 60 firms representing 279 firm-year observations. Panel C of Table 1 shows the number of patents granted to these companies in Taiwan. A total of 503 patents were granted to companies in Taiwan for those in design and 394 to those

Table 1. Sample Distribution.

Panel A: Firm-year distribution in the semiconductor industry value chain

Year	1990–1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	Total
All	16	7	13	14	21	28	32	40	50	58	279
Design	3	1	1	2	5	7	8	11	19	24	82
Manufacture	8	3	6	6	7	9	11	11	11	11	82
Packaging & testing	5	3	6	6	9	12	13	18	20	23	115

Panel B: The frequency of sample firms in the study period

Frequency	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
Number of firms	12	9	7	5	8	6	1	7	1	1	0	1	3	60

Year	1990–1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	Total
Panel C: Patents granted in Taiwan											
All	39	126	354	452	716	764	1,125	2,200	2,540	1,834	10,152
%	0.4	1.2	3.5	4.5	7.1	7.4	11.1	21.7	25.0	18.1	100.0
Design	2	1	1	4	5	9	18	60	174	229	503
%	0.4	0.1	0.1	0.8	1.0	1.8	3.7	11.9	34.7	45.5	100.0
Manufacture	35	124	343	430	702	742	1,096	2,094	2,213	1,476	9,255
%	0.4	1.4	3.7	4.6	7.7	8.0	11.8	22.6	23.9	15.9	100.0
Packaging & testing	5	1	10	18	9	13	11	46	153	129	394
%	1.3	0.3	2.4	4.6	2.3	3.2	2.8	11.7	38.7	32.7	100.0
Panel D: Patents granted in US											
All	29	59	198	304	232	562	855	1,245	1,469	1,145	6,095
%	0.4	0.9	3.3	4.9	3.8	9.3	14.0	20.5	24.1	18.8	100.0
Design	0	0	0	0	0	10	9	28	40	72	159
%	0.0	0.0	0.0	0.0	0.0	6.2	5.7	17.6	25.2	45.3	100.0
Manufacture	26	59	195	302	231	546	839	1,205	1,386	998	5,787
%	0.4	1.0	3.3	5.3	3.9	9.5	14.6	20.8	23.9	17.3	100.0
Packaging & testing	3	0	3	2	1	16	7	19	43	75	169
%	1.8	0.0	1.8	1.2	0.6	9.5	4.1	11.2	25.4	44.4	100.0

Panel E: The difference in the number of patents granted in Taiwan and US

	Taiwan		US		Difference
	N	%	N	%	z-statistics
All	10,152	100.0	6,095	100.0	
Design	503	4.9	159	2.6	1.378
Manufacture	9,255	91.2	5,787	94.9	1.592
Packaging & testing	394	3.9	149	2.5	1.137

companies in the packaging and testing end of the value chain. We included packaging and testing as one entity even though we differentiated them in the introduction. Sher and Yang (2005) observe that, though initially distinct, over time these two activities have blended into one entity and are conducted by the same companies due to the necessity to meet market demands quickly and efficiently. Once the packaging process is complete, virtually all companies also test the product rather than waste time and money sending it out for inspection. The majority of patents (9,255) were granted to those in the manufacturing end of the value chain.

Panel D of Table 1 shows the number of patents granted to these companies in the United States. Once again the bulk of the patents (5,787) were granted to those in the manufacturing end of the value chain. Companies in the design end of the value chain were granted 159 patents while those in packing and testing granted a total of 169 patents. However, panel E of Table 1 indicates that difference in the percentage of granted patents at the various stages of semiconductor value chain is insignificant. As shown in Table 1, design percentages granted were 4.9% in Taiwan relative to 2.6% in the U.S. manufacture percentages in Taiwan and U.S. were respectively, 91.2% and 94.9%; and packaging and testing 3.9% and 2.5% between the respective countries. The differences were not, as previously mentioned, statistically significant. This proves that, the companies in our sample are a relatively homogenous group and not significantly different in terms of patenting/innovative activity.

Table 2 provides descriptive statistics of the variables used in our sample broken down by different stages in the semiconductor industry value chain. Each panel consistently indicates that number of patents granted in Taiwan is greater than that granted in U.S. However, the estimated value of patents granted in the United States is greater than that in Taiwan (Section 6.3.1 provides the details of patent value estimation).

6. METHODOLOGY AND RESULTS

6.1. Tests of Hypotheses 1 and 2

The variables used in our study and measurements are shown in Exhibit 1.

All variables shown in Exhibit 1 were selected from prior literature. For example, some studies postulate that level of advertising could influence valuation (Morck & Yeung, 1991; Simon & Sullivan, 1993; Wu & Bjornson,

Table 2. Descriptive Statistics.

Variables	Mean	Standard Deviation	Quartile		
			25%	50%	75%
Panel A: All stages					
Tobin's q	1.273	1.156	0.523	0.917	1.582
PAT_TW	1.000	1.450	0.110	0.505	1.199
PAT_US	0.946	1.827	0.000	0.102	0.873
Advertising expense	0.002	0.005	0.000	3.681×10^{-4}	0.002
Leverage	0.335	0.143	0.236	0.322	0.444
SIZE	6.839	0.677	6.296	6.700	7.449
R&D	0.079	0.051	0.021	0.057	0.114
Patent value-Taiwan	6.785×10^{6}	3.820×10^{7}	1.050×10^{5}	7.920×10^{6}	1.199×10^{7}
Patent value-US	7.495×10^{6}	4.782×10^{7}	0.000	9.772×10^{6}	1.841×10^{7}
Panel B: Design stage					
Tobin's q	1.707	1.170	0.804	1.444	2.473
PAT_TW	1.000	1.269	0.348	0.685	1.100
PAT_US	0.915	1.905	0.000	0.000	0.573
Advertising expense	0.002	0.004	0.000	3.893×10^{-4}	0.003
Leverage	0.261	0.152	0.137	0.241	0.335
SIZE	6.449	0.472	6.110	6.471	6.715
R&D	0.129	0.077	0.086	0.111	0.146
Patent value-Taiwan	8.743×10^{6}	5.848×10^{7}	2.080×10^{6}	1.245×10^{7}	2.406×10^{8}
Patent value-US	1.014×10^{7}	5.704×10^{8}	0.000	1.524×10^{7}	3.674×10^{8}
Panel C: Manufacturing stage					
Tobin's q	1.205	1.276	0.511	0.838	1.326
PAT_TW	1.000	1.445	0.077	0.251	0.884
PAT_US	1.000	1.380	0.119	0.429	1.119
Advertising expense	0.002	0.003	0.000	4.276×10^{-4}	0.002
Leverage	0.371	0.113	0.285	0.372	0.464
SIZE	7.593	0.543	7.422	7.692	7.875
R&D	0.092	0.087	0.330	0.794	0.124
Patent value-Taiwan	6.925×10^{6}	3.402×10^{7}	1.503×10^{5}	7.925×10^{6}	3.589×10^{6}
Patent value-US	7.581×10^{6}	3.602×10^{7}	0.000	9.798×10^{6}	1.535×10^{7}
Panel D: Packaging and testing stage					
Tobin's q	0.935	0.783	0.411	0.717	1.230
PAT_TW	1.000	1.581	0.069	0.339	1.440
PAT_US	0.930	2.051	0.000	0.000	0.542
Advertising expense	0.002	0.006	0.000	3.264×10^{-4}	0.001
Leverage	0.363	0.137	0.257	0.365	0.458
SIZE	6.579	0.404	6.244	6.521	6.789
R&D	0.034	0.049	0.014	0.022	0.037
Patent value-Taiwan	1.002×10^{6}	1.928×10^{7}	7.617×10^{5}	1.998×10^{6}	5.086×10^{6}
Patent value-US	2.348×10^{6}	2.011×10^{7}	0.000	2.932×10^{6}	7.904×10^{6}

Exhibit 1. Definition and Measurement of Variables.

Variables	Definition	Measurement
Q_{it}	Tobin' *q* score	(Book value of total assets + market value of equity minus the book value of equity)/Book value of total assets
PAT_TW_{it}	The industry adjusted patent count granted in Taiwan	Number of patents granted in Taiwan for firm *i* in year *t*, scaled by the industry average number of patents in the same year
PAT_US_{it}	The industry adjusted patent count granted in USA	Number of patents granted in US for firm *i* in year *t*, scaled by the industry average number of patents in the same year
AE_{it}	Advertising expense ratio	Advertising expense/Net sales for firm *i* in year *t*
LEV_{it}	Leverage ratio	Total liabilities/Total assets for firm *i* in year *t*
SIZE_{it}	Firm size	Logarithm of the total assets
$\text{D}_1\text{\_TW}_{it}$	Design sector dummy	$\text{D}_1\text{\_TW}_{it}$ is a dummy variable that represents 1 if the firm is in the design sector and 0 if the firm is in other sector
$\text{D}_2\text{\_TW}_{it}$	Manufacturing sector dummy	$\text{D}_2\text{\_TW}_{it}$ is a dummy variable that represents 1 if the firm is in the manufacturing sector; 0 if the firm is in other sectors
$\text{D}_1\text{\_US}_{it}$	Design sector dummy	$\text{D}_1\text{\_US}_{it}$ is a dummy variable that represents 1 if the firm is in the design sector and 0 if the firm is in other sectors
$\text{D}_2\text{\_US}_{it}$	Manufacturing sector dummy	$\text{D}_2\text{\_US}_{it}$ is a dummy variable that represents 1 if the firm is in the manufacturing sector; 0 if the firm is in other sectors
RD_{it}	R&D intensity	Total research and development expense/Net sales for firm *i* in year *t*
YEAR_t	Dummy of year	Yearly dummy variable, $t = 1990, 1991, \ldots, 2001$

1996; Bharadwaj et al., 1999). Other studies postulated that leverage should be negatively correlated with firm valuation as higher levels of debt may be viewed more negatively by investors (Morck, Shleifer, & Vishny, 1988). Some studies postulate that research and development expenditures should be value relevant to investors (Cockburn & Griliches, 1988). All studies noted above included log of assets to control for effects of firm size and we did the same. We did not use any new variables and limited the analysis to variables used in prior studies. An argument could be made that tests do not rule out any change in firm performance due to other factors besides increasing patenting activity. One solution is to use a control group of firms without patents. Another is to standardize all variables. We chose the latter strategy and scaled all variables in our study by the industry average number of patents in the same year.

Table 3 shows the correlation between the independent variables in our study. As shown in the correlation matrix, the only significant correlations relate to the association between research and development expenditure (R&D) and advertising expenditure and the association between R&D and leverage. We conducted a variance inflation factors test and the results of this test do not indicate the existence of potential multicollinearity problems.

In order to test the first two hypotheses we ran three regression models. The dependent variable was Tobin's q. Tobin's q was calculated using book value of total assets plus market value of equity minus the book value of equity as the numerator and book value of total assets as the denominator consistent with Doidge, Karolyi, and Stulz (2004). Further, consistent with earlier studies using Tobin's q, advertising expense, research and development expenditure and leverage ratios were included as control variables (Park, Jaworski, & Macinnes, 1986; Aaker, 1991; Megna & Mueller, 1991;

Table 3. Pearson Correlation Matrix.

Variables	Q	PAT_TW	PAT_US	AE	LEV	SIZE
PAT_TW	0.270					
	(0.000)					
PAT_US	0.491	0.025				
	(0.000)	(0.127)				
AE	0.147	0.256	0.048			
	(0.014)	(0.000)	(0.421)			
LEV	−0.323	−0.164	−0.051	0.069		
	(0.000)	(0.006)	(0.394)	(0.246)		
SIZE	−0.130	0.285	0.278	0.048	0.167	
	(0.030)	(0.000)	(0.000)	(0.424)	(0.005)	
RD	0.094	0.012	0.079	0.147	−0.161	0.008
	(0.118)	(0.848)	(0.188)	(0.014)	(0.072)	(0.895)

Note: This table presents the Pearson correlation coefficients of variables and the p-values of the coefficients (in parentheses). See Exhibit 1 for definition of other variables. The sample consists of 279 observations for years 1990–2002.

Morck & Yeung, 1991; Simon & Sullivan, 1993; Wu & Bjornson, 1996; Bharadwaj et al., 1999). The three models are shown below:

$$Q_{it} = \alpha_0 + \alpha_1 \mathrm{PAT\_TW}_{it} + \alpha_2 \mathrm{AE}_{it} + \alpha_3 \mathrm{LEV}_{it} + \alpha_4 \mathrm{SIZE}_{it} + \alpha_5 \mathrm{RD}_{it} + \sum_{y=1990}^{2001} \alpha_{6y} \mathrm{YR}_y + \varepsilon_{it} \tag{1}$$

$$Q_{it} = \alpha_0 + \alpha_1 \mathrm{PAT\_US}_{it} + \alpha_2 \mathrm{AE}_{it} + \alpha_3 \mathrm{LEV}_{it} + \alpha_4 \mathrm{SIZE}_{it} + \alpha_5 \mathrm{RD}_{it} + \sum_{y=1990}^{2001} \alpha_{6y} \mathrm{YR}_y + \varepsilon_{it} \tag{2}$$

$$Q_{it} = \alpha_0 + \alpha_1 \mathrm{PAT\_TW}_{it} + \alpha_2 \mathrm{PAT\_TW}_{it} + \alpha_3 \mathrm{AE}_{it} + \alpha_4 \mathrm{LEV}_{it} + \alpha_5 \mathrm{SIZE}_{it} + \alpha_6 \mathrm{RD}_{it} + \sum_{y=1990}^{2001} \alpha_{7y} \mathrm{YR}_y + \varepsilon_{it} \tag{3}$$

where,

$\mathrm{PAT\_TW}_{it}$ = Number of patents granted in Taiwan for firm *i* in year *t*
$\mathrm{PAT\_US}_{it}$ = Number of patents granted in the United States for firm *i* in year *t*
AE_{it} = Advertising expense/Net Sales for firm *i* in year *t*
LEV_{it} = Leverage ratio as measured by total liabilities/total assets for firm *i* in year *t*
SIZE_{it} = Logarithm of the total assets
RD_{it} = Total research and development expense/Net sales for firm *i* in year *t*.
Year = Dummy variable representing for a specific year commencing 1990 up to 2001.

In the first model shown in Eq. 1, independent variables are the number of patents granted in Taiwan ($\mathrm{PAT\_TW}_{it}$), advertising expense (AE_{it}), a variable to measure the company's solvency position (LEV_{it}) and firm size representing total assets (SIZE_{it}), research and development expenditure (RD_{it}), and a dummy variable (Year_t) representing the year in our sample. We include advertising, solvency, firm size, and research and development expenditure as control variables since past studies have shown that these variables can influence Tobin's *q*. Since we are dealing with panel data, each of our models is, in effect, a fixed effects model controlling for the years in our sample. The use of the random effect model as a sensitivity test is discussed in greater detail in Section 6.3.3. In the second model shown in Eq. 2, we replace the number of patents granted in Taiwan (PAT_TW_{it}) with patents granted in the United States (PAT_US_{it}). In the third model shown in Eq. 3 we include both patents granted in the United States and patents granted in Taiwan. The above three models allow us to investigate differences in the relationship between patenting activity and firm performance in Taiwan and the United States. For parsimony in terms of presentation, we do not report the coefficient estimates and significance levels associated with the year dummy variables. The results of the three regression models are shown in Table 4.

In the column (1) of Table 4, the coefficient of the variable PAT_TW_{it}, namely the patents granted in Taiwan (0.047) is significant at the 5% level, suggesting innovation measured by granted patents is positively related to firm's performance as predicted in our hypothesis H1a. The coefficient of the

Table 4. A Comparative Analysis of Patent Granted in Taiwan and US on Tobin's q.

Variables	Predicted Sign	(1)	(2)	(3)
α_0		3.416	3.786	3.636
		(2.81)***	(3.11)***	(2.97)***
PAT_TW_{it}	+	0.047		0.096
		(1.83)**		(1.70)**
PAT_US_{it}	+		0.055	0.109
			(1.82)**	(1.85)**
AE_{it}	+	4.630	8.407	2.771
		(0.18)	(0.32)	(0.11)
LEV_{it}	–	−4.718	−4.564	−4.728
		(5.56)***	(5.47)***	(5.58)***
$SIZE_{it}$	–	−0.064	−0.014	−0.025
		(0.35)	(0.08)	(0.17)
RD_{it}	+	1.997	2.238	2.163
		(1.39)*	(1.51)*	(1.46)*
$PAT\_US_{it} > PAT\_TW_{it}$				$F = 2.498$***
N		279	279	279
F-statistics		7.542	7.627	6.559
Adj. R^2		0.105	0.106	0.107

Notes: For parsimony in terms of presentation, we do not report the coefficient estimates and significance levels associated with the year dummy variables.
Statistical significance is based on two-tailed test:
*signifies statistical significance at 10% level.
**signifies statistical significance at 5% level.
***signifies statistical significance at 1% level.

advertising expense variable is not significant. The result is consistent with findings by Chin, Lin, and Hong (2003). The coefficient of the leverage variable is negative and statistically significant at the 1% level. This again is consistent with past studies that show a negative association between leverage and Tobin's q. The coefficient of the asset variable is not statistically significant. The coefficient of the research expenditure variable (RD_{it}) can be considered to be only marginally significant (significant at the 10% level). In the column (2) of Table 4, the coefficient of the variable PAT_US_{it}, namely the number of patents granted in the United States (0.055) is significant at 5% level. This provides evidence that patents filed in the United States also have a positive association with Tobin's q as predicted in our hypothesis H1b. Once again we find no evidence to indicate that level of advertising influenced firm performance as measured by Tobin's q. Leverage is once again significant at the 1% level. The results are consistent as before with firm size (SIZE) not being statistically significant and research and development expenditure (RD_{it}) being only marginally significant at the 10% level.

In the column (3) of Table 4, we include both PAT_TW_{it} and PAT_US_{it}. The coefficient of PAT_US_{it} (0.109) is significant at the 5% level while the coefficient of PAT_TW_{it} (0.096) is also significant at the 5% level. These results further corroborate H1a and H1b. The results for the other variables are similar as before. Furthermore, the results of the F-test comparing the coefficients of PAT_US_{it} and PAT_TW_{it} indicate that the coefficient of PAT_US_{it} is significantly greater than that of PAT_TW_{it}. This finding shows that patents granted in the United States have a greater influence on a firm's market valuation relative to patents granted in Taiwan. The patents granted to Taiwan semiconductor firms in the United States are seen as stronger evidence of innovative activity. Thus, consistent with our hypothesis H2 prediction, our empirical evidence lends credence to the theory that patents filed and granted in the United States surrogate for more intense innovative activity and is viewed more strongly by the market.

6.2. Tests of Hypotheses 3 and 4

We ran two more regressions as stated above to test our third and fourth hypotheses. A design firm is knowledge intensive and is considered to be more value-added than other types of semiconductor firms in the manufacturing and/or packaging and testing operation. The packaging and testing firms are labor intensive in nature and are less value-added than design and manufacture firms. The two regression equations are presented below:

$$Q_{it} = \alpha_0 + \alpha_1 \mathrm{PAT\_TW}_{it} + \alpha_2 \mathrm{AE}_{it} + \alpha_3 \mathrm{LEV}_{it} + \alpha_4 \mathrm{SIZE}_{it} + \alpha_5 \mathrm{RD}_{it} + \alpha_6 \mathrm{D}_1\mathrm{\_TW}_{it} + \alpha_7 \mathrm{D}_2\mathrm{\_TW}_{it} + \sum_{y=1990}^{2001} \alpha_{8y} \mathrm{YR}_y + \varepsilon_{it} \quad (4)$$

$$Q_{it} = \alpha_0 + \alpha_1 \mathrm{PAT\_US}_{it} + \alpha_2 \mathrm{AE}_{it} + \alpha_3 \mathrm{LEV}_{it} + \alpha_4 \mathrm{SIZE}_{it} + \alpha_5 \mathrm{RD}_{it} + \alpha_6 \mathrm{D}_1\mathrm{\_US}_{it} + \alpha_7 \mathrm{D}_2\mathrm{\_US}_{it} + \sum_{y=1990}^{2001} \alpha_{8y} \mathrm{YR}_y + \varepsilon_{it} \quad (5)$$

In Eq. 4, $\mathrm{D}_1\mathrm{\_TW}_{it}$ is a dummy variable that represents 1 if the firm is in the design sector and 0 if the firm is in other sectors. $\mathrm{D}_2\mathrm{\_TW}_{it}$ is a dummy variable that represents 1 if the firm is in the manufacturing sector; 0 if the firm is in other sectors. In Eq. 5, $\mathrm{D}_1\mathrm{\_US}_{it}$ is a dummy variable that represents 1 if the firm is in the design sector and 0 if the firm is in other sectors. $\mathrm{D}_2\mathrm{\_US}_{it}$ is a dummy variable that represents 1 if the firm is in the manufacturing sector; 0 if the firm is in other sectors. All other variables are the same as previously discussed. The results of Eq. 4 are shown in the first column of Table 5 and the results of Eq. 5 are shown in the second column.

In the first column of Table 5, the coefficient of the variable $\mathrm{D}_1\mathrm{\_TW}_{it}$ (1.357) is significant at the 1% level. This indicates that patents granted in Taiwan to Taiwanese firms in the design sector have much stronger association with Tobin's *q* relative to those granted to firms in the packaging & testing sector. Similarly, the coefficient of the variable $\mathrm{D}_2\mathrm{\_TW}_{it}$ (0.556) is also significant at the 1% level. This indicates that patents granted to Taiwanese firms in the manufacturing sector have greater association with Tobin's *q* relative to firms in the packaging & testing sector. However, the coefficient of the variable $\mathrm{D}_1\mathrm{\_TW}_{it}$ is larger than the coefficient of the variable $\mathrm{D}_2\mathrm{\_TW}_{it}$. The results of an *F*-test examining the difference between the coefficients of design and manufacturing indicate that the difference is significant (i.e. $\mathrm{D}_1\mathrm{\_TW}_{it} > \mathrm{D}_2\mathrm{\_TW}_{it}$). Thus, we conclude that the association between Tobin's *q* and $\mathrm{D}_1\mathrm{\_TW}_{it}$ is stronger relative to the association between Tobin's *q* and $\mathrm{D}_2\mathrm{\_TW}_{it}$. This finding lends support to H3, namely, that patents granted in the design sector have greater influence on Tobin's *q* relative to patents granted in other sectors.

The results for Eq. 5 shown in the second column of Table 5 indicate that the coefficient of $\mathrm{PAT\_US}_{it}$ is significant at the 5% level. This indicates a

Table 5. A Comparative Analysis of Patent Granted in Taiwan and US on Tobin's *q* with Industry Value Chain Stage Dummy.

Variables	Sign	Taiwan	US
α_0		2.167	1.134
		(2.00)**	(0.75)
PAT_TW_{it}	+	0.134	
		(1.57)*	
PAT_US_{it}	+		0.147
			(1.98)**
AE_{it}	+	13.928	55.519
		(0.90)	(1.01)
LEV_{it}	−	−2.043	−3.776
		(−3.20)***	(−3.39)***
SIZE_{it}	−	−0.128	−0.313
		(1.05)	(1.25)
RD_{it}	+	1.628	0.112
		(1.06)	(0.06)
$\text{D}_1\text{\_TW}_{it}$	+	1.357	
		(11.44)***	
$\text{D}_2\text{\_TW}_{it}$	+	0.556	
		(9.22)***	
$\text{D}_1\text{\_US}_{it}$	+		0.668
			(1.66)**
$\text{D}_2\text{\_US}_{it}$	+		0.228
			(1.43)*
$\text{D}_1\text{\_TW}_{it} > \text{D}_2\text{\_TW}_{it}$		$F = 59.83$***	
$\text{D}_1\text{\_US}_{it} > \text{D}_2\text{\_US}_{it}$			$F = 4.25$***
N		279	279
F-statistics		9.24	13.22
Adj. R^2		0.233	0.279

Notes: For parsimony in terms of presentation, we do not report the coefficient estimates and significance levels associated with the year dummy variables.
Statistical significance is based on two-tailed test:
*signifies statistical significance at 10% level.
**signifies statistical significance at 5% level.
***signifies statistical significance at 1% level.

positive and significant association between patents granted to Taiwanese firms by the U.S. patents office and Tobin's *q*. The conclusion is that patents granted in the U.S. convey a positive signal to investors. Similarly, the coefficient of leverage (−3.77), a control variable, is significant at the 1% level. This is once again in accordance with literature that the market perceives low leverage positively. The coefficient of the other control variables, namely, advertising, asset size, and R&D expenditure are not significant. Similar to $D_1\_TW_{it}$ and $D_2\_TW_{it}$ in the first column of Table 5, the coefficient of the dummy variable $D_1\_US_{it}$ (0.668) and $D_2\_US_{it}$ (0.228) are significant at the 5% and 10% level, respectively. This implies that, for Taiwanese firms in the design sector patents granted by the U.S. patents office have a significantly greater influence on Tobin's q relative to patents granted in the packaging & testing sector. The result of an F-test indicates that the coefficient of design sector ($D_1\_US_{it}$) is more statistically significant than that of manufacturing sector ($D_2\_US_{it}$) at the 1% level (i.e. $D_1\_US_{it} > D_2\_US_{it}$). Thus, we have sufficient evidence to support H4.

6.3. Sensitivity Tests

6.3.1. Additional Surrogate for Innovative Ability

In this study we used patent frequency as a surrogate for innovative ability. However, patent frequency does not take into account the value of the patent granted. Patent value rather than simple patent frequency could be construed to be an important indicator of innovative ability. We estimated patent values using the Cobb–Douglas production function as used by Seethamraju (2003) and estimated the value of each patent at different value chain stages. Based on the Seethamraju's model, the Cobb–Douglas production framework is transformed as follows[7]:

$$\log \mathrm{SALE}_t = \alpha_0 + \alpha_1 \log C_t + \alpha_2 \log L_t + \alpha_3 \log \mathrm{PAT}_t + \alpha_4 \log \mathrm{RD}_t + \varepsilon_t \quad (6)$$

where SALE_t is the output, measured as sale in year t; C_t the physical measure as fixed assets at end of year t; L the labor, measured as salaries of employees in year t; PAT the number of patents owned by a firm at the end of year t; RD the research and development expense in year t.

Patents represent only a subset of innovation, because a firm can choose other legal forms, such as copyright and trade secrets to its intellectual property from competitors (e.g. Tabak & Barr, 1998; Balkin, Markman, & Gomerz-Mejia, 2000). The design that combines patents and RD helps better capturing firm's innovation efforts, with RD measuring investments in innovation and number of patents indicating innovation outputs (e.g. Balkin et al., 2000).

The incremental sales attributable to new patents granted in year t can be calculated as follows:

$$\Delta \mathrm{SALE}_t = \alpha_3 \times \mathrm{PCHPAT}_t \times \mathrm{SALE}_t \quad (7)$$

where α_3 is the estimated coefficients on the log PAT variable from Eq. (6); ΔSALE_t the incremental sale attributable to patents in year t; PCHPAT_t the percentage change in patents in year t; SALE_t the actual sales in year t.

ΔSALE_t is a series of cash flows attributable to the patents, we assume that this cash flow goes to remaining legal year n in year t, and the cost of capital is 6%.[8] Estimated market value of all patents in year t:

$$\text{PATMV}_t = \Delta\text{SALE}_t \times \frac{\text{PATLEVEL}_t}{\Delta\text{PAT}_t} \times (1 + A_{\overline{n}|6\%}) \tag{8}$$

where PATLEVEL is the number of firm's patents as of the end of fiscal year t, ΔPAT_t the number of new patents granted by the firm in year t. $A_{\overline{n}|6\%}$ is the present value of annuity for n years at 6% discounted rate. We re-conduct our tests using this measure and our results are shown in Table 6.

Table 6 presents the comparison between the mean and median estimated value of patents granted in the United States, and in Taiwan at the three industry levels. As evidenced in panel A, the estimated value of patents is still significantly greater in the United States relative to Taiwan at the design stage, manufacture stage, and the packaging and testing stage and overall at the 5% level. Further, our findings in panel B show that within countries the estimated value of patents is higher for design and declines as we move down the value chain to manufacturing and packaging and testing sectors. The estimated value for packaging and testing is the lowest for both countries. These results further reinforce our hypotheses.

6.3.2. Test for Influence of "ADS"

Prior research indicates that trading in American Depository Shares (ADS) influences patent values discussed in Doidge et al. (2004). Thus one could attribute the results of our tests shown in Table 4 to the fact that firms traded in ADS. Hence we excluded firms which traded in ADS and reestimated our regressions. The results are not significantly different from that of Table 4.

6.3.3. Test for Panel Data Autocorrelation

In studies similar to ours which use panel data, autocorrelation could bias the results. We used a random effects model and a fixed effect model to test for the presence of serial correlation due to panel data as suggested by Greene (2003). The Hausman test is used to see if the significance of parameter estimates is consistent between two models. The statistical results, $X^2(5) = 0.79 (P = 0.8236)$, indicate that the deviation between two models is insignificant. Given the unbalanced panel in our sample, we conducted a

Table 6. A Comparative Analysis of Estimated Patent Value across Industry Value Chain Stages by Country.

Panel A: Differences in mean and median estimated values between US and Taiwan by industry

Industry Value Chain Stages	Mean Estimated Patent Value[a] in US (Median)	Mean Estimated Patent Value[a] in Taiwan (Median)	Difference in the Mean and Median of Estimated Patent Value between in US and in Taiwan	*t*-Values (*z*-Statistics)
Design (D)	\$13,147,624	\$9,437,291	D_{us}–D_{tw}	2.713***
	(\$16,284,501)	(\$13,529,101)		(2.437**)
Manufacture (M)	9,484,772	7,002,113	M_{us}–M_{tw}	2.324**
	(10,214,110)	(8,321,582)		(2.298**)
Package & test (PT)	3,847,992	2,213,113	PT_{us}–PT_{tw}	2.113**
	(4,138,510)	(2,868,294)		(2.289**)
Total	\$9,429,217	\$7,463,424	US_{patent}–TW_{patent}	2.204**
	(\$10,332,247)	(\$8,779,242)		(2.268**)

Panel B: Differences in mean and median estimated values between industries in US and Taiwan

Industry Value Chain Stages	Mean Estimated Patent Value[a] in US (Median)	Mean Estimated Patent Value[a] in Taiwan (Median)	Difference in the Mean and Median of Estimated Patent Value by Stages	US Patent *t*-Values (*z*-Statistics)	Taiwan Patent *t*-Values (*z*-Statistics)
Design	\$13,147,624	\$9,437,291	D-M	3.413***	2.992***
(D)	(\$16,284,501)	(\$13,529,101)		(4.975***)	(3.602***)
Manufacture	9,484,772	7,002,113	D-PT	9.749***	8.814***
(M)	(10,214,110)	(8,321,582)		(7.298***)	(7.010***)
Package & test	3,847,992	2,213,113	M-PT	5.362***	6.624***
(PT)	(4,138,510)	(2,868,294)		(4.731***)	(5.012***)

Statistical significance is based on two-tailed test:
*signifies statistical significance at 10% level.
**signifies statistical significance at 5% level.
***signifies statistical significance at 1% level.
[a]The estimated patent value is amounted by New Taiwan dollars and calculated according to section 6.3.1 above.

sensitivity analysis and repeated the tests using a random effects model. We found results similar to the prior analysis.

6.3.4. Additional Surrogates for Firm Performance

Firm performance is a very complex construct and a criticism of this study is that relying on one performance indicator could be problematic. Hence, we replicate our models in Eqs. (4) and (5) using stock price as a dependent variable. The results shown on Table 7 are consistent with our current findings using Tobin's q in Table 5.

Table 7. A Comparative Analysis of Patent Granted in Taiwan and US on Stock Price with Industry Value Chain Stage Dummy.

Variables	Sign	Taiwan	US
α_0		2.167	3.367
		(2.00)**	(0.15)
PAT_TW_{it}	+	0.696	
		(3.25)***	
PAT_US_{it}	+		1.868
			(1.59)*
AE_{it}	+	494.516	304.274
		(0.05)	(2.32)**
LEV_{it}	−	−3.799	−27.214
		(−0.52)	(−2.62)***
$SIZE_{it}$	−	−1.26	−3.365
		(0.71)	(−1.18)
RD_{it}	+	0.398	14.012
		(0.02)	(0.66)
$D_1\_TW_{it}$	+	8.419	
		(3.63)***	
$D_2\_TW_{it}$	+	0.545	
		(1.44)*	
$D_1\_US_{it}$	+		5.309
			(2.47)***
$D_2\_US_{it}$	+		0.266
			(1.51)*
$D_1\_TW_{it} > D_2\_TW_{it}$		$F=34.88$***	
$D_1\_US_{it} > D_2\_US_{it}$			$F=9.46$***
N		279	279
F-statistics		27.92	31.23
Adj. R^2		0.180	0.209

Notes: The dependent variable (P_{it}) is a closing price at the fiscal year end. For parsimony in terms of presentation, we do not report the coefficient estimates and significance levels associated with the year dummy variables.
Statistical significance is based on two-tailed test:
*signifies statistical significance at 10% level.
**signifies statistical significance at 5% level.
***signifies statistical significance at 1% level.

7. CONCLUSIONS

In this study we focus on the semiconductor industry because it is an important industry that is vibrant and technologically dynamic characterized by continuous innovation. We concentrate on the semiconductor industry in Taiwan. This is because the Taiwan semiconductor industry value chain is different with respect to vertical specialization relative to the rest of the world. In Taiwan firms specialize in one stage of the industry value chain, unlike the U.S. where all stages are attempted by the same company. Thus, it is possible to examine how the market incorporates added value as we move along the industry chain.

Our sample comprised 279 Taiwan semiconductor firm-years covering the period 1990 to 2002. The total number of patents granted in Taiwan was 10,152 and that in the United States was 6,095. We categorized our sample by the stage in the industry chain, namely, design, manufacture, and packaging and testing. Our conclusions are that extent of value added as measured by patenting activity in this study influences market's perception (as measured by Tobin's q) of added value. Overall, we find that patents filed and granted in the United States surrogate for more intense added value activity. The cross-national patenting is greatly valued by the market. Finally, we find evidence that position in the industry chain affects such value-relevance association. In particular, design stage value added activity has a stronger association with Tobin's q; this association declines as we move from upstream to the downstream end of the industry value chain. The design stage is characterized by relatively higher knowledge intensity; more closer interaction with end users; constant upgrading of latest technology to enhance design capabilities; and quick response to market changes. These characteristics decline as we move down the value chain. It appears these factors are considered by the market and accentuate the association of innovative capability and firm performance. This accentuation is magnified by location of the patent filing. It is greater for firms filing in the United States relative to those filing in Taiwan.

NOTES

1. Taiwan's patent filings per million capita in the United States rank second only to U.S. companies. Out of more than 67,000 patent applications, over 47,000 patents were granted in 2002. These figures represent an 8% increase in invention patent applications and registrations compared with year 2000. Meanwhile, Taiwan's outbound patent filings in the United States rank fourth (4,667 patents in 2000), only lower than the U.S. (85,072), Japan (31,296), and Germany (10,234) (see Eastwood & Shiue, (2002) for more details about intellectual property protection in Taiwan).

2. The patents granted to Taiwan semiconductor industry account for 42% out of total patents in Taiwan. In addition, 56% out of all Taiwan patents granted by United States Patent and Trademark Office (USPTO) was granted to Taiwan semiconductor firms (see Chin, Lin, & Chi, 2004).

3. These innovative efforts that enhance productivity and create the potential for competitive advantage in the semiconductor industry have been lauded and

supported by Morris Chang, the chairman of Taiwan Semiconductor Manufacturing Co. Ltd. (TSM), the world's largest dedicated semiconductor foundry, and he emphasizes that "It is true that new economy is driven by technology, but that is not all. A spirit of innovation and initiative is also fueling its progress" (see "Innovators are the biggest winners", published by Common Wealth Monthly, 7/19/2002).

4. According to World Semiconductor Trade Statistics (2003), United States is the biggest semiconductor market in the world and is also the primary market to Taiwan. Sixty-five percent of Taiwan semiconductor products are exported to North America market.

5. Because firm level patent citations data are currently unavailable from the resources of academic research in Taiwan, only the number of patent is used in the current study.

6. Please see http://www.nber.org/patents/ for the NBER U.S. patent data files.

7. Seethamraju (2003) used number of trademarks (TM) and advertising expense (ADV) rather than patents and research and development. Furthermore, he assumed that cash flows were set to infinity, not remaining legal years.

8. We also use costs of capital 8% and 10%, respectively, and the empirical results remain unchanged.

ACKNOWLEDGMENTS

We gratefully acknowledge associate editor and two anonymous reviewers for very insightful comments on earlier version of this chapter. All errors and mistakes are our responsibility.

REFERENCES

Aaker, D. (1991). *Managing brand equity: Capitalizing on the value of a brand name*. New York: Free Press.

Amir, E., & Lev, B. (1996). Value relevance of nonfinancial information: The wireless communications industry. *Journal of Accounting and Economics*, *22*, 3–30.

Balkin, D. B., Markman, G. D., & Gomez-Mejia, L. R. (2000). Is CEO pay in high-technology firms related to innovation? *Academy of Management Journal*, *43*, 1118–1129.

Bharadwaj, A. S., Bharadwaj, S. G., & Konsynski, B. R. (1999). Information technology effects on firm performance as measured by Tobin's q. *Management Science*, *45*, 1008–1024.

Bosworth, D. L., & Mahdian, H. (1999). Returns to intellectual property in the pharmaceuticals sector. *Economique Appliquee*, *52*, 69–93.

Brown, S., Lo, K., & Lys, T. (1999). Use of R^2 in accounting research: Measuring changes in value relevance over the last four decades. *Journal of Accounting and Economics*, *28*, 83–115.

Chang, C. J., & Hwang, N. C. R. (2002). The effects of country and industry on implementing value chain cost analyses. *International Journal of Accounting*, *37*, 123–141.

Chang, P. L., & Tsai, C. T. (2000). Evolution of technology development strategies for Taiwan's semiconductor industry: Formation of research consortia. *Industry and Innovation*, *7*, 185–197.

Chen, S. H. (2002). Global production networks and information technology: The case of Taiwan. *Industry and Innovation*, *9*, 249–265.

Chin, C. H., Lin, H. W., & Hong, J. S. (2003). Investigating the extent intangible assets help explain IPO underpricing and post-IPO insider holding changes. *International Journal of Accounting Studies*, *36*, 23–53.

Chin, C. L., Lee, P., Chi, Y. H., & Anandarajan, A. (2006). Patent citation, R&D spillover, and Tobin's Q: Evidence from the Taiwan semiconductor industry. *Review of Quantitative Finance and Accounting*, *26*, 67–84.

Chin, C. L., Lin, H. W., & Chi, H. Y. (2004). Value relevance of patent: A test of life cycle hypothesis. *Journal of Management*, *21*, 175–197Pressed in Taipei, Taiwan.

Cockburn, I., & Griliches, Z. (1988). Industry effects and appropriability measures in the stockmarket's valuation of R&D and patents. *American Economic Review*, *78*, 419–423.

Conner, K. (1991). Theory of the firm: Firm resources and other economic theories. *Journal of Management*, *17*, 121–154.

Connolly, R. A., & Hirschey, M. (1988). Market value and patents: A Bayesian approach. *Economics Letters*, *27*, 83–87.

Connolly, R. A., & Hirschey, M. (1990). Firm size and R&D effectiveness. *Economic Letters*, *32*, 277–281.

Dekker, H. C. (2003). Value chain analysis in interfirm relationships: A field study. *Management Accounting Research*, *14*, 1–23.

Doidge, C., Karolyi, G. A., & Stulz, R. M. (2004). Why are foreign firm, listed in U.S. worth more? *Journal of Financial Economics*, *71*, 205–238.

Eastwood, J., & Shiue, S. (2002). Intellectual property protection in Taiwan: The year in review. http://www.winklerpartners.com/htmlffiles-english/Publications/articles/2002IPProfiles Taiwan.PDF

Francis, J., & Schipper, K. (1999). Have financial statements lost their relevance? *Journal of Accounting Research*, *37*, 319–352.

Greene, J. P., Stark, A. W., & Thomas, H. M. (1996). UK evidence of the market valuation of research and development expenditures. *Journal of Business Finance and Accounting*, *23*, 191–216.

Greene, W. H. (2003). *Econometric analysis*. Upper Saddle River, NJ: Prentice Hall.

Griliches, Z. (1981). Market value, R&D and patents. *Economics letters*, *7*, 183–187.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, *28*, 1661–1707.

Griliches, Z., Hall, B., & Pakes, A. (1991). R&D, patents and market value revisited: Is there a second (technological opportunity factor). *Economics of Innovation and New Technology*, *1*, 1183–1201.

Griliches, Z., Pakes, A., & Hall, B. (1987). The value of patents as indicators of inventive activity. In: P. Dasgupta & P. Stoneman (Eds), *Economic policy and technological performance* (pp. 97–124). Cambridge: Cambridge University Press.

Grupp, F. H., & Schmoch, U. (1999). Patent statistics in the age of globalization: New legal procedures, new analytical methods, new economic interpretation. *Research Policy*, *28*, 377–396.

Hall, B. (1993). The stock market valuation of R&D investment during the 1980s. *American Economic Review*, *83*, 259–264.

Hall, B. (2000). Innovation and market value. In: R. Barrell, G. Mason & M. O'Mahoney (Eds), *Productivity, innovation and economic performance* (pp. 177–198). Cambridge: Cambridge University Press.

Hall, B. H., & Ziedonis, R. H. (2001). The patent paradox revisited: An empirical study of patenting in the U.S. semiconductor industry, 1979–1995. *RAND Journal of Economics*, *32*, 101–128.

Hirschey, M., Richardson, V. J., & Scholz, S. (2001). Value relevance of nonfinancial information: The case of patent data. *Review of Quantitative Finance and Accounting*, *17*, 223–235.

Hitt, M., Hoskisson, R. E., & Ireland, R. D. (1994). A midrange theory of the interactive effects on innovation and firm performance in product diversified firms. *Academy of Management Journal*, *40*, 767–798.

Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits and market value. *American Economic Review*, *76*, 984–1001.

Macher, J. T., Mowery, D. C., & Simcoe, T. S. (2002). E-Business and disintegration of the semiconductor industry value chain. *Industry and Innovation*, *9*, 155–181.

Mathews, J. A. (1997). A silicon valley of the east: Creating Taiwan's semiconductor industry. *California Management Review*, *39*, 26–54.

Megna, P., & Klock, M. (1993). The impact of intangible capital on Tobin's q in the semiconductor industry. *American Economic Review*, *83*, 265–269.

Megna, P., & Mueller, D. C. (1991). Profit rates and intangible capital. *Review of Economics and Statististics*, *73*, 632–642.

Ministry of Economic Affairs. (2001). *Report on information industry research*. http://mic.iii.org.tw/products/products4.asp Industrial Technology Intelligence Services, Taipei, Taiwan.

Ministry of Economic Affairs. (2003). *Semiconductor industry yearbook*. Taipei, Taiwan: Industrial Technology Intelligence Services.

Morck, R., Shleifer, A., & Vishny, R. W. (1988). Management ownership and market valuation: An empirical analysis. *Journal of Financial Economics*, *20*, 293–315.

Morck, R., & Yeung, B. (1991). Why investors value multinationality. *Journal of Business*, *64*, 165–187.

Park, C. W., Jaworski, B. J., & Macinnes, D. J. (1986). Strategic brand concept-image management. *Journal of Marketing*, *50*, 135–145.

Pegels, C. C., & Thirumurthy, M. V. (1996). The impact of technology strategy on firm performance. *IEEE Transactions on Engineering Management*, *43*, 246–249.

Seethamraju, C. (2003). The value-relevance of trademarks. *Intangible assets, values, measures and risks* (pp. 80–96), Oxford Management Readers.

Shane, H., & Klock, M. (1997). The relation between patent citations and Tobin's Q in the semiconductor industry. *Reveiew of Quantitative Finance and Accounting*, *9*, 131–146.

Shank, J. K., & Govindarajan, V. (1992). Strategic cost management and the value chain. *Journal of Cost Management*, *6*(3), 5–21.

Sher, P. J., & Yang, P. Y. (2005). The effects of innovative capabilities and R&D clustering on firm performance: The evidence of Taiwan's semiconductor industry. *Technovation*, *25*, 33–43.

Simon, C. J., & Sullivan, M. W. (1993). The measurement and determinants of brand equity: A financial approach. *Marketing Science*, *12*, 28–52.

Stoneman, P., & Toivanen, O. (1997). *Innovation and market value in the UK*. Coventry: Warwick Business School, University of Warwick.

Tabak, F., & Barr, S. H. (1998). Innovation attributes and category membership: Explaining intention to adopt technological innovations in strategic decision making contexts. *Journal of High Technology Management Research*, *9*, 17–33.

Tobin, J. (1978). Monetary policies and the economy: The transmission mechanism. *Southern Economic Journal*, *37*, 421–431.

Toivanen, O., Stoneman, P., & Bosworth, D. (2002). Innovation and the market value of UK firms, 1989–1995. *Oxford Bulletin of Economics and Statistics*, *64*, 39–61.

Tung, A. (2001). Taiwan's semiconductor industry: What the state did and did not. *Review of Development Economics*, *5*, 266–288.

Watanabe, C., Tsuji, Y. S., & Griffy-Brown, C. (2001). Patent statistics: Deciphering a 'real' versus a 'pseudo' proxy of innovation. *Technovation*, *21*, 783–790.

World Semiconductor Trade Statistics. (2003). Semiconductor market forecast. http://www.wsts.org/press.html

Wu, W., & Bjornson, B. (1996). Value of advertising by food manufactures as investment in intangible capital. *Agribusiness*, *12*, 147–156.

Zahra, S. A., Ireland, R. D., & Hitt, M. A. (2000). International expansion by new venture firms: International diversity, mode of market entry, technology learning and performance. *Academy of Management Journal*, *43*, 925–950.

AN EXAMINATION OF FACTORS ASSOCIATED WITH THE TYPE AND NUMBER OF INTERNAL CONTROL DOCUMENTATION FORMATS

James Bierstaker, Diane Janvrin and D. Jordan Lowe

ABSTRACT

Auditing standards require that auditors obtain and document their understanding of internal control on every engagement, but do not specify the type or number of documentation formats auditors should adopt. We investigate the association of selected factors on the type and number of formats chosen. Data were collected from 181 auditors representing Big 4, national, regional, and local firms. Results suggest that auditors are most likely to use narratives followed by questionnaires. Firm size, client IT complexity, and auditor IT expertise are associated with auditors' format choice. Furthermore, while auditors use multiple formats, they tend to emphasize one format more than others. These findings have implications for audit effectiveness, since prior research suggests that documentation format may impact audit judgment, and auditors who rely on a single format may overlook significant internal control deficiencies.

Advances in Accounting, Volume 23, 31–48

ISSN: 0882-6110/doi:10.1016/S0882-6110(07)23002-0

INTRODUCTION

International and U.S. auditing standards require that auditors obtain and document a "sufficient understanding of internal control" on every engagement as a basis to plan the audit and determine the nature, timing, and extent of audit testing (ISA 400 IFA, 2003; U.S. SAS No. 78 AICPA, 1995). In addition, Section 404 of the Sarbanes–Oxley Act of 2002 (SOX) mandates that all publicly traded companies include management's assessment of internal control effectiveness over financial reporting in their annual report and that independent auditors attest to management's evaluation of internal control (U.S. House of Representatives, Committee on Financial Services, 2002). Further, SOX has widespread influence, since all foreign registrants must comply with SOX or lose access to U.S. capital markets. As a result, auditing standard organizations throughout the world are currently considering what changes to internal control documentation practices are warranted (Bierstaker & Wright, 2004).

At present, there is no standard format requirement for auditors to follow when documenting client internal control procedures (PCAOB, 2004-A-11; AICPA, 2006, AU 319.61; ISA 400 IFA, 2003). Although prior research has examined the use of different documentation formats, these studies involved large firms and were generally conducted prior to the recent rapid growth in client information technology (IT) (Mock & Turner, 1981; Cushing & Loebbecke, 1986; Bierstaker, 1999a). Little, if any research has examined the documentation practices of smaller firms. Recently, this has become an important issue, since auditors and management are considering which, and how many formats they should use to document controls in response to Section 404 of SOX. Our results provide benchmarks to practitioners from both large and small firms as they evaluate their own internal control documentation practices. For example, as national firms acquire more large publicly traded companies (International Accounting Bulletin, 2005), they may consider adapting their internal control documentation processes if necessary.

The choice of documentation format is an important issue because each format structures the data-collection process in a different way, which in turn may affect auditors' internal control evaluations and control risk assessments, and ultimately audit quality (Carscallen, 1982; Mock & Willingham, 1983). The documentation format guides auditors as to the quantity and type of internal control information that auditors gather (Purvis, 1989), and may influence the way auditors encode, retrieve, and recognize internal control information (Plumlee, 1985; Frederick, 1991; Plumlee, Tuttle, & Moeckel, 1998; Bierstaker, 1999b).[1] Furthermore, internal control documentation format may influence how auditors evaluate internal control strengths and weaknesses (Bierstaker, 1999b, 2003). Recent research suggests that auditors from Big 4 firms who completed questionnaires identified more internal control weaknesses, and assessed control risk higher than those auditors who prepared narratives (Bierstaker & Thibodeau, 2006). These results may be explained by differences in auditors' problem representation development due to the internal control documentation format used. Thus, in summary, differences in documentation format may affect auditors' problem representations, which could then influence their subsequent judgments and decisions.

Internal control documentation is also an important issue with regard to clients' development of more complex IT financial reporting systems, as documentation format may facilitate how auditors understand and evaluate complex IT systems (AICPA, 2001). That is, the nature and complexity of the client's controls and IT may be associated with the form and extent of internal control documentation. For example, audit standards suggest that auditors consider using flowcharts, questionnaires, and/or decision tables when documenting their understanding of the internal controls of a complex client system in which a large volume of transactions are electronically initiated, recorded, processed, or reported (AICPA, 2006, AU 319.61). For an information system making limited or no use of IT or for which few transactions are processed, documentation in the form of a narrative may be sufficient. Generally, the more complex the client's internal control and the more extensive the procedures performed by the auditor, the more extensive the auditors' documentation should be (AICPA, 2006, AU 319.61). Similarly, auditors' IT expertise may also be associated with their choice of internal control documentation.

The purpose of this study is to investigate current internal control document format choice and its association with firm size, client IT complexity, and auditor IT expertise. A field-based questionnaire was used to collect post-SOX data from 181 auditors of public companies representing Big 4, national, regional, and local firms. The overall findings indicate that auditors use narratives most often, followed by questionnaires, with other formats used less frequently. The results further suggest that Big 4 firms favor narratives, flowcharts, matrices, and process mapping more often than other firms, whereas smaller firms (i.e., national, regional, and local) use questionnaires more extensively. Interestingly, results also indicate that in contrast to audit standards, client IT complexity does not appear to be associated with narrative and questionnaire usage. However, auditors are more likely to use flowcharts, matrices, and process mapping for high IT complexity clients. Additionally, auditors with greater IT expertise are more likely to use flowcharts than are novices. Finally, we found that while auditors generally use multiple formats, they most often document internal controls with narratives or questionnaires.

RESEARCH QUESTION DEVELOPMENT

Internal Control Documentation Format

Several internal control documentation formats exist (AICPA, 2006, 319.61). For example, *narratives* provide written descriptions of the auditor's understanding of the client's internal control structure. Narratives are the most flexible but least structured of the formats since key decisions, such as which controls to include and the extent of control coverage, are left to the discretion of the individual auditor (Purvis, 1989). In contrast, *questionnaires* consist of a series of yes or no questions designed to assess a client's control environment, accounting system, and control activities. *Flowcharts* provide a symbolic, diagrammatic representation of the client's documents and tend to focus on procedural and organizational controls related to the flow of documents. Therefore, they may enhance the recall of internal control strengths and weaknesses as compared to narratives (Mock & Willingham, 1983; Bierstaker, 1999a). Finally, auditors may use formats that have received sparse research attention, such as matrices and/or process maps. *Matrices* are organized with the control objectives across the top and control activities listed down the side. *Process maps* are hierarchical methods for displaying processes that illustrate how a product or transaction is processed. Given the differences in the way these formats organize internal control information, the format(s) used to document internal control information is (are) likely to impact auditors' decision outcomes (Boritz, 1985).

Factors Influencing Auditors' Choice of Formats

To date, most prior research concentrates on identifying what formats Big 4 auditors use *without* regard for the factors that are associated with these choices (Bierstaker, 1999b; Bierstaker & Wright, 2004). Few studies have explored factors that may be associated with auditors' choice of documentation format(s). Thus, in addition to identifying current documentation practices, our study examines the association of firm size, client IT complexity, and auditor IT expertise on the type and number of internal control documentation formats selected by auditors.

Firm Size

Audit research often uses firm size to proxy for firm resources (e.g., Palmrose, 1986; Gist & Davidson, 1999). Firm size varies greatly within the audit profession, from local one-office firms to international Big 4 firms (Brierley & Gwilliam, 2001). Prior research finds that audit firm IT usage is more extensive in Big 4 firms as compared to smaller firms (Manson, McCartney, & Sherer, 1997; Ho, Vera-Munoz, & Chow, 2002). Similarly, firm resources may influence auditors' choice of internal control documentation format. For example, auditors employed by large firms may have more training with, and access to a wider array of formats, or more sophisticated tools for documenting controls, than auditors with smaller firms. In addition, large firms may be more focused on controls since their clients may have more sophisticated control systems to rely on. Documentation format differences between small and large firms may have significant audit effectiveness implications since documentation format may influence auditors' judgments. Since prior research has not addressed this issue, we ask:

RQ1. Will firm size be associated with the frequency of relying on various control documentation formats?

Client IT Complexity

As noted earlier, internal control documentation is increasingly important as clients develop more complex IT financial reporting systems given that documentation format may facilitate how auditors understand and evaluate complex IT systems (AICPA, 2001). To adequately understand the client's internal control, the auditor may need to adopt a documentation method that is best suited to the nature of the client's internal control system. Therefore, audit standards recommend that certain formats, such as flowcharts and questionnaires, may be a better fit for documenting sophisticated electronic internal control systems while narratives may be sufficient for clients with simpler reporting systems (AICPA, 2006, AU 319.61). The second research question follows:

RQ2. Will client IT complexity be associated with the frequency of relying on various control documentation formats?

Auditor IT Expertise

Another factor that may be associated with auditors' choice of format is their IT expertise. The greater the IT expertise of an individual, the more likely he or she will adopt IT (Venkatesh, Morris, Davis, & Davis, 2003). Similarly, auditors with higher levels of IT expertise may be more likely to adopt newer methods of internal control documentation such as process mapping and matrices than auditors with less IT expertise. The third research question follows:

RQ3. Will auditor IT expertise be associated with the frequency of relying on various control documentation formats?

Number of Formats

In addition to identifying what formats auditors use, researchers are interested in whether auditors are more likely to use single or multiple formats. Recent competitive pressures may have led to changes in the internal control documentation process, such as using fewer formats to document controls, or relying on formats that are less time-consuming (Bierstaker & Wright, 2004; Gramling, 1999; Houston, 1999). For instance, Bierstaker and Wright (2004) find that the average number of formats used by auditors decreased from 2.38 in 1995 to 1.45 in 2000, and that 62 percent of respondents typically relied on a single format in 2000 as compared with 12 percent in 1995. They suggest that this change may have been driven primarily by concerns about audit *efficiency*.

However, since the passage of SOX, auditors may be more concerned about adequate internal control documentation and audit *effectiveness*, perhaps excessively so (Reason, 2005). Purvis (1989) suggests that if a single documentation format is used, the relative benefits of other formats may be lost. This, in turn, could jeopardize audit effectiveness if significant internal control weaknesses are not identified. For example, Purvis (1989) finds that auditors using questionnaires focus on controls related to the organization and flow of information to a greater extent than auditors using flowcharts or narratives. Moreover, auditors who use narratives are more likely to overlook internal control weaknesses than auditors who use questionnaires (Bierstaker & Thibodeau, 2006), perhaps because questionnaires provide a comprehensive approach to assist auditors in gathering information about internal control. Given these recent changes in the auditing environment, it is difficult to predict whether auditors are more apt to use single or multiple documentation formats. This leads to the fourth research question:

RQ4. Will auditors be more likely to use single or multiple control documentation formats?

METHOD

Participants

Participants included a total of 181 auditors representing Big 4, national, regional, and local firms. All participants were auditors of public companies. One author attended the American Institute of Certified Public Accountants (AICPA) National Advanced Accounting and Auditing Technical Symposium to obtain responses from 109 auditors. We also contacted local offices of each Big 4 firm and one national firm. From these offices, we obtained responses from 72 auditors.

As shown in Table 1, respondents' average age was 36.5 years. Individuals had 12.7 years of experience with those from local and regional firms having more experience.[2] Thirty-one percent of the respondents were employed by Big 4 firms, 17 percent by national firms, 15 percent by regional firms, and 37 percent by local firms.[3] The highest education level for a significant majority (82.8 percent) was a bachelor's degree. Over 86 percent of the respondents held CPA certificates. Seventy-one percent were males. Participants varied in IT expertise with 70.5 percent indicating intermediate IT expertise, 16.7 percent were novice IT expertise, and 12.8 percent stated that they were IT experts.[4]

Instrument Development and Validation

To obtain credible evidence on internal control documentation usage, we included a documentation format question on an extensive field-based questionnaire regarding the role of technology in the audit process.[5] To increase construct validity (Cook & Campbell, 1979), we conducted two rounds of pilot testing. First, four researchers with significant audit and systems knowledge examined the instrument. We then pilot tested the revised instrument with eight auditors from four firms (Big 4, national, regional, and local) who had an average of 4.5 years of experience. Based on pilot testing feedback, we added an example to ensure participants understood how to code multiple documentation format usage.[6]

Independent Variables

Respondents provided information about firm size, client IT complexity, and IT expertise. Firm size is measured as a categorical variable with

Table 1. Participant Demographics ($n = 181$).

	Frequencies	Mean % (Standard Deviation)
Years as an external auditor[a]		12.7
		(9.4)
Age[a]		36.5
		(9.9)
Firm size[a]		
Big 4	55	31.1
National	31	17.5
Regional	26	14.7
Local	65	36.7
Highest education level[a]		
Bachelor's degree	149	82.8
Master's degree	29	16.1
Coursework beyond master's degree	2	1.1
Certification[a,b]		
Certified internal auditor	1	
Certified public accountant	156	
Certified information systems auditor	0	
Certified management accountant	1	
Certified financial executive	8	
Certified financial planner	0	
Other certification	1	
Gender[a]		
Male	127	71.0
Female	52	29.0
IT expertise		
Novice	30	16.7
Intermediate	127	70.5
Expert	23	12.8

[a]One or more participants did not respond.
[b]Participants could list more than one certification.

four categories: Big 4, national, regional, and local. Client IT complexity is measured by asking respondents to indicate the extent they agree (from 1 = strongly disagree to 7 = strongly agree) with the following statement, "My clients generally use complex information technology." Client IT complexity responses were split at the median (4.0), resulting in two groups – *low and high client IT complexity*. We adopted Hackbarth, Grover, and Mun's (2003) IT expertise measurement approach as participants rated their own IT expertise using three categories: novice, intermediate, and expert.

Dependent Variables

Following prior research (Bierstaker & Wright, 2004), our documentation format question was designed to elicit a wide variety of formats. Participants were asked to estimate how frequently (0–100 percent) they use each of the following formats to document internal controls on a typical audit engagement: narrative, questionnaire, flowchart, matrix, and process mapping. There was also a blank in which respondents could add other formats (very few did so). Respondents were not limited to listing only one format. The directions included an example that read, ‘‘If you use narratives for each engagement and questionnaires for 40 percent of your engagements, you would code it as: 100 percent narrative, 40 percent questionnaire, 0 percent flowchart, 0 percent matrix etc.’’

RESULTS

As shown in Table 2, respondents estimated that they used narratives on 71.4 percent of audit engagements, compared to 51.0 percent for questionnaires, 11.8 percent for flowcharts, 11.6 percent for matrices, and 9.1 percent for process maps.

Firm Size

To examine whether an association between internal control documentation choice and firm size exists, we ran analysis of covariance (ANCOVA)

Table 2. Internal Control Documentation Format Usage.

Format	*n*	Mean %
Narrative[a]	181	71.4
Questionnaire[b]	177	51.0
Flowcharts[b]	178	11.8
Matrix[b]	175	11.6
Process map[b]	174	9.1
Other[b]	174	1.1

[a]Respondents were asked to estimate how frequently (0 to 100%) they use each format to document internal controls on a typical audit engagement.
[b]Some respondents did not answer this part of the question.

analyses with experience as the covariate and firm size as the independent variable.[7] When formats varied by firm size, we conducted planned contrasts[8] to determine whether the specific format varied between (1) Big 4 and non-Big 4 firms, (2) Big 4 and national firms, and (3) national and smaller firms (i.e., regional and local).[9]

Firm size was also found to be an important factor in documentation format usage (Table 3). Respondents from Big 4 firms are more likely to use narratives (86.7 vs. 67.1, 64.5, and 64.9 percent) ($F = 5.84$; $p = 0.00$), matrices (28.2 vs. 5.0, 4.5, and 4.9 percent) ($F = 12.06$; $p = 0.00$), and process mapping (24.3 vs. 5.2, 4.0, and 1.5 percent) ($F = 12.08$; $p = 0.00$) than respondents from national, regional, and local firms. Conversely, respondents from local (63.1 percent) and national firms (66.4 percent) are more likely to use questionnaires ($F = 10.42$; $p = 0.00$) than respondents from Big 4 firms (31.0 percent). Finally, auditors employed by national firms were more apt to use similar documentation formats as auditors from smaller firms, rather than those employed by Big 4 firms.

Client IT Complexity

As shown in Table 4, auditors that examine clients with high IT complexity are more likely to use flowcharts (15.2 vs. 6.2 percent; $t = 3.17$; $p = 0.01$), matrices (16.8 vs. 3.9 percent; $t = 3.85$; $p = 0.00$), and process mapping (14.0 vs. 1.8 percent; $t = 4.30$; $p = 0.00$) than auditors examining clients with low IT complexity. However, client IT complexity is not associated with internal control documentation format for narratives and questionnaires at statistically significant levels, despite the suggested guidance in the auditing standards.

Auditor IT Expertise

To examine whether an association between internal control documentation and auditor IT expertise exists, we ran ANCOVA analyses with experience as the covariate and auditor IT expertise as the independent variable. When formats varied by auditor IT expertise, we conducted planned contrasts to identify specific differences. Results indicate that respondents with IT expertise are more likely to use flowcharts (13.0 vs. 5.5 percent; $F = 2.96$; $p = 0.04$). No other statistically significant results were found for auditor IT expertise.

Table 3. Firm Size Results.

	Big 4 Mean % ($n=55$)	National Mean % ($n=31$)	Regional Mean % ($n=26$)	Local Mean % ($n=65$)	ANCOVA	Big 4 vs. Others Contrast	National vs. Big 4 Contrast	National vs. Smaller Firm Contrast
Narrative[a]	86.7	67.1	64.5	64.9	0.00**	**	**	N/S
Questionnaire	31.0	66.4	45.2	63.1	0.00**	**	**	N/S
Flowchart	19.7	10.6	3.8	8.5	0.00**	**	**	N/S
Matrices	28.2	5.0	4.5	4.9	0.00**	**	**	N/S
Process map	24.3	5.2	4.0	1.5	0.00**	**	**	N/S
50 percent[b]	1.9	1.6	1.3	1.4	0.00**	**	**	N/S

Note: N/S, Non-significant two-tailed p-value.

** Two-tailed p-value < 0.01.

[a]Respondents were asked to estimate how frequently (0–100%) they use each format to document internal controls on a typical audit engagement.

[b]Number of formats used on average on at least 50 percent or more of audit engagements.

Table 4. Client IT Complexity Results.

	High Client IT Complexity[a] Mean % ($n = 100$)	Low Client IT Complexity Mean % ($n = 67$)	t-Statistics	p-Value
Narrative[b]	73.3	68.9	0.91	0.37
Questionnaire	48.5	54.8	−1.08	0.28
Flowchart	15.2	6.2	3.17	0.01**[d]
Matrices	16.8	3.9	3.85	0.00**[d]
Process map	14.0	1.8	4.30	0.00**[d]
50 percent [c]	1.70	1.36	2.74	0.01*[d]

* Two-tailed p-value < 0.05.
** Two-tailed p-value < 0.01.
[a]Respondents were asked to indicate the extent to which they agreed with the statement "My clients generally use complex information technology" on a scale of 1 = strongly disagree to 7 = strongly agree. Client IT complexity responses were split at the median (4.0), resulting in two groups.
[b]Respondents were asked to estimate how frequently (0–100%) they use each format to document internal controls on a typical audit engagement.
[c]Number of formats used on average on at least 50 percent or more of audit engagements.
[d]Results of Cochran t-test due to unequal variances.

Number of Formats

Results indicate that a single format was used by 9.9 percent of auditors. Given that respondents were allowed to provide the percentage of audits that they use each format, we analyzed the data further to determine whether auditors using multiple formats were still more likely to rely heavily on one format as opposed to the other(s). We found that auditors used 1.57 formats on average on at least 50 percent or more of audit engagements, which is comparable to the 1.45 formats commonly used by auditors as reported in Bierstaker and Wright (2004). Since we found that firm size, client IT complexity, and auditor IT expertise are associated with documentation format type, we also examined whether these factors were associated with single format usage and number of formats used. Results indicate that Big 4 auditors are likely to use more formats than smaller firms. In addition, auditors examining clients with complex IT tend to use more formats than those auditing clients with simpler IT. No other statistically significant associations between individual factors and single format usage or number of formats used were found.[10]

CONCLUSION, IMPLICATIONS, AND FUTURE RESEARCH

This study investigated current internal control documentation format usage and the association of internal control documentation with firm size, client IT complexity, and auditor IT expertise. Our overall findings suggest that narratives are the most common format for documenting internal controls, followed by questionnaires. Auditors may prefer narratives because they are relatively easy to prepare and update compared to other formats, and as a result they enhance flexibility and audit efficiency (Bierstaker & Wright, 2004). This finding is of concern, since recent research (Bierstaker & Thibodeau, 2006) suggests that auditors identify more internal control weaknesses and assess control risk higher when using a questionnaire as compared to a narrative. In addition, Bierstaker (1999a) finds that flowcharts enhance the recall of internal control strengths and weaknesses as compared to narratives. Therefore, while reliance on narratives for internal control documentation may enhance audit efficiency, the use of other formats, such as questionnaires and flowcharts, may lead to greater audit effectiveness (Purvis, 1989). Furthermore, while auditors indicated that they use more than one format across a variety of audits, they still tend to primarily document internal controls using narratives or questionnaires as opposed to other formats. These results are indicative of an emphasis on audit efficiency, which is consistent with other recent research (Bierstaker & Wright, 2004).

Our findings have important implications for both academia and practice. Although internal control documentation format can impact both audit effectiveness and efficiency, we are not aware of any other studies examining documentation format used by smaller firms. We found that smaller firms are more likely to use questionnaires, while Big 4 firms are more likely to use narratives. Since prior research indicates that auditors may acquire less internal control information (Purvis, 1989), and identify fewer internal control deficiencies (Bierstaker & Wright, 2004) when they use narratives as opposed to questionnaires, our findings raise concerns about the internal control documentation practices of larger audit firms that may ultimately have implications for audit quality.

Current audit standards suggest that client IT complexity influences internal control documentation format choice (AICPA, 2006, AU 319.61). Specifically, auditors are advised to use flowcharts and questionnaires to document internal controls of clients with complex IT, but use narratives for clients with simpler IT. Interestingly, we found that client IT complexity is not associated with auditors' choice to document internal controls with narratives or questionnaires. However, our findings do suggest that auditors use flowcharts, matrices, and process mapping more often on high IT complexity clients.

With regard to practice, auditors and management are currently debating which, and how many formats they should use to document controls in response to Section 404 of SOX. Recently, chief financial officers have criticized auditors for excessive levels of internal control documentation in response to SOX (Reason, 2005). This study provides descriptive information regarding several factors that may be associated with format choice, as well as formats that are commonly used. Given that our study collected data from auditors representing several firms of different sizes, our results provide benchmarks to practitioners from both large and small firms as they evaluate their own internal control documentation practices.

Finally, this study has some methodological and scope limitations that warrant consideration in interpreting the findings and/or offer promising opportunities for future consideration. First, we acknowledge certain construct limitations. Due to data limitations, we asked auditors to self-report the extent of their internal control documentation usage. To increase construct validity, a better proxy may be actual documentation usage (Straub, Limayem, & Karahanna, 1995; Devaraj & Kohli, 2003; Venkatesh et al., 2003). Similarly, to obtain data from a wide variety of firms, we collected data at the individual level, consistent with prior research (Bierstaker & Wright, 2004). Future research using firm-level data is needed to confirm our findings. Finally, internal control documentation choice may depend upon prior year choice and/or firm policy. Additional research examining the impact of prior year documentation choice and firm policy would provide insight into this influence.

Given the lack of prior research on the issues examined, we view the current study as exploratory in nature (Peecher & Solomon, 2001). Since little is known about auditors' choice of internal control documentation formats in the post-SOX environment, the descriptive data reported in this paper should motivate future research in a variety of ways. For instance, future research is thus needed to corroborate our findings by examining archival data from auditors' working papers. These data would be useful for corroborating the extent to which each documentation format is used for each audit client. In addition, future research could examine documentation formats used by internal auditors, and the extent to which internal control documentation is shared between external and internal auditors. Future research could also determine why firm size appears to be a major factor in determining document format. For instance, are firm size differences primarily due to the corresponding size of clients and related complexity? Alternatively, are differences based more on the increased emphasis that Big 4 firms place on internal controls? Finally, as auditors begin to issue opinions on internal control effectiveness as mandated by Section 404 of SOX, researchers can build upon our descriptive results to explore the relationship between internal control documentation formats, the judgments of managers and auditors on internal control effectiveness, and specific material weaknesses identified.

NOTES

1. Moreover, Agoglia, Kida, and Hanno (2003) find that the structure of justification memos in audit work papers influences preparers' and reviewers' conclusions about the strength of the control environment.
2. The demographics of our respondents are similar to the demographics of participants in Vendrzyk and Bagranoff (2003). Since experience varied by firm size, we include experience as a covariate when analyzing the impact of firm size on internal control documentation format. In spite of these experience differences, we find that our results are driven primarily by firm size differences.
3. Three participants did not provide their firm size designation.
4. We describe how we measure IT expertise in the Independent Variables section.
5. Given that one co-author has developed an extensive research program examining work paper review methods and the impact of documentation format on audit judgment, one goal of our field-based questionnaire was to obtain up-to-date information about internal control documentation formats from both large and small firms.
6. We describe the example verbiage in the Dependent Variables section.
7. We ran the ANCOVA tests using weighted least squares regression to account for differences in firm size sample cells.
8. We also ran Krushal-Wallis non-parametric test since the Brown Forsythe *F*-test of homogeneous variances indicated that our data contained non-homogeneous variances. In addition, we used bootstrap methods for narratives and flowcharts since the experience co-variate was significant. Results of both Krushal-Wallis and bootstrap methods are qualitatively similar to the reported results.
9. Before grouping responses from regional and local firms together, we ran an initial planned contrast to identify any differences between these responses. Results indicated that only questionnaire usage differed between regional and local firms.
10. Generalized weighted least squares regression models (Format Usage = f (firm size, client IT complexity, and auditor IT expertise)) were also run to determine if results hold when the factors associated with documentation format usage are examined simultaneously. Results suggest that the effect of firm size is stronger than client IT complexity for flowcharts, matrices, and process mapping. In addition, the firm size effect is stronger than auditor IT expertise for flowcharts.

ACKNOWLEDGMENTS

We appreciate the helpful comments of Joseph Brazel, Julie Smith David, Julie Anne Dilley, Gary Hackbarth, Frank Hodge, Cynthia Jeffrey, Eric Johnson, Brian Mennecke, Ed O'Donnell, Kurt Pany, G. Premkumar, Jay Rich, and Tony Townsend. Finally, we gratefully acknowledge the assistance of Will Baumann, Sarah Erdmann, Krissy Gronborg, Allan Trapp, Omar Torrens, Pat Wagaman, Katie Wallace, the AICPA, and the study participants.

REFERENCES

Agoglia, C. P., Kida, T., & Hanno, D. (2003). The effects of alternative justification memos on the judgments of audit reviewees and reviewers. *Journal of Accounting Research*, *41*(March), 33–47.

AICPA – American Institute of Certified Public Accountants. (1995). *Consideration of the internal control structure in a financial statement audit, statement of auditing standards no. 78*. New York: AICPA.

AICPA – American Institute of Certified Public Accountants. (2001). *The effect of information technology on the auditor's consideration of internal control in a financial statement audit, statement of auditing standards no. 94*. New York: AICPA.

AICPA – American Institute of Certified Public Accountants. (2006). *Codification of auditing standards*. New York: AICPA.

Bierstaker, J. (1999a). Internal control documentation: Which format is preferred? *The Auditor's Report* (Winter), 12–13.

Bierstaker, J. (1999b). A test of the split attention effect in a professional context. *Journal of Business and Behavioral Sciences* (Fall), 177–189.

Bierstaker, J. (2003). Auditor recall and evaluation of internal control information: Does task-specific knowledge mitigate part-list interference. *Managerial Auditing Journal*, *18*(2), 90–99.

Bierstaker, J., & Thibodeau, J. (2006). The effect of documentation format and task-specific experience on auditor internal control evaluation. *Managerial Auditing Journal*, *21*(9), 877–891.

Bierstaker, J., & Wright, A. (2004). The impact of the adoption of a business risk audit approach on internal control documentation and testing practices: A longitudinal investigation. *International Auditing Journal*, *8*(March), 67–78.

Boritz, E. F. (1985). The effect of information presentation structures on audit planning and review judgments. *Contemporary Accounting Research*, *2*, 193–218.

Brierley, J. A., & Gwilliam, D. R. (2001). *Human resource management issues in accounting and audit firms: A research perspective*. Aldershot, England: Ashgate.

Carscallen, M. P. (1982). Audit efficiency. *CPA Journal* (May), 20–22.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation design & analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.

Cushing, B. E., & Loebbecke, J. K. (1986). *Comparison of audit methodologies of large accounting firms*. Sarasota, FL: American Accounting Association.

Devaraj, S., & Kohli, R. (2003). Performance impacts of information technology: Is actual usage the missing link? *Management Science*, *49*(March), 273–289.

Frederick, D. M. (1991). Auditors' representation and retrieval of internal control knowledge. *The Accounting Review*, *66*(April), 240–258.

Gist, W. E., & Davidson, R. A. (1999). An exploratory study of the influence of client factors on audit time budget variances. *Auditing: A Journal of Practice and Theory*, *18*(Spring), 101–116.

Gramling, A. A. (1999). External auditors' reliance on work performed by internal auditors: The influence of fee pressure on this reliance decision. *Auditing: A Journal of Practice and Theory*, *18*(Supplement), 117–136.

Hackbarth, G., Grover, V., & Mun, Y. (2003). Computer playfulness and anxiety: Positive and negative mediators of the system experience effect on perceived ease of use. *Information & Management*, *40*(January), 221–232.

Ho, J., Vera-Munoz, S., & Chow, C. (2002). *Towards a framework for understanding the antecedents of knowledge sharing in large CPA firm audits*. Working paper. University of California-Irvine.

Houston, R. W. (1999). The effect of fee pressure and client risk on audit seniors' time budget decisions. *Auditing: A Journal of Practice and Theory*, *18*(Fall), 70–86.

International Accounting Bulletin. (2005). Interview-BDO Seidman: Grabbing new opportunities. *International Accounting Bulletin*, September 8, p. 5, London.

International Federation of Accountants. (2003). *Handbook of international auditing, assurance and ethics pronouncements*. New York, NY: IFA.

Manson, S., McCartney, S., & Sherer, M. (1997). *Audit automation: The use of information technology in the planning, controlling and recording of audit work*. Edinburgh: Institute of Chartered Accountants of Scotland.

Mock, T., & Turner, J. L. (1981). *Internal accounting control evaluation and auditor judgment.* Audit Research Monograph, No. 3. New York: AICPA.

Mock, T., & Willingham, W. (1983). An improved method of documenting and evaluating a system of internal accounting controls. *Auditing: A Journal of Practice and Theory*, *2*(Spring), 91–99.

Palmrose, Z. (1986). Audit fees and auditor size: Further evidence. *Journal of Accounting Research*, *24*(Spring), 97–111.

Peecher, M. E., & Solomon, I. (2001). Theory and experimentation in studies of audit judgments and decisions: Avoiding common research traps. *International Journal of Auditing*, *5*(November), 193–203.

Plumlee, R. D. (1985). The standard of objectivity for internal auditors: Memory and bias effects. *Journal of Accounting Research*, *23*, 683–699.

Plumlee, R. D., Tuttle, B., & Moeckel, C. L. (1998). *Auditors' evidence documentation and judgment: The mediating role of memory*. Working paper. University of Utah.

Public Company Accounting Oversight Board (PCAOB). (2004). An audit of internal control over financial reporting performed in conjunction with an audit of financial statements, Release No. 2004-001, March 9.

Purvis, S. E. (1989). The effect of audit documentation format on data collection. *Accounting, Organizations and Society*, *14*(5–6), 551–563.

Reason, T. (2005). Is Sarbanes-Oxley worth the cost? *CFO Magazine* (May), 1–8.

Straub, D., Limayem, M., & Karahanna, E. (1995). Measuring system usage: Implications for IS theory testing. *Management Science*, *41*(August), 1328–1342.
U.S. House of Representatives, Committee on Financial Services. (2002). Sarbanes–Oxley Act of 2002 (SOX). Public Law No. 107-204. Washington, DC, Government Printing Office.
Vendrzyk, V. P., & Bagranoff, N. A. (2003). The evolving role of IS audit: A field study comparing the perceptions of IS and financial auditors. *Advances in Accounting*, *20*, 141–163.
Venkatesh, V., Morris, M., Davis, G., & Davis, F. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(September), 425–478.

RE-DEFINING "MATERIALITY": AN EXERCISE TO RESTORE ETHICAL FINANCIAL REPORTING

Govind Iyer and Stacey Whitecotton

ABSTRACT

This paper uses a survey and experimental case methodology to examine whether financial statement users/stakeholders agree with the underlying assumptions of SAB 99, a staff bulletin issued by the Securities and Exchange Commission (SEC) in response to numerous high-profile cases of abusive earnings management practices by major corporations. SAB 99 lists several qualitative factors that should render an otherwise quantitatively immaterial item to be material in nature. Results of the survey and the experimental case studies confirm that both current and prospective members of management (as well as current audit seniors) are in general agreement with the materiality guidelines put forth in SAB 99.

INTRODUCTION

The concept of materiality is of paramount importance to the integrity of the financial reporting process under both generally accepted accounting principles (GAAP) and federal security laws. In presenting financial statements to shareholders, management is required to assert that the

Advances in Accounting, Volume 23, 49–83

ISSN: 0882-6110/doi:10.1016/S0882-6110(07)23003-2

financial statements are free of errors in all "material" respects, and auditors in providing unqualified reports are required to attest similarly. The typical unqualified report states that

> In our opinion, such consolidated financial statements present fairly, in all material respects, the financial position of the Company at June 30, 2005 and 2004, and the results of its operations and cash flows for each of the three years in the period ended June 30, 2005, in conformity with accounting principles generally accepted in the United States of America.[1]

Terms such as "fairly" and "material" require the exercise of professional judgment and interpretation given surrounding circumstances. Experience is clear that simple rules and definitions do not adequately protect the public under all conditions. It was such experiences that led federal legislators to mandate the engagement in the financial reporting process of "audit professionals" who by definition must bring to bear "professional judgment" in the performance of their "professional" role in society. While the term materiality is commonplace in accounting and financial reporting, there accordingly is no single definition of materiality. The meaning and interpretation of materiality nonetheless has been the subject of considerable research, professional commentary, authoritative pronouncements and case law including two Supreme Court decisions.[2] In Statement of Financial Accounting Concepts No. 2, the U.S. Financial Accounting Standards Board (1980) reflecting on past commentary defined the essence of the concept of materiality as follows:

> The omission or misstatement of an item in a financial report is material if, in the light of surrounding circumstances, the magnitude of the item is such that it is probable that the judgment of a reasonable person relying upon the report would have been changed or influenced by the inclusion or correction of the item … (however) magnitude by itself, without regard to the nature of the item and the circumstances in which the judgment has to be made, will not generally be a sufficient basis for a materiality judgment.

The focus of the literature and authoritative pronouncements has consistently emphasized the profession's fiduciary responsibility to society and user/stakeholders.[3] This is the very reason for the user-orientation found in definitions of materiality worldwide.[4] Given this focus, it is not surprising the legislative and regulatory reforms of the accounting profession have typically followed failures of the profession and the marketplace to adequately protect the interests of the public. Indeed, the most recent authoritative guidance on materiality (SEC Staff Accounting Bulletin SAB 99, 1999) was issued following numerous high visibility speeches and articles by then SEC Chairman Arthur Levitt on the topic of increasingly frequent and abusive earnings management practices by major corporations. With rare clairvoyance, regulators foresaw the pending meltdown occasioned by such colossal failures as Enron, World.com, Adelphi, Parmalat, Arthur Andersen and others. In denouncing several increasingly common abuses, Chairman Levitt reserved his harshest criticism for the widespread use of the strictly quantitative interpretations of materiality (Levitt, 1998, 2000). Strictly quantitative interpretations of materiality represented in the mind of the Chairman, the growing abrogation of professional responsibilities and business ethics by corporate management and the public accounting profession. The increasingly common application of strictly quantitative interpretations of materiality epitomized for the regulatory watchdog the increasingly common refusal of the profession to fully embrace their legislated fiduciary responsibilities to protect the public by application "professional judgment" in interpretation of the concepts of "fairness" and "material."

Chairman Levitt accused the management of numerous high-profile corporations of intentionally misleading the public by shielding known errors in financial statements from auditors' actions by staying within strictly applied quantitative materiality rules-of-thumb. These rules-of-thumb had evolved in practice in part through the influence of corporate management itself. Thus, it was increasingly common that financial statements contained known errors, arguably immaterial quantitatively, but large enough to allow the firm to meet targets important to management such as meeting financial analysts' consensus earnings estimates, maintenance of earnings trend, or qualification for management incentive compensation bonuses. According to Chairman Levitt, if meeting or not meeting these targets was important to management it was also important to the investing public. Further, it was alleged this lack of corporate integrity in financial reporting to the public (aided and abetted by auditors) was destroying the investor community's faith in the U.S. capital markets system. During Chairman Levitt's tenure, the SEC conducted vigorous investigations of many public corporations and issued a record number of enforcement actions on wrongdoers. Between July 31, 1997 and July 30, 2002, the SEC filed 515 enforcement actions for financial reporting and disclosure violations arising out of 227 Division of Enforcement investigations (SEC, 2003). It included 164 companies and 705 individuals. Even more striking was that of the 227 enforcement matters, 157 (70%) involved legal actions against at least one senior manager. Charges were brought against 75 chairmen of the board, 111 chief executive officers, 111 presidents, 105 chief financial officers, 21 chief operating officers, 16 chief accounting officers and 27 vice presidents of finance. In addition, the Commission also brought charges against 18 auditing firms and 89 individual auditors.[5] These figures provide staggering evidence and overwhelming testimony to the erosion of corporate integrity in financial reporting and auditing (even before the Enron and subsequent financial disasters of 2000 and thereafter). The evidence of diminished standards of conduct convinced the SEC of the need to take corrective actions even before Sarbanes–Oxley reforms. Those actions included the promulgation of SAB 99.

Staff Accounting Bulletin (SAB) 99 provided operational interpretative guidance by the staff of SEC regarding application of materiality standards. SAB 99 emphasized that exclusive reliance on quantitative thresholds in evaluating materiality had no basis in accounting literature, law or regulation; and corporate management and independent auditors (CPAs) must consider both ‘‘quantitative’’ and ‘‘qualitative’’ factors in assessing an item's materiality. The objective of SAB 99 was not to provide hard and fast rules that individuals might technically circumvent, but rather to place once again the burden of ethical conduct on management and CPAs with prescriptions to exercise professional judgment consistent with ‘‘fair’’ representations of corporate performance. It was deemed absolutely imperative that a reduction occur in the gap between standards of reporting expected by user/stakeholders and what was being provided. Six years later, with all the intervening distractions, there remains the question of whether the spirit and substance of SAB 99 has been embraced.

BACKGROUND

Materiality and its Role in Financial Reporting

The concept of materiality stems from the concern that in the absence of subjective judgment by auditors and managers, shareholders and investors may be buried in an avalanche of trivial disclosures that is not conducive to informed decision making. That is, shareholders and investors need to concern themselves only with ‘‘material’’ matters. In one of the first definitions of materiality, the Supreme Court stated that ‘‘[T]he question of materiality, it is universally agreed, is an objective one, involving the significance of an omitted or misrepresented fact to a reasonable investor.’’ Financial Accounting Concept No. 2 (quoted above) contains a similar definition of materiality. Consideration of materiality permeates every stage of the financial reporting and auditing process. In the planning stage, an auditor makes a preliminary estimate of the amount of error or irregularity that users would consider material to the overall financial statement. This directly impacts the amount of audit evidence gathered and the scope of tests and inquiries. Errors by definition are unintentional whereas irregularities are defined as intentional. Auditors are also required in the planning stage of an audit to assess the integrity of management and adjust the extent and scope of tests accordingly. In the reporting stage of an audit, the auditor is required to aggregate and evaluate uncovered errors and irregularities and correct those that are material. The auditor must decide whether the errors or irregularities, if not disclosed, are large enough or subjectively important enough, either individually or collectively, to alter the decision of a reasonable financial statement user, that is, place them in harm's way. If the auditor concludes that the effect of the error(s) is not likely to do harm, the errors are considered "not material" and an unqualified opinion is issued on the unadjusted financial statement.

As envisaged by the Court, the SEC and the Financial Accounting Standards Board (FASB), materiality analysis is *focused on the investor* not management, legal counsel, the corporate accountant or the auditor. The concept is purposefully vague; and as such carries legal risk for the parties rendering judgments. In the absence of specific standards, accountants and auditors informally created their own standard to help reduce their legal risks and increase consistency of interpretation. As a result, a rule-of-thumb, commonly known as the "5% rule" emerged. The logic behind the 5% rule was that *reasonable* investors should not be influenced in their investment decisions by a fluctuation in net income of 5% or less. Nor should the investor be swayed by a fluctuation or series of fluctuations of less than 5% in income statement line items as long as the net change was less than 5%. While the 5% rule was not formally codified as a standard, it was a widely accepted working rule. The standard was cited by the court in the Northway case ruling; although not acknowledged as definitive. The accounting profession has applied the rule-of-thumb increasingly as a hard and fast rule. Often cited as support is the Escott v. Bar Chris Construction Corporation case, in which Judge McClean argued as follows: "Accountants should not be held to a higher standard than that recognized in their profession."[6] Thus, legal protection was available according to Judge McClean, if one professional acts as other professionals act. Auditors sought legal protection by adhering to a rule-of-thumb with the hopes that it would be interpreted as an authoritative rule.

Chairman Levitt and the SEC however held that the 5% "rule-of-thumb" was never an authoritative rule, nor an acceptable standard for responsible action; and it did not protect the public, nor should it protect the auditor or management in court. SAB 99 asserted that (a) even if an item is quantitatively immaterial, it may still be material if one of more qualitative factors apply, and (b) intent can be viewed as constituting evidence of materiality. Thus, matters casting significant doubt on the integrity of management were in and of themselves material irrespective of the quantitative effect on current financial performance reports. SAB 99 lists a number of qualitative factors that may render a quantitatively immaterial fact material.

Item 1 in the SEC list addresses the nature of the misstatement or omission and the intent of management. Without doubt, a burden of assessing the integrity of management's actions was being hoisted on the auditor for all to see. Different standards were being attributed to errors versus irregularities. This was consistent with the mantle of "public watchdog" that Supreme Court Chief Justice Warren Burger previously had bestowed on the auditing profession. In United States v. Arthur Young and Co. Chief Justice Burger observed, "By certifying the public reports that collectively depict a corporation's financial status, the independent auditor assumes a public responsibility transcending any employment relationship with the client. The independent public accountant [465 U.S. 805, 818] performing this special function owes ultimate allegiance to the corporation's creditors and stockholders, as well as to the investing public. This "public watchdog" function demands that the accountant maintain total independence from the client at all times and requires complete fidelity to the public trust."[7] There is evidence that Chief Justice Warren's impression on the auditor as a watchdog is also shared by other members of the public. For instance, Reckers, Lowe, Jennings, and Pany (2005) report that judges, law students, MBA students and auditors alike agree that the role of the auditor is primarily that of a watchdog.[8] Other, but not all items in the list, relate to whether the misstatement or omission would "result in a significant stock market reaction;" (see item 8). Item 8 would seem to subsume item 3, and a number of other items would seem to be subsumed under item 3: whether an item would hide a failure to meet analysts' consensus estimates. Among those subsumed under meeting consensus analysts' predictions, one could argue are items 2, 4 and 5.

Item 6 speaks of the issue of whether a misstatement or omission would have the effect of qualifying management inappropriately for bonus incentive compensation. Item 6 would not seem to be a sub-element of

Item 3. Bonuses could neither be expected to directly and significantly affect the income of a public traded corporation, in and of themselves, nor to affect the market price of the company's stock, except potentially through the public's interest in the integrity of management. The intent and integrity of management appears to be the focus of item 6. Similarly, the focus of item 7 appears to reside with management integrity. In fact, item 7 harkens back to congressional prescriptions related to the Foreign Corrupt Practices Act of 1977. Thus, the items included in the SEC lists arguably relate to one or two issues: stock market consequences on user/stakeholders to whom fiduciary responsibility is owed or intent and integrity of management or both. All have clear ties to the ethics of corporate management, and the public accounting profession.

By introducing the intent and integrity of management as important factors in evaluation of materiality, SAB 99 requires auditors to perform *an additional one or potentially two-step analysis* for quantitatively immaterial amounts. First, the auditor must assess the intent of the management. If the misstatement or omission is deemed intentional, the auditor is pressed to require adjustment. It remains unclear whether it is necessary that a reasonable investor consider such intent highly indicative of low management integrity exercised elsewhere. The first question concerns mens rea or the state of the mind of the management. In this case an auditor is expected to be judge or jury like and evaluate management's state of mind based on collected evidence. The second question concerns what value investors attach to integrity and ethics of management in concluding that some "intents" are material while others are not.[9]

Prior Research on Materiality

Accounting literature is replete with earnings management and materiality studies. The former focuses on the relationship between economic incentives and incidents of earnings management; and the latter focuses on the imputation of thresholds for materiality assessments. Extant research on materiality also can be categorized into three groups: (a) experimental and archival examination of whether and what quantitative materiality thresholds are used by auditors, (b) experimental and archival examination of the extent of auditor reliance on subjective evidence in determination of materiality and (c) experimental and archival evaluation of whether materiality standards applied to income statement and balance sheet information differs from standards applied to footnote information. Evaluations of quantitative materiality thresholds generally consist of investigating whether ''rules-of-thumb'' are consistently applied in practice. The general conclusion of this stream of research has been that the size of the misstatement relative to net income (and in many cases gross or net assets or revenue) was an important determinant of whether an auditor assessed a misstatement to be material or not. For example, Holstrum and Messier (1982) indicate that (a) errors with a magnitude of more than 10% of income consistently are considered material and (2) errors with a magnitude of less than 5% *most frequently* are considered immaterial. Carmichael and Willingham (1989) reported that *for the most part* auditors use quantitative baselines as guidelines for determining whether errors are material or immaterial (see also Holder, 1983; Blocher & Cooper, 1988; Kinney & McDaniel, 1989; DeFond & Jimbalvo, 1991). This research arguably contributed to a circularity of events where research findings drove practice increasingly toward exclusive reliance on quantitative rules-of-thumb which succeeding rounds of research confirmed and reinforced. This line of research with a concentration on what is done *most frequently* or *for the most part* ignores the necessary exceptions and is easily misinterpreted as support for not making exceptions. Who can argue against consistency? Well, potentially user/stakeholders who rely on those who have fiduciary responsibility to them to do them no harm and to make fair representations of financial performance to them (and not hide behind rules-of-thumb subject to manipulation).

Research evaluating the effect of subjective factors on materiality decisions have tended to be limited to a very few factors. One subjective factor examined frequently is the nature of the alleged misstatement, e.g., violation of a technically specific generally accepted accounting standard, violation of a subjectively interpretable generally accepted accounting standards or an estimate of a revenue, expense, gain, loss or contingency. Auditors reportedly are less likely to require adjustments to management's financial statements when amounts are estimates and/or when judgments are required regarding the likelihood of a future event. Unfortunately, these are the very areas to which unscrupulous members of management resort to manipulate performance reports; and these are the items high on Chairman Levitt's chronicle of abuses. Examples of such adjustments include premature revenue recognition and manipulation of loss or expense reserves estimates. Another factor found to influence auditors' judgments was whether the contested item had an income increasing or income decreasing effect (Wright & Wright, 1997; Braun, 2001). Other findings from prior research suggest some auditors take into consideration information such as consistency with industry trends, management's cooperativeness (perceived indicative of intent/integrity), and risk of potential litigation in forming materiality assessments (Frishkoff, 1970; Firth, 1980; Chewning, Pany, & Wheeler, 1989). Finally, research on evaluation of the "format" of disclosure concluded that auditors evaluated materiality consistently irrespective of whether the information was recognized in the income statement versus disclosed in a footnote irrespective of "equal prominence" judicial standards (Maines & McDaniel, 2000).

Background: Ethics and Materiality

A limited number of studies have addressed the assessment of materiality from an ethical perspective. Shafer (2002) introduces Jones' (1991) theory of moral reasoning to explain the prevalence of earnings management in practice. Based on a study of CPAs employed as senior executives, Shafer (2002) concluded that organizational pressures impact the likelihood of executives manipulating financial results. Furthermore, Shafer noted that the reporting decision of financial executives continue to be influenced by quantitative materiality heuristics even when misstatements are clearly material on qualitative grounds. These results are particularly noteworthy since the study was conducted after the issuance of SAB 99, which clearly stated that quantitative thresholds alone should not influence assessment of materiality (Fig. 1).

Other recent studies have investigated the perceptions of management and auditors of the ethicality of earnings management. Shafer, Morris, and Ketchand (1999) and Ketchand, Morris, and Shafer (1999) noted that the dollar amount of misstatements but not qualitative variables exert a significant influence on auditors' judgments of the *ethical acceptability* of acquiescing in clients' reporting schemes. Kinney, Burgstahler, and Martin (2000) reported results consistent with those of Shafer (2002), Shafer, Morris, and Ketchand (1999) and Ketchand, Morris, and Shafer (1999) observing that auditors were less likely to require correction of quantitatively immaterial errors if those errors would cause the company's earnings to fall below analyst-forecasted targets. Conflicting research has been forthcoming regarding the significance to which stockholders attribute importance to the ethicality of management as evidenced in stock price differentials across firms (Balsam, Bartov, & Marquardt, 2002; Rao & Brooke, 1996).

1. The motivation of the otherwise quantitatively immaterial misstatement or omission (i.e., intentional manipulation, disagreement of opinion, unintentional error.)
2. Whether the misstatement or omission, otherwise quantitatively immaterial, conceals a change in earnings trend.
3. Whether the misstatement or omission, otherwise quantitatively immaterial, hides a failure to meet analysts' consensus expectations.
4. Whether the misstatement or omission, otherwise quantitatively immaterial, changes a loss into income.
5. Whether the misstatement or omission, otherwise quantitatively immaterial for the firm as a whole, pertains to a division of the firm that has been identified as especially important to the success of the firm.
6. Whether the misstatement or omission, otherwise quantitatively immaterial, has the effect of qualifying management for bonus incentive compensation.
7. Whether the misstatement or omission, otherwise quantitatively immaterial, involves concealment of an unlawful act.
8. Whether the auditor expects a known misstatement or omission, otherwise quantitatively immaterial, will result in a significant stock market reaction.
9. Whether a known misstatement or omission is offset by another misstatement of opposite income consequences.

Fig. 1. Qualitative Characteristics that May Impact Assessment of Materiality.

No public research of which we are aware has been forthcoming subsequent to the Enron fiasco and related reforms that examine which potential cues (to management ethicality) auditors believe to be relevant and reliable. Nor has there been research regarding if and how auditors incorporate potential cues to the ethicality of management into decisions regarding the extent and scope of planned audit examinations or materiality assessments for contested items in financial reports. If accountants and auditors pay little heed to qualitative factors in materiality assessments, as evidence in the past would have us believe, auditors and auditor firms do otherwise in their client acceptance and retention policies currently. Before accepting a new client or deciding to continue as auditor with an existing client, KPMG openly acknowledges that consideration is given to corporate management philosophy, operating style and integrity, all of which are incorporated in its KRISK methodology and software (KPMG, 2005). If management integrity is a key factor in client acceptance and retention, arguably it should be an equally important factor in assessing materiality and the need to adjust financial statements to purge these intentional self-serving errors. However, materiality, per se, is a user centric concept. Do users believe there is a need to purge/adjust management's financial statements for quantitatively immaterial items that nonetheless exhibit qualitatively important characteristics per SAB 99? This is the question we address in our research.

In this study, we ask financial statement users directly whether and which of the SAB 99 qualitative factors they believe render quantitatively immaterial amounts material, and therefore require adjustment or a qualified audit report. Our purpose is to confirm or deny the assumptions on which SAB 99 policy is based. We also asked financial statement users to assess the materiality of selected contested items embedded in four scenarios to determine if their direct survey responses are reliable. That is, will their materiality assessments actually incorporate the qualitative factors that survey results indicate they say are important? Hopefully, our findings provide a glimpse of users' expectations of auditors, their public "watchdogs," to mitigate intentional albeit quantitatively immaterial misstatements of financial reports by management to achieve personal gains.

RESEARCH QUESTION, DESIGN AND TESTABLE HYPOTHESES

Research Question

The unethical reporting practices of certain captains of industry have received tremendous press attention in recent years. Attempts to improve the current capital markets system were inevitable. Similar disasters have consistently led to legislated or regulatory reforms (e.g., the formation of SEC following the great stock market crash, the pension reform act and foreign corrupt practices act following respective widespread corporate malfeasance in those areas). For its part, the SEC issued SAB 99 and reminded accountants and auditors to consider qualitative factors in assessing materiality. However, the SEC too failed to provide any evidence to support their claim that certain qualitative factors may render quantitatively immaterial items material *in the minds of investor stakeholders, or other stakeholders.* Whether investors believe the qualitative factors identified in SAB 99 would affect their investment decisions is an empirical question. If investor stakeholders participating in our study deem these items material, the implication which we made explicit in the study is that auditors will be held responsible to do likewise and failing to do so, auditors become legally liable. We therefore sought our participants to answer the following Research Question:

> **Research Question.** To what extent do you as a financial statement user believe the qualitative factors identified in SAB 99 are germane to assessments of materiality and "should influence assessments of materiality ex-ante by auditors and ex-post by jurors?"

Research Design

Participants

The participants in the study were 108 MBA students enrolled in the evening program of a large metropolitan university. The profile of the participants made them attractive to a study of this kind because they fit the mold of a typical reasonably informed investor. The average age of participants in this study was 30 years (ranging from 22 to 46 years) with an average work experience of eight years (ranging from 2 to 32 years). The sample consisted of 33 females and 77 male participants. Eighty percent of the participants currently or previously owned stock. The participants also had a baseline understanding of accounting and finance. On the average, the participants had taken four accounting courses and three finance courses (mostly introductory), about half taken as part of their undergraduate programs in years past and half in their current graduate MBA program of study (see Table 1 for descriptive statistics).

Table 1. Descriptive Statistics.

Characteristic	Number/Frequency	Mean	SD
Participants	110 (100%)		
Male participants	77 (70%)		
Female participants	33 (30%)		
Age		28.07	4.35
Years of experience		8.25	5.79
Number of accounting courses		3.98	2.58
Number of finance courses		2.57	2.10
Number of participants with past or current investment experience	88 (80%)		

It is also important to note that the participants identified themselves as part of management. All were undertaking a program of study to advance their management careers and the program of study was underwritten by their employers. To this extent, participants provide a "strong" test of our hypotheses, potentially tilted against favoring SEC mandates for greater management disclosures. That is, the SEC mandates target management for reform and participants identify themselves as part of management.

Direct Survey

Participants were presented with the list of nine qualitative factors identified in SAB 99. Each participant was asked to identify on a 10-point scale (1 = strongly disagree; 10 = strongly agree) the degree to which they agreed that each of the factors should influence materiality assessment. Specifically, the participants were asked to use the following scale:

> Using the following scale, please indicate the extent to which you believe the following qualitative characteristics should influence assessments of materiality. (ex-ante by auditors and ex-post by jurors)
>
> Strongly Disagree 1 … 2 … 3 … 4 … 5 … 6 … 7 … 8 … 9 … 10 Strongly Agree

The participants had previously been advised as to the prevailing generic definition of materiality that allowed for but was operationally vague regarding qualitative factors. Specifically, they were told:

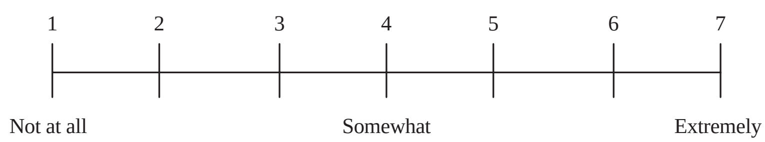
> The omission or misstatement of an item in a financial report is material if, in the light of surrounding circumstances, the magnitude of the item is such that it is probable that the judgment of a reasonable person relying upon the report would have been changed or influenced by the inclusion or correction of the item … (however) magnitude by itself, without regard to the nature of the item and the circumstances in which the judgment has to be made, will not generally be a sufficient basis for a materiality judgment. (Financial Accounting Standards Board)

If the SEC is correct in SAB 99, the judgments of financial statement users should be significantly influenced. Thus, for this phase of the study, we hypothesize:

Materiality Impact Hypothesis. User/stakeholders will agree that the SEC identified qualitative factors "should influence assessments of materiality (ex-ante by auditors and ex-post by jurors)." Agreement is operationally defined as a response significantly greater than a mid-point score of 5 on the provided 10-points scale.

Indirect Assessments: Case Scenarios

Four case scenarios were also developed to address the influence of selected qualitative and quantitative factors identified by the SEC on the materiality assessments of certain omissions and misstatements by financial statement users (our participants). In each of the four cases, respondents were asked to assess whether the cited misstatement or omission was material in light of the qualitative surrounding circumstances. Responses were made using the following scale:



Research Cases and Related Hypotheses

Four cases involving seven SAB 99 qualitative factors and one quantitative factor were presented to financial statement users. The objective of this phase of the study was to determine whether the SAB 99 factors do indeed "experimentally" impact materiality judgments in a fashion consistent with responses to the direct survey. That is, participants may say that qualitative factors are or are not important in the survey and their responses to the case scenarios will document whether they impound or do not incorporate qualitative factors in their judgments.

Wincort Case. *Wincort* is a manufacturing company in existence for over 50 years being audited by the same audit firm for the last 10 years. *Wincort* managers and their external auditors (CPAs) are noted to be in a disagreement over the accounting treatment for certain revenues and expenditures. The nature of the misstatement in this case is a reporting error (i.e., interpretation and application of authoritative standards, otherwise known as "generally accepted accounting standards" (GAAP). GAAP is established by pronouncements of the FASB and the SEC. In the Wincort case, auditors believe they have uncovered an instance of *intentional* manipulation of financial statements by management. The magnitude of the misstatement however is only 2% of net income and thus well below the 5% rule-of-thumb. The case further clarifies that, irrespective of whether the financial statements are adjusted for the misstatement or not adjusted for the misstatement, the company will show a profit and meet consensus earnings expectation.

Four variants of the case were prepared to accommodate the experimental manipulation of two SAB 99 qualitative factors. In two of the four versions of the case, the *earnings trend* of the company only will be maintained if management's figures are reported. In the other two versions of the case, trends are not dependent on management's figures; adjusted numbers would still maintain past earnings trends. The other manipulation related to whether *management's qualification for an incentive compensation bonus* was dependent on the misstated figures. Thus, we manipulated two SAB 99 factors: (i) whether the misstatement or omission, otherwise quantitatively immaterial, conceals a change in earnings trend, and (ii) whether the misstatement or omission has the effect of qualifying management for bonus incentive compensation. We used a 2×2 ANOVA design.

The Wincort case was designed to constitute a strong test of SAB 99 assumptions. It is a strong case because many controlled variables argue against materiality. Among those factors are the small amount of the misstatement relative to income (i.e., the 2% item is less than half the 5% rule-of-thumb), and the irrelevance of the item to meeting the consensus earnings estimates of Wall Street financial analysts (i.e., the outcome is unlikely to affect stock prices). To be judged material under these circumstances would provide strong support for SAB 99 contentions that bonus and/or trend qualitative variables are important.

For the *Wincort* case, specific hypotheses were:

Bonus Hypothesis. Materiality assessments of a 2% intentional reporting error will be significantly greater if the misstatement serves to qualify management for an incentive compensation bonus.

Trend Hypothesis. Materiality assessments of a 2% intentional reporting error will be significantly greater if the misstatement serves to perpetuate past earnings trends.

Qualtec Case. *Qualtec* is also a manufacturing company in existence for 50 years being audited by the same audit firm for the past 10 years. Qualtec management and auditors are in disagreement over the accounting for certain revenues and expenses. In this case, the nature of the misstatement and disagreement is one of *estimates* (subjective determination of certain numbers). Past research has shown that auditors are less likely to require adjustments to estimates than to reporting errors, under a rationale of lack of materiality. It is also known that members of management lacking integrity know this and use this to their advantage; that is, they know that manipulated estimates are easier to get past auditors than contested interpretations of authoritative standards as included in the prior Wincort case. In all four versions of the Qualtec case, the situation is that management's unadjusted figures meet consensus analysts' earnings expectations while statements adjusted to the auditors' estimates would not meet consensus analysts' earnings expectations. Thus, the question of management's motivation is made salient. Failure to meet consensus analysts' expectations typically results in a drop in stock price, sometimes a precipitous drop in stock prices. Information inconsistent with consensus earning expectations constitutes new information to the market because prevailing theory is that consensus earnings expectations incorporate all extant public information.

Four experimental versions of the case were developed. In two versions, auditors' assessment of the motivation (of the management) for the misstatement was noted to be *intentional* while in the other two versions, the motivation for the misstatement was deemed to be *unintentional* by the auditors. Disagreements between corporate accountants and their external auditors over accounting estimates are not uncommon and often do not reflect untoward intentions. Accordingly, it is entirely possibly that disputes over estimates may be judged to be intentional manipulations by auditors or unintentional misinterpretations or misuse of data. The other factor manipulated in the case was the *relative size of the disputed estimate*. Thus, the Qualtec case pitted one qualitative factor against one quantitative factor. In two versions of the case the magnitude of the misstatement was set at 2% of net income and in the other two versions the magnitude of the misstatement was set at 4%. Note that in all versions, the magnitude of the misstatement is set below the traditional materiality threshold of 5%. The SAB 99 qualitative factor considered in this case is the motivation of the otherwise quantitatively immaterial misstatement or omission. The magnitude and intent factors were manipulated by employing a 2×2 ANOVA design. For the *Qualtec* case, we hypothesize:

Intention Hypothesis. The materiality assessment of a disputed accounting estimate will be greater if the estimate is judged to be intentionally misstated rather than the result of an unintentional motive.

Magnitude Hypothesis. The materiality assessment of a disputed accounting estimate will be greater if the disputed amount is 4% of net income rather than 2% of net income.

Toplink Case. The base facts in the *Toplink Corporation* case are similar to the *Wincort* and *Qualtec* cases. The nature of the misstatement in this case was a reporting error (similar to the Wincort case), but this time equal to 4% of net income. The magnitude of the misstatement thus was again below the 5% rule-of-thumb that the SEC contests. The motivation for the misstatement was unknown. That is, in contrast to prior cases, the participants are not provided any direct input on management's motives. For instance, in the Wincort case, participants were told that auditors believed the reporting error was intentional: in the Qualtec case intentionality was manipulated to test the significance of intentional versus unintentional motivations. In the Toplink case, motive would have to be imputed from other information if the participants deemed motive to be relevant to their judgment. Both adjusted and unadjusted financial statements meet consensus analysts' expectations in the Toplink case. And, management does not qualify for a bonus based on either an adjusted or unadjusted statement.

Four versions of the case were prepared in which two SAB 99 qualitative factors were manipulated using a 2×2 ANOVA design. The two SAB 99 factors were (i) whether the misstatement or omission, otherwise quantitatively immaterial, involves concealment of an unlawful act and (ii) whether the misstatement or omission, otherwise quantitatively immaterial, changes a loss into income. In two versions, the misstatements would conceal unlawful conduct while in the other cases there is no unlawful conduct. Likewise, in two versions both adjusted and the unadjusted statements indicated that the company was profitable while in the other two versions, the unadjusted statement reports the firm is profitable whereas an adjusted statement would report a loss. For the two aforementioned SAB 99 qualitative factors, we test the following hypotheses:

Unlawful conduct Hypothesis. Materiality assessments of a 4% reporting error will be significantly greater if the misstatement serves to conceal unlawful conduct.

Profit/Loss Hypothesis. Materiality assessments of a 4% reporting error will be significantly greater if the misstatement makes a difference between reporting a profit or a loss.

Ameriday Case. The base facts in the *Ameriday Corporation* case are similar to the others. The nature of the misstatement in this case was a contested estimate (similar to the Qualtec case) but equal to 4% of net income under all conditions. The magnitude of the misstatement thus was again below the 5% rule-of-thumb that the SEC contests. The motivation for the misstatement was again unknown, similar to the Toplink case. In contrast to the Wincort and Qualtec cases, but similar to the Toplink case, the participants are not provided any direct input on management's motives. Motive would have to be imputed from other information if the participants deemed motive to be relevant to their judgment. Both adjusted and unadjusted financial statements would report a loss and fail to meet consensus analysts' expectations. Management does not qualify for a bonus based on either an adjusted or unadjusted statement. Given the financially distressed (loss) condition of the firm, one might speculate that even a contested estimate would be deemed material.

Four versions of the case were prepared in which two SAB 99 qualitative factors were manipulated using a 2×2 ANOVA design. The two SAB 99 factors were (i) whether the misstatement or omission, otherwise quantitatively immaterial, would serve to reduce the loss or exacerbate the loss and (ii) whether the misstatement or omission, otherwise quantitatively immaterial, offsets another misstatement. No authoritative literature exists that suggests items otherwise similar are more material because directionally they serve to increase/decrease income or loss. Nonetheless, a major accounting abuse often cited by SEC Commissioner Levitt was the creation (and use) of "cookie jar reserves." Cookie jar reserves involve the intentional creation of expense reserves in one year (thus lowering that year's income or reporting larger losses) with the intention of reversing the reserves allowances in later years to boost income and create perceptions of stable increasing earnings trends. The qualitative issue of whether a known misstatement or omission is offset by another misstatement of opposite income consequence is another item in the SEC qualitative factors list. Arguably, if management has advanced two offsetting misstatements, it may be difficult to attribute to management a motive of income manipulation, because net income would not be affected. Nonetheless, to the extent that two items offset and neither is reported, disclosure to the public is reduced. The significance of the knowledge of the two offsetting items might not be equivalent. On the other hand, if a 4% item is income increasing and not offset, a clear income effect is manifest and motive might be more easily if not always correctly inferred. Specific hypotheses are as follows:

Income Direction Hypothesis. Materiality assessments of a 4% disputed accounting estimate will be significantly greater if the misstatement serves to reduce reported losses, as opposed to increase losses. (Normatively, there should be no difference).

Offset Hypothesis. Materiality assessments of a 4% disputed accounting estimate will be significantly greater if the misstatement is not offset by another misstatement of opposite income/loss consequence.

Tolerance for Ambiguity

Tolerance for ambiguity is the psychological characteristic that affects how people react to ambiguity. Individuals who display high tolerance for ambiguity are more comfortable dealing with complex, probable, unstructured and contradictory information. They tend not to deny, nor distort complexity or incongruity. A person who is intolerant of ambiguity, on the other hand, prefers simpler solutions and rules-of-thumb, items easily interpretable. They may prefer information expressed in numbers imputing a greater factual grounding. When faced with complex information sets, their decisions reflect simplistic rules (e.g., use defaults, use averages and/or ignore inconsistency). Because issues relating to materiality involve complex and contextual issues, individuals with varying tolerance for ambiguity might be expected to respond differently to our four cases. Accordingly, for control purposes, all participants were asked to complete the standard 20 question Tolerance for Ambiguity Scale questionnaire (TFA 20, MacDonald, 1970). Use of the Tolerance for Ambiguity Scale is widespread in the business and accounting research. Based on their responses participants were partitioned into two groups: low tolerance and high tolerance. This partitioned factor was included in the ANOVA analyses reported for our four cases.

RESULTS AND ANALYSIS

Survey Results

Participants were presented with the nine qualitative factors identified in SAB 99 as being important in assessing materiality. The participants were asked to identify on a 10-point scale (1 = strongly disagree; 10 = strongly agree) the extent to which they agreed that each of the factors should influence materiality assessment. In general, participants strongly concurred with SAB 99 and indicated that eight of nine qualitative factors were highly germane to assessing materiality. For each question (except question 9), the mean score was at or above 7.2 (four items exceeded a score of 8) and statistically significantly different from 5 (the point of indifference and scale mid-point) indicating support for the Materiality Impact Hypothesis.

Respondents were ambivalent regarding whether a known misstatement or omission is offset by another misstatement of opposite income consequences (qualitative factor 9). This issue is addressed again in the Ameriday Case. The results are presented in Table 2.

An explanation for the low score for item 9 is tenuous at best. Arguably, the concept itself is too complex to easily interpret. On the one hand, offsetting misstatements would seem to argue against any intentional effort to manipulate income in the current year, everything else being equal. On the other hand, one now confronts a situation of not one misstatement but two, which may not be equivalent in nature or type, albeit equivalent in effect on bottom-line net income. Given the potential ambiguity of the question, participants were partitioned into two groups: those scoring lower on the standard Tolerance for Ambiguity Scale and those scoring higher. Somewhat stereotypically, low-tolerance individuals hedged their responses producing a mean response of 5, precisely the scale mid-point. Higher tolerance participants had a mean response of 5.7. This difference in scores is not statistically significant. Thus, we are left with the conclusion, based on the survey results, that the last qualitative item in the SEC list was not seen as particularly germane to questions of materiality. Phase 2 of this research study re-examines this issue.

User/stakeholders' responses were also compared with those of 48 audit seniors obtained at a national firm CPE training program conducted by a BIG 4 public accounting firm. Table 2 includes these scores. Notable similarities can be found across the two groups. First, the highest scoring and lowest scoring items are the same. Second, the average (mean) scores for each item is above the scalar mid-point (5 on the 10-point scale) and again significantly so statistically for all item except item 9. Thus, our findings are such that both current and prospective members of management and current audit seniors (the two groups targeted by SEC reforms) indicate general agreement with the basic tenets of SAB 99, based on direct survey responses. For eight of the nine items, mean responses for both groups were significantly greater than a scale mid-point of 5, at .01 level.

Table 2. Agreement with Factors Identified by SAB 99 as Important for Assessing Materiality.

Qualitative Factor	MBAs Mean	MBAs SD	CPAs Mean	CPAs SD
The motivation of the otherwise quantitatively immaterial misstatement or omission (i.e., intentional manipulation, disagreement of opinion, unintentional error)	7.62	2.08	6.96	3.11
Whether the misstatement or omission, otherwise quantitatively immaterial, conceals a change in earnings trend	8.25	1.50	7.06	3.14
Whether the misstatement or omission, otherwise quantitatively immaterial, hides a failure to meet analysts' consensus expectations	7.25	1.92	6.98	3.31
Whether the misstatement or omission, otherwise quantitatively immaterial, changes a loss into income	8.05	1.79	6.98	3.28
Whether the misstatement or omission, otherwise quantitatively immaterial for the firm as a whole, pertains to a division of the firm that has been identified as especially important to the success of the firm	7.22	1.80	6.46	2.96
Whether the misstatement or omission, otherwise quantitatively immaterial, has the effect of qualifying management for bonus incentive compensation	7.82	2.05	7.19	3.33
Whether the misstatement or omission, otherwise quantitatively immaterial, involves concealment of an unlawful act	9.15	1.61	7.52	3.46

Table 2. (*Continued*)

Qualitative Factor	MBAs Mean	MBAs SD	CPAs Mean	CPAs SD
Whether the auditor expects a known misstatement or omission, otherwise quantitatively immaterial, will result in a significant stock market reaction	8.13	1.89	7.08	3.54
Whether a known misstatement or omission is offset by another misstatement of opposite income consequences	5.37	2.38	5.48	3.15

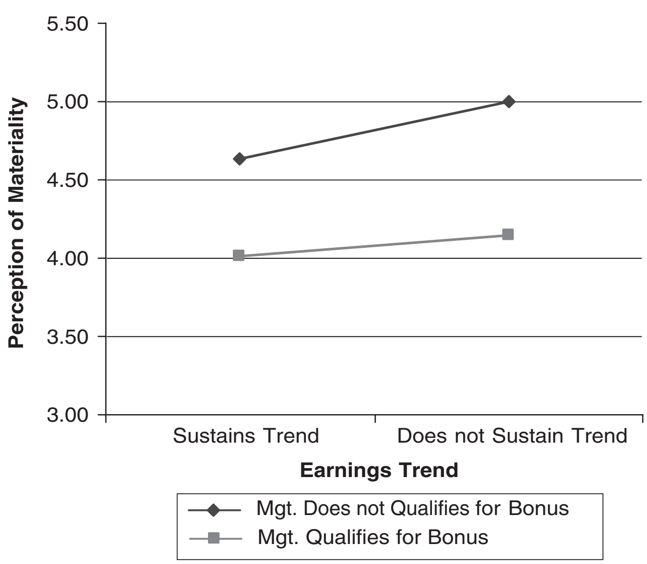
Empirical (Case Study) Results

Wincort Case. Recall that in the *Wincort* case the two experimentally manipulated SAB 99 factors are: (i) whether the misstatement or omission, otherwise quantitatively immaterial, has the effect of *qualifying management for bonus* incentive compensation and (ii) whether the misstatement or omission, otherwise quantitatively immaterial conceals a *change in earnings trend*. Results of this case are presented in Table 3. *ANOVA result indicated strong support for the Bonus Hypothesis* (*F*=*5.372*, *p*=*.022*). *The Trend Hypothesis was not supported* (*F*=*1.024*, *p*=*.314*). In retrospect, the lack of support for the Trend Hypothesis is not surprising. Note that in this case, both the adjusted and the unadjusted statements meet analysts' consensus earnings expectation, which tends to be more important than trends. Given that analyst expectations are met by the adjusted statements as well as by the unadjusted (misstated) figures, savvy users may not be too concerned with earnings trend. Analysts' consensus earnings estimates factor in some earnings trend variance and thus limited trend variation is unlikely to affect stock prices. Strong support for the Bonus Hypothesis reveals users' sensitivity to management integrity. Users appear to be troubled by management's exploitation of GAAP loopholes in their own self-interest and for their personal economic benefit. *Participants expect* "*watchdog*" *auditors to stop untoward behavior wherever they find it* (*and have influence*) *even for quantitatively immaterial items*. This result is not surprising because of the increased press attention focused on the largesse that top managers have been receiving even in economically challenging times and even in years

Table 3. Wincort Case Analysis.

Source	Type III Sum of Squares	df	Mean Square	*F*	Significant Values
Corrected model	20.745(a)	7	2.964	1.421	.205
Intercept	2067.526	1	2067.526	991.081	.000
Ambiguity	.248	1	.248	.119	.731
Bonus	11.207	1	11.207	5.372	.022
Trend	2.135	1	2.135	1.024	.314
Ambiguity × bonus	.408	1	.408	.196	.659
Ambiguity × trend	3.599	1	3.599	1.725	.192
Bonus × trend	.519	1	.519	.249	.619
Ambiguity × bonus × trend	.000	1	.000	.000	.998
Error	208.613	100	2.086		
Total	2358.250	108			
Corrected total	229.359	107			

	Mean	SE	95% Confidence Interval	
			Lower bound	Upper bound
Bonus				
No	4.783	.200	4.387	5.180
Yes	4.127	.201	3.729	4.525
Trend				
No	4.599	.200	4.202	4.996
Yes	4.312	.200	3.915	4.709



when a company has performed dismally. This result is particularly noteworthy because the misstatement level is only 2% of net income (well below the traditional 5% threshold). *Users nonetheless believe that management of earnings for self economic benefit even though quantitatively immaterial rises to a level of somewhat material* (mean of 4.63 when trend is sustained and 5.0 when trend is not sustained, on a 7-point scale; both are statistically different than the scale mid-point of 4 at a .01 level). Less versus more ambiguity-tolerant participants did not record significantly different responses. *Findings in the Wincort case indicates that selected qualitative factors are relevant to financial statement users*; *in this case specifically*, *management's illicit qualification for a bonus. Also*, *under various conditions*, *participants' responses significantly* (*statistically*) *exceed the scale mid-point of 4 clearly indicating selected quantitatively immaterial items indeed were deemed material.*

Qualtec Case. In the *Qualtec* case, we manipulated the intent of management and the magnitude of misstatement to gauge their effect on user judgments regarding materiality assessments. Results of the case are presented in Table 4. *ANOVA results indicate the Intention Hypothesis* (*intentional v. unintentional*) *is supported* ($F = 4.918$, $p = .029$) *but the Magnitude Hypothesis* (*2% v. 4%*) *is not supported* ($F = 1.794$, $p = .183$). The intention manipulation is most salient among ambiguity intolerant participants. (Note a significant intent by ambiguity interaction: $F = 10.161$, $p = .002$.) We also find a marginally significant intent by amount interaction where the intent factor is most salient under conditions of a 4% misstatement. Intentional misstatements elicit a mean response above 5 (on a 7-point scale) for ambiguity intolerant participants. This is statistically different than the scale mid-point of 4 at a .01 level. Intentional misstatements across both subject groups at the 4% level also generate mean responses above 5 (5.168; statistically significant at .01 level), indicating the misstatement is deemed material. *Similar to findings in the Wincort case*, *findings in the Qualtec case indicate that qualitative factors are relevant to financial statement users*; in this case specifically, intent. *Also*, *under various conditions*, *participants' responses significantly exceed the scale mid-point of 4 clearly indicating selected quantitatively immaterial items were deemed material.*

The lack of significance for the amount manipulation may be due to the fact that irrespective of whether the misstatement is 2% vs. 4%, adjustment of the misstatement at either level will cause the firm to fail to meet the consensus earnings estimate.

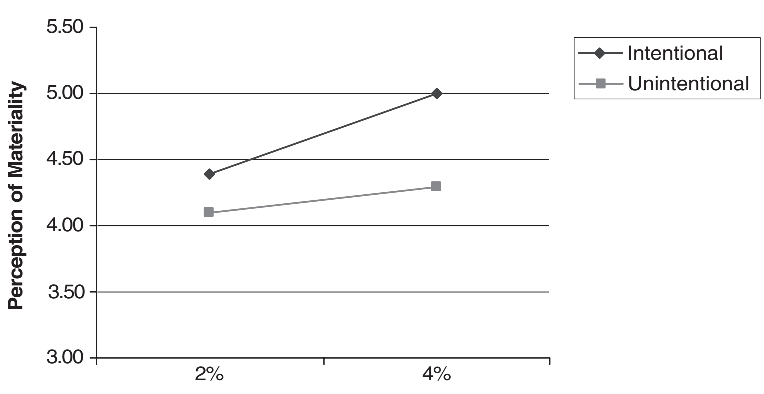
Table 4. Qualtec Case Analysis.

Source	Type III Sum of Squares	df	Mean Square	F	Significant Values
Corrected model	41.507(a)	7	5.930	3.232	.004
Intercept	2120.042	1	2120.042	1155.440	.000
Intent	9.024	1	9.024	4.918	.029
Amount	3.292	1	3.292	1.794	.183
Ambiguity	2.636	1	2.636	1.437	.234
Intent × amount	4.866	1	4.866	2.652	.107
Intent × ambiguity	18.644	1	18.644	10.161	.002
Amount × ambiguity	2.186	1	2.186	1.192	.278
Intent × amount × ambiguity	1.478	1	1.478	.805	.372
Error	183.483	100	1.835		
Total	2403.000	108			
Corrected total	224.991	107			

Intent	Mean	SE	95% Confidence Interval	
			Lower bound	Upper bound
No	4.191	.187	3.820	4.562
Yes	4.776	.186	4.407	5.145

Intent	Ambiguity	Mean	SE	95% Confidence Interval	
				Lower bound	Upper bound
No	Low	3.613	.273	3.071	4.154
	High	4.770	.255	4.263	5.276
Yes	Low	5.038	.266	4.511	5.566
	High	4.514	.261	3.996	5.031

Intent	Amount	Mean	SE	95% Confidence Interval	
				Lower bound	Upper bound
No	4%	4.153	.270	3.617	4.688
	2%	4.229	.259	3.716	4.742
Yes	4%	5.168	.261	4.650	5.685
	2%	4.385	.266	3.858	4.912



Toplink Case. In the *Toplink* case, the two experimentally manipulated SAB 99 factors are: (i) concealment of unlawful conduct and (ii) whether the misstatement has the effect of changing loss into profit. The results are presented in Table 5. *ANOVA results indicate very strong support for the Unlawful Conduct Hypothesis* ($F = 29.401$, $p = .001$) *and for the Profit/Loss Hypothesis* ($F = 11.993$, $p = .001$). Note that the magnitude of the misstatement is set at 4% (below the traditional 5% threshold). Despite the 4% level, unlawful acts register a mean response of 5.967; and items that change a loss into a profit elicit a mean response of 5.732. Both these figures are statistically different from the scale mid-point of 4 at the .01 level. These results indicative users' beliefs that management's ethics and integrity are highly relevant in assessing materiality. A marginally significant interaction of ambiguity tolerance and lawful actions ($F = 2.895$, $p = .092$) conforms to findings of past research that ambiguity-tolerant individuals exhibit greater sensitivity to context relevant stimuli. In this instance, ambiguity tolerant participants register both the lowest mean score (4.460) for lawful conduct and the highest mean score (6.167) for unlawful conduct. *Similar to findings in the prior cases, findings in the Toplink case indicate that qualitative factors are relevant to financial statement users, consistent with survey responses. Further, quantitatively immaterial items (4%) in general elicit responses significantly greater than the scale mid-point of 4 (perceptions of lawful acts by ambiguity tolerant participants being the exception).*

Ameriday Case. In the *Ameriday* case, the two experimentally manipulated SAB 99 factors are: (i) offsetting versus not offsetting misstatements and (ii) their increasing income versus decreasing income effect. The results are presented in Table 6. ANOVA results indicate neither factor is significant directly or interactively. In addition, no differences were exhibited between ambiguity-tolerant and ambiguity-intolerant individuals. *Non-significance for the income increasing versus decreasing qualitative factor is significant in and of itself. This is a normatively correct response.* Untoward behavior can underlie intentional misstatements in either direction.

The offsetting item manipulation, on the other hand, was predicted to elicit a reaction. It did not. Still, this is *consistent with participants' responses to the direct survey* in Phase 1 of this study, where participants ranked the offset qualitative dimension as relatively unimportant to assessing materiality and not significantly greater than the scale mid-point. Phase 1 of the study was conducted independent of and subsequent to Phase 2 in an effort to avoid any demand effects that the survey might impose on the experiments.

Table 5. Toplink Case Analysis.

Source	Type III Sum of Squares	df	Mean Square	*F*	Significant Values
Corrected model	82.211(a)	7	11.744	7.869	.000
Intercept	2940.326	1	2940.326	1969.966	.000
Ambiguity	.001	1	.001	.001	.975
Lawful act	43.884	1	43.884	29.401	.000
Profit/loss	17.901	1	17.901	11.993	.001
Ambiguity × lawful act	4.321	1	4.321	2.895	.092
Ambiguity × profit/loss	.583	1	.583	.391	.533
Lawful profit/loss	3.335	1	3.335	2.234	.138
Ambiguity × lawful × profit/loss	2.158	1	2.158	1.446	.232
Error	147.765	99	1.493		
Total	3229.250	107			
Corrected total	229.977	106			

Lawful	Mean	SE	95% Confidence Interval	
			Lower bound	Upper bound
No	5.967	.166	5.638	6.295
Yes	4.667	.173	4.324	5.011
Profit/loss				
Loss	5.732	.172	5.391	6.073
Profit	4.902	.167	4.571	5.233

Ambiguity	Lawful	Mean	SE	95% Confidence Interval	
				Lower bound	Upper bound
Low tolerance	No	5.766	.235	5.300	6.233
	Yes	4.875	.253	4.373	5.377
High tolerance	No	6.167	.233	5.704	6.630
	Yes	4.460	.236	3.991	4.929

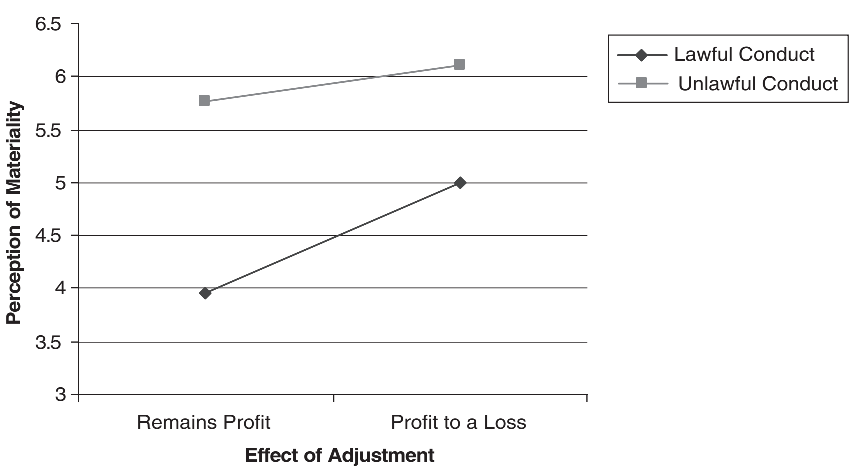


Table 6. Ameriday Case Analysis.

Source	Type III Sum of Squares	df	Mean Square	F	Significant Values
Corrected model	4.431(a)	7	.633	.274	.963
Intercept	1612.526	1	1612.526	697.164	.000
Ambiguity tolerance	.003	1	.003	.001	.973
Offsetting items	2.594	1	2.594	1.122	.292
Income increasing/decreasing	.179	1	.179	.077	.781
Ambiguity tolerance × offsetting items offset	.797	1	.797	.344	.559
Ambiguity tolerance × income increasing/decreasing	.015	1	.015	.007	.935
Offsetting items × income increasing/decreasing	1.060	1	1.060	.458	.500
Ambiguity tolerance × offsetting items income increasing/decreasing	.209	1	.209	.090	.765
Error	231.298	100	2.313		
Total	1904.250	108			
Corrected total	235.729	107			

	Mean	SE	95% Confidence Interval	
			Lower bound	Upper bound
Offset				
No	3.790	.210	3.375	4.206
Yes	4.107	.213	3.684	4.531
Income				
Decrease	3.907	.208	3.495	4.319
Increase	3.990	.215	3.563	4.417

SUMMARY AND CONCLUSION

Allegedly, members of the corporate management community and the public accounting profession have made exclusive use of strictly applied quantitative definitions of materiality to justify and cloak fraudulent earnings management. In doing so, they have allegedly failed in their fiduciary responsibilities to protect the public and brought disrepute to their businesses or profession. Following a series of speeches and articles by SEC Chairman Arthur Levitt in which such abuses were exposed, the SEC proposed reforms in SAB 99. SAB 99 delineated a number of contextual and qualitative factors which the SEC argued raised quantitatively immaterial items to the level of materiality. In this paper, we examine whether selected user/stakeholders (whose interest the SEC is acting to protect) agree with the staff of the SEC regarding a needed recalibration/definition of what is material. The views of 110 business-experienced and investment savvy MBAs were both surveyed and experimentally examined to gauge the extent to which they agreed with SAB 99 proposals. The results were consistent across the two research methods; *the SEC delineated qualitative factors should influence assessments of materiality ex-ante by auditors and ex-post by jurors.*

In ancillary analyses, 48 Big 4 auditors were given the same survey (but not experimental cases). Perhaps surprising, the views of these *practitioners agreed with our MBA participants*: *materiality should be interpreted within the context of qualitative factors and not turn simply and exclusively on quantitative measures.*

The policy issues addressed in SAB 99 and in this paper are no less important and arguably more important than many of the recent Sarbanes–Oxley reforms initiated in response to the financial markets meltdown following the Enron bankruptcy and Andersen indictment. The concept of materiality is the most fundamental standard of financial reporting. Materiality is a concept integrally linked to the fairness of corporate financial reporting. Allegedly, it has become a primary tool in the arsenal of the unethical to take advantage of the American public. As such, the issues addressed herein warrant further serious inquiry. Severe limitations attach to any one research study. We thus invite and encourage others to explore this topic. This is an especially important time to do so. We are entering a period of international harmonization of accounting standards. However, without harmonization of the concept of materiality, harmonization of accounting standards is limited in meaning, if not hollow.

NOTES

1. From the Report of the Independent Registered Public Accounting Firm (Deloitte and Touche LLP) for P&G year ended June 30, 2005.
2. TSC Industries, Inc. v. Northway, Inc. 426 US 438 (1976) and Basic Inc. v. Levinson 485 US 224 (1988).
3. See for example "Objectives of Financial Statements" (AICPA, 1974) were it is stated "An objective of financial statements is to serve primarily those users who have limited authority, ability, or resources to obtain information and who rely on

financial statements as their principal source of information about an enterprise's economic activities'' (on p. 17).

4. For example, in the Framework for Preparation and Presentation of Financial Statements, the Institute of Chartered Accountants of India defines materiality as follows:

> ''Information is material if its misstatement (i.e., omission or erroneous statement) could influence the economic decisions of users taken on the basis of the financial information. Materiality depends on the size and nature of the item or error, judged in the particular circumstances of its misstatement.'' Furthermore ICAI states that ''Although the auditor ordinarily establishes an acceptable materiality level to detect quantitatively material misstatements, both the amount (quantity) and nature (quality) of misstatements need to be considered.''

5. From the SEC Report Pursuant to Section 704 of the Sarbanes–Oxley Act of 2002.

6. Escott v. Barchris Constr. Corp., 340 F.2d 731 (2d Cir.).

7. U.S. v. Arthur Young & Co. 465 US .805 (1984).

8. On a scale of 1–10 where 1 indicated strong disagreement and 10 indicated strong agreement, judges reported a mean score of 7.9, law students 7.3, MBAs 7.2 and auditors 7.1.

9. Auditing standards also reflect on the importance of high management integrity. For example, at AU 325.21, it is stated that ''Evidence of intentional override of IC by those in authority to the detriment of the overall objectives of the system is an example of a possible reportable condition.'' AU 319.28 states ''Concerns about the integrity of the entity's management may be so serious as to cause the auditor to conclude that the risk of management misrepresentation in the financial statements is such that an audit cannot be conducted.'' Further, the Public Corporations Accounting Oversight Board (PCAOB) constituted in the wake of the Enron and related financial disasters observes in their standard on internal control that ''Identification of fraud of any magnitude on the part of senior management is at least a significant deficiency and a strong indicator that a material weakness exists.'' (on p. A-61). See American Institute of Certified Public Accountants (AICPA) (2004).

REFERENCES

American Institute of Certified Public Accountants (AICPA). (1974). *Objectives of financial statements*. New York: NY.

American Institute of Certified Public Accountants (AICPA). (2004). *Codification of auditing standards* (including AICPA and PCAOB auditing and attestation atandards). New York: NY.

Balsam, S., Bartov, E., & Marquardt, C. (2002). Accruals management, investor sophistication, and equity valuation: Evidence from 10-Q filings. *Journal of Accounting Research*, *40*(4), 987–1012.

Blocher, E., & Cooper, J. C. (1988). A study of auditors' analytical review performance. *Auditing: A Journal of Practice and Theory*, *7*(2), 1–28.

Braun, K. W. (2001). The disposition of audit-detected misstatements: An examination of risk and reward factors and aggregation effects. *Contemporary Accounting Research*, *18*(1), 71–100.
Carmichael, D. R., & Willingham, J. J. (1989). *Auditing concepts and methods* (5th ed.). New York: McGraw-Hill.
Chewning, G., Pany, K., & Wheeler, S. (1989). Auditor reporting decisions involving accounting principle changes: Some evidence on materiality thresholds. *Journal of Accounting Research* (Spring), 78–96.
DeFond, M., & Jimbalvo, J. (1991). Incidence and circumstances of accounting errors. *The Accounting Review* (July), 643–655.
Escott v. Bar Chris Constr. Corp., 340 F.2d 731 (2d Cir.).
Financial Accounting Standards Board. (1980). *Statement of financial accounting concepts No. 2*, FASB (1980). Stamford: Connecticut.
Firth, M. (1980). A cross-sectional analysis of qualified audit reports. *International Journal of Accounting* (Spring), 47–59.
Frishkoff, P. (1970). An empirical investigation of the concept of materiality in accounting. *Journal of Accounting Research*, *8*(Supplement), 116–129.
Holder, W. W. (1983). Analytical review procedures in planning the audit: An application study. *Auditing: A Journal of Practice and Theory*, *2*(2), 100–107.
Holstrum, G. L., & Messier, W. F., Jr. (1982). A review and integration of empirical research on materiality. *Auditing: A Journal of Practice and Theory* (Fall), 45–63.
Jones, T. M. (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of Management Review*, *16*(2), 366–395.
Ketchand, A. A., Morris, R. E., & Shafer, W. E. (1999). An analysis of the role of moral intensity in auditing judgments. *Research on Accounting Ethics*, *5*, 249–269.
Kinney, R. K., Jr. (1987). Attention-directing analytical review using accounting ratios: A case study. *Auditing: A Journal of Practice and Theory*, *6*(2), 59–73.
Kinney, R. K., Jr., & McDaniel, L. S. (1989). Characteristics of firms correcting previously reported quarterly earnings. *Journal of Accounting and Economics*, *11*, 71–93.
Kinney, W., Burgstahler, D., & Martin, R. (2000). *The materiality of earnings surprise*. Working Paper. The University of Texas at Austin.
KPMG. (2005). *West area faculty symposium, Los Angeles*, July 2005.
Levitt, A. (1998). *The numbers game.* Speech delivered at the New York University Center for Law and Business (September 28).
Levitt, A. (2000). *Renewing the covenant with investors.* Speech delivered at the New York University Center for Law and Business (May 10, 2000).
MacDonald, A. P., Jr. (1970). Revised scale for ambiguity tolerance: Reliability and validity. *Psychological Reports*, *26*, 791–798.
Maines, L. A., & McDaniel, L. S. (2000). Effects of comprehensive-income characterstics on nonprofessional investors' judgments: The role of financial-statement presentation format. *The Accounting Review*, *75*(2), 179–207.
Rao, S. M., & Brooke, H. J., III. (1996). The effect of published reports of unethical conduct on stock prices. *Journal of Business Ethics*, *15*(12), 1321–1330.
Reckers, P., Lowe, J., Jennings, M., & Pany, K. (2005). *Judges' attitudes toward the public accounting profession: A longitudinal analysis.* Working Paper, Arizona State University.
Securities & Exchange Commission. (1999). "Materiality", Staff Accounting Bulletin No. 99, Washington, DC, August 12.

Shafer, W. E. (2002). Effects of materiality, risk, and ethical perceptions on fraudulent reporting by financial executives. *Journal of Business Ethics*, *38*, 243–262.
Shafer, W. E., Morris, R. E., & Ketchand, A. A. (1999). The effect of formal sanctions on auditor independence. *Auditing: A Journal of Practice and Theory*, *18*(Suppl.), 85–101.
U.S. v. Arthur Young & Co. 465 US 805 (1984).
Wright, S., & Wright, A. M. (1997). The effect of industry experience on hypothesis generation and audit planning decisions. *Behavioral Research in Accounting*, *9*, 273–294.

APPENDIX. ILLUSTRATION OF EXPERIMENTAL INSTRUMENTS

Qualtec Case

QualTec Corporation manufactures and services a variety of monitoring devices used in the home and in industrial quality control operations. The firm is 50 years old and historically is known for high standards of product reliability. The firm has been an audit client of the international public accounting firm of Haskins, Ross & Montgomery CPAs for 10 years.

QualTec management and their auditors currently are in disagreement over the accounting for certain revenues and expenses. The SEC opines that quantitative magnitude by itself may not be a sufficient basis for materiality judgment. In each of the following cases, you will be asked to assess whether the cited misstatement or omission is material in light of the qualitative surrounding circumstances.

Magnitude of item as a percent of net income	(*2%*) *4%*
Nature of misstatement (reporting error vs. contested estimates)	Contested estimates
Motivation for misstatement (intentional vs. unintentional)	(*Un*)*intentional*

Financial Statement Effects	Unadjusted Statements	Adjusted Statements
Reports profit/loss	Profit	Profit
Sustains 3-year earnings growth trend	Yes	Yes
Meets consensus analysts' earnings expectation	Yes	No

Management qualifies for bonus	No	No
Unlawful conduct concealed	No	No
Offset by another misstatement of similar size	No	No

Wincort Case

Wincort Corporation manufactures and services a variety of sensing devices used for home and commercial electronic applications. The firm is 50 years old and historically is known for high standards of product reliability. The firm has been an audit client of the international public accounting firm of Bailey, Sells & Mitchell CPAs for 10 years.

Wincort management and their auditors currently are in disagreement over the accounting for certain revenues and expenses. The SEC opines that quantitative magnitude by itself may not be a sufficient basis for materiality judgment. In each of the following cases, you will be asked to assess whether the cited misstatement or omission is material in light of the qualitative surrounding circumstances.

Magnitude of item as a percent of net income	2%
Nature of misstatement (reporting error vs. contested estimates)	Reporting error
Motivation for misstatement (intentional vs. unintentional)	Intentional

Financial Statement Effects	Unadjusted Statements	Adjusted Statements
Reports profit/loss	Profit	Profit
Sustains 3-year earnings growth trend	Yes	*Yes (No)*
Meets consensus analysts' earnings expectation	Yes	Yes
Management qualifies for bonus	Yes	*Yes (No)*
Unlawful conduct concealed	No	No
Offset by another misstatement of similar size	No	No

Toplink Case

Toplink Corporation manufactures and services home and commercial heating and air-conditioning equipment. The firm has been in business since 1952 and has a reputation for quality design and workmanship. The firm has been an audit client of the international public accounting firm of Whinney & Lybrand CPAs for 10 years.

Toplink management and their auditors currently are in disagreement over the accounting for certain revenues and expenses. The SEC opines that quantitative magnitude by itself may not be a sufficient basis for materiality judgment. In each of the following cases, you will be asked to assess whether the cited misstatement or omission of is material in light of the qualitative surrounding circumstances.

Magnitude of item as a percent of net income	4%
Nature of misstatement (reporting error vs. contested estimates)	Reporting error
Motivation for misstatement (intentional vs. unintentional)	Unknown

Financial Statement Effects:	Unadjusted Statements	Adjusted Statements
Reports profit/loss	Profit	*Profit* (*Loss*)
Sustains 3-year earnings growth trend	No	No
Meets consensus analysts' earnings expectation	Yes	Yes
Management qualifies for bonus	No	No
Unlawful conduct concealed	(*Yes*) *No*	No
Offset by another misstatement of similar size	No	No

Ameriday Case

Ameriday Corporation manufactures a variety of home maintenance and commercial construction products. The firm, founded in 1948, distributes its products under several well-known brand names. The firm has been an audit client of the international public accounting firm of Clarkson Brothers & Oliver CPAs for 10 years.

Ameriday management and their auditors currently are in disagreement over the accounting for certain revenues and expenses. The SEC opines that quantitative magnitude by itself may not be a sufficient basis for materiality judgment. In each of the following cases, you will be asked to assess whether the cited misstatement or omission of is material in light of the qualitative surrounding circumstances.

Magnitude of item as a percent of net income	4%
Nature of misstatement (reporting error vs. contested estimates)	Contested estimates
Motivation for misstatement (intentional vs. unintentional)	Uncertain
Income increasing/decreasing	*Increasing* (*Decreasing*)

Financial Statement Effects:	Unadjusted Statements	Adjusted Statements
Reports profit/loss	Loss	Loss
Sustains 3-year earnings growth trend	No	No
Meets consensus analysts' earnings expectation	No	No
Management qualifies for bonus	No	No
Unlawful conduct concealed	No	No
Offset by another misstatement of similar size	*Yes* (*No*)	No

This page intentionally left blank

EFFECTS OF SUBORDINATE LIKEABILITY AND BALANCED SCORECARD FORMAT ON PERFORMANCE-RELATED JUDGMENTS

Steven E. Kaplan, Michael J. Petersen and Janet A. Samuels

ABSTRACT

Previous appraisal research has found that subordinate manager likeability influences appraisal-related judgments. We hypothesize that when performance measures are presented in an unstructured fashion, evaluators will use an affect-consistency heuristic to simplify the task. Alternatively, when a balanced scorecard (BSC) format is used, the structure of the BSC will guide the evaluations so as to lessen the influence of subordinate likeability on evaluators' performance-related judgments. Unexpectedly, we find that the effect of subordinate manager likeability on performance-related judgments is not lessened by the format and structure of the BSC. We also provide supplemental analysis, which demonstrates that subordinate likeability has both a direct and an indirect effect on bonus allocations.

Advances in Accounting, Volume 23, 85–111

ISSN: 0882-6110/doi:10.1016/S0882-6110(07)23004-4

INTRODUCTION

The balanced scorecard (BSC) was developed by Kaplan and Norton (1992, 1996a, 1996b) for multidimensional performance measurement. A BSC typically contains a set of measures spanning four broad performance categories: financial performance, customer relations, internal business processes, and the organization's learning and growth activities. Researchers have examined a variety of issues related to how a BSC-based system is used to evaluate the performance of subordinate managers and the biases that might influence the performance evaluations (Banker, Chang, & Pizzini, 2004; Dilla & Steinbart, 2005; Libby, Salterio, & Webb, 2004; Lipe & Salterio, 2000, 2002; Roberts, Albright, & Hibbets, 2004; Wong-On-Wing, Guo, Li, & Yang, 2007). This work portrays evaluators as information processors and examines how their cognitive capabilities and shortcomings may systematically influence their performance-related judgments. Management accounting scholars are interested in the quality of performance evaluations because they are typically used by organizations for salary changes, promotions, and providing key feedback to the manager.

Previous experimental research on the use of multiple performance measures, however, has not explored the role, if any, of affect toward subordinate managers on overall performance evaluations.[1] Because of their ongoing professional relationship, evaluators typically will have formed an impression of subordinate managers' likeability. In this regard, subordinate managers possess positive likeability when their personal attributes are generally regarded as having favorable implications. Alternatively, subordinate managers possess negative likeability when their personal attributes are generally regarded as having unfavorable implications. To the extent that performance evaluations are more valid when based exclusively on behaviors and associated performance measures (Balzer, 1986; Dipboye, 1985; Lefkowitz, 2000), consideration of a subordinate's likeability when evaluating the subordinate's performance represents a source of bias and may be dysfunctional. Performance evaluation researchers outside of accounting (Bates, 2002; DeNisi, Robbins, & Summers, 1997; Robbins & DeNisi, 1994, 1998) have shown that subordinate likeability influences evaluators' processing of information about subordinates' performance and their subsequent evaluations of subordinates. For example, research in this area has found that the evaluator's personal liking for a subordinate is associated with both rating accuracy (Cardy & Dobbins, 1986) and rating errors (Tsui & Barry, 1986).

In this chapter we examine whether information about the subordinate's likeability influences performance evaluations in a setting with a large number of performance measures with targets and outcomes. Research in this setting may be distinguished from appraisal research in two important respects. First, the large number of performance measures included in BSCs is intended to represent "an integrated set of leading and lagging performance measures designed to capture the organization's strategy" (Lipe & Salterio, 2000, p. 285). Because the multiple measures provide information about the accomplishments related to the broad range of a firm's strategic objectives, Ullrich and Tuttle (2004) characterize the BSC as a "comprehensive reporting system." To the extent that evaluators use information about a subordinate's likeability because performance-related information is incomplete, subordinate likeability should influence performance evaluations much less under a "comprehensive" set of performance measures. That is, there is much smaller need to infer missing performance-related outcomes from information about a subordinate's likeability under a measurement system intended to be "comprehensive." Second, research investigating the role of subordinate likeability on appraisal judgments generally occurs in a setting without performance criterion (e.g., targets) (Cardy & Dobbins, 1986; Tsui & Barry, 1986; Turban, Jones, & Rozelle, 1990). Baltes and Parker (2000) contend that subordinate likeability will have a stronger influence on evaluators' judgments in settings without performance criterions. BSCs contain a target for each performance measure, which also should reduce one's tendency to rely on information about subordinate likeability when evaluating performance.

Regardless, we expect subordinate likeability to influence evaluators' performance evaluations in settings including a large number of performance-related measures (e.g., outcomes and targets). Lipe and Salterio (2000) characterize the task of evaluating performance based on a large number of performance measures as ambiguous and complex. Consequently, evaluators generally simplify their task through the use of heuristics, which in turn, systematically influence the subjective weightings of performance measures. While BSCs include a large number of performance measures, commonly evaluators have discretion to assign weights to performance measures subjectively (Ittner, Larcker, & Meyer, 2003; Kaplan & Norton, 1996a).

We contend that subordinate affect will trigger the use of an affect-consistency heuristic when evaluating performance based on a large number of performance measures. Under this approach, evaluators attend to and weight information that is consistent with subordinate's likeability (Robbins &

DeNisi, 1994). Thus, even with a "comprehensive" set of performance measures that includes targets, subordinate likeability is expected to influence evaluators' judgments through the discretion given to evaluators to subjectively attend to and weight the multiple performance measures. Further, we propose that the structure of the BSC, in comparison to the same information listed in an unstructured fashion, will lessen the influence of subordinate likeability on evaluators' performance-related judgments of subordinates. Lipe and Salterio (2002) characterize the BSC as a structuring mechanism that simplifies and guides information processing. Thus, an unstructured listing of performance measures represents a more complex task compared to the same information presented and organized within the structure of the BSC. Without the structure of the BSC to simplify and guide information processing, we expect subordinate likeability to have an especially strong influence on directing evaluators' attention and weighting toward measures that are consistent with subordinate likeability.

Lipe and Salterio (2002) contend that the structure of the BSC triggers the use of a divide and conquer approach to simplify the task of evaluating performance based on a large number of performance measures (Shanteau, 1988), and that this approach will be guided by the structure of the BSC (Lipe & Salterio, 2002). Evaluators will simplify the task by dividing it into smaller tasks using the structure provided by the four BSC categories. That is, the task of forming four separate evaluations based on the performance measures presented under each of the four BSC categories and then combining these four evaluations into an overall evaluation is a cognitively less demanding task than processing the entire set of performance measures.

In addition to simplifying the task, we also expect the structure of the BSC to lessen the influence of subordinate likeability on evaluators' performance evaluations. By simplifying the task into smaller tasks, the structure of the BSC is expected to make it more likely that evaluators will attend to each performance and make performance similarities and differences within a BSC more transparent. For example, imagine a setting where a BSC category includes four measures and that two subordinates perform similarly based on the four measures. We contend that the structure of the BSC, presenting the four measures side by side, will make this similarity in performance more apparent to evaluators, and, consequently, lessen the influence of subordinate likeability on subordinate performance-related judgments. Overall, our research is intended to complement and extend research on whether and how the structure of the BSC affects evaluators' subordinate-related performance judgments (Lipe & Salterio, 2002).

Using an experimental approach similar to previous BSC studies (Banker et al., 2004; Dilla & Steinbart, 2005; Libby et al., 2004; Lipe & Salterio, 2000, 2002; Roberts et al., 2004), evaluators were provided with background information about the company and strategy along with performance information for multiple performance measures and outcomes for two managers. Evaluators rated the performance of two division managers, indicated their strength of conviction in their ratings for each manager, and allocated a performance bonus to each manager. The study manipulated performance measures organization (e.g., presence or absence of the BSC format) and information about the subordinates' likeability (e.g., Like Manager A/Dislike Manager B and vice versa) in a fully crossed design.

The results indicate that subordinate likeability influences evaluators' performance evaluation judgments. However, the results do not support our expectation that the structure of the BSC lessens the influence of subordinate likeability on evaluators' performance evaluation judgments. Instead, subordinate likeability was found to significantly influence evaluators' subordinate performance-related judgments regardless of the performance measure format. Additionally, as part of a supplemental analysis, we find that while performance ratings partially mediate the effect of likeability on the bonus allocated to each manager, likeability still has a marginally significant direct effect on the bonus allocation.

Our finding that the influence of subordinate likeability on evaluation judgments persists in a setting with a large number of performance measures that include outcomes and targets, and is not lessened by the structure of the BSC, suggests that the subordinate likeability will matter as long as evaluators have the discretion to attend to and weight performance measures subjectively. This is an important finding because a critical issue facing firms using BSCs is whether to assign explicit weights to each performance measure or to allow subjectivity in weighting each measure (Ittner et al., 2003). Thus, when firms are facing this decision, they should be aware that evaluators, when given discretion to subjectively weight performance measures, are likely to consider subordinate likeability when evaluating performance.

The rest of the chapter is organized as follows. The next section reviews relevant literature. In the subsequent sections we present our hypotheses, describe our research method, and present our results, including supplemental analysis. In the last section of the chapter we then discuss the implications and limitations of our research, and offer suggestions for further research.

BACKGROUND

Affect refers to feeling states and encompasses two broad categories: undifferentiated and differentiated (Park, Sims, & Motowildo, 1987). Undifferentiated affect refers to mood states and is not directed at a particular entity. For example, feelings of joy or sadness are mood states. An evaluator experiencing one of these moods might be expected to inflate or deflate evaluations of any subordinate. Thus, while undifferentiated affect might bias evaluations, this form of bias is expected to affect evaluations of all subordinates equally (Cardy & Dobbins, 1994). Alternatively, differentiated affect constitutes a like–dislike reaction to an individual and has been characterized as "the major currency in which social intercourse is transacted" (Zajonc, 1980). In contrast to undifferentiated affect, this form of affect differentiates among subordinates.

In this study we examine the effect of differentiated affect between two subordinate managers on evaluators' performance-related judgments in a multiple performance evaluation setting. Specifically, one subordinate manager is described using personal attributes indicative of a likeable person whereas the other subordinate manager is described using personal attributes indicative of an unlikeable person. Without sufficient information about the subordinate manager's actual performance on performance relevant outcomes, using information about the manager's likeabilty to base or inform evaluations would be appropriate. That is, likeability is generally positively associated with performance relevant outcomes. Managers possessing likeable personal attributes would generally be expected to outperform managers possessing unlikeable personal attributes on performance relevant outcomes. Consequently, in the absence of actual performance outcomes, evaluators would be expected to consider likeability information relevant to reach a judgment about a subordinate's performance.

Because a BSC contains many performance relevant metrics tied to the company's strategy that is intended to be a "comprehensive reporting system" (Ullrich & Tuttle, 2004), information about manager likeability might be expected to have a limited influence on performance evaluation. To the extent likeability influences performance evaluations because it informs evaluators about unmeasured performance outcomes, the effect of subordinate likeability should diminish as the set of performance measure becomes more complete. However, subordinate likeabilty information may be expected to influence performance evaluations under a BSC even when the set of BSC measures represents a relatively complete set of performance measures. Commonly, no explicit weights are assigned to each measure included in a BSC such that forming an overall performance evaluation is largely subjective. That is, while many, if not all, of the individual measures may be relatively quantifiable and considered "objective," discretion is left on how to combine measures into an overall performance evaluation of subordinate managers. In such a setting, subordinate likeability is expected to influence the attention and subjective weighting of performance measures, which in turn, will effect performance evaluations. Consistent with this view, early research by Alexander and Wilkins (1982) found that subordinate likeability had a greater effect when subjective rather than more objective performance measures were used.

Lipe and Salterio (2002), consistent with their earlier work (Lipe & Salterio, 2000), assert that the use of a large number of performance measures to evaluate performance represents a complex task. This assertion is based on research that has found a positive relationship between task complexity and number of cues or attributes (Bonner, 1994; Wood, 1986). More specifically, Lipe and Salterio (2002, p. 532), state that separately processing, weighing, and combining a large number of performance measures into an overall evaluation "is cognitively, a very difficult thing to do."

Within cognitive psychology, individuals are portrayed as having limited information-processing capacity. For example, research findings indicate that individuals are only able to simultaneously process approximately seven to nine information cues (Miller, 1956). A broad range of research supports this view (Baddeley, 1994). When based on a large number of performance measures, performance evaluations are likely to be subject to cognitive or other types of heuristics that systematically influence the subjective weighting of performance measures. For example, research has generally found that when evaluating two managers, unaided performance ratings reflect a common-measures bias (Banker et al., 2004; Dilla & Steinbart, 2005; Libby et al., 2004; Lipe & Salterio, 2000). That is, a performance measure that is common between two managers is subjectively weighted more heavily than a performance measure that is unique to each manager. However, use of strategy maps (Banker et al., 2004), a "disaggregated BSC" (Roberts et al., 2004), or process accountability or third-party assurance about performance measures (Libby et al., 2004) have been found to mitigate the common-measures bias.

Lipe and Salterio (2002) examine the conditions under which the structure imposed by the BSC (e.g., grouping measures into one of four categories) influences performance evaluation judgments.

When evaluating performance based on a large number of performance measures, which represents a complex task, evaluators will invoke a heuristic such as a divide and conquer information processing approach (Lipe & Salterio, 2002). Under a divide and conquer approach, each of the performance outcomes is assigned to one of a small number of groups. Next, performance is assessed for each group based upon the measures that have been assigned to the group. Because each group contains a relatively small number of measures, evaluation at the group level represents a less cognitively demanding task. Finally, separate group evaluations are combined into an overall evaluation.

Lipe and Salterio (2002, p. 533) contend that when faced with a complex evaluation task, evaluators will invoke a heuristic such as a divide and conquer information processing approach and that "the organization of the BSC lends itself quite naturally to this kind of mental approach." That is, performance measures are each assigned to one of the four categories of the BSC. Whether the structure of the BSC influences overall performance evaluations, however, depends upon the specific pattern of outcomes among performance measures. In particular, Lipe and Salterio (2002) predict that the structure of the BSC will influence performance evaluation judgments when the measures within a BSC category consistently reflect that performance by one subordinate was better than another subordinate. Based on a divide and conquer approach, simultaneously processing a small number of measures that are grouped together should facilitate evaluators' ability to see relations among the measures. Such perception, in turn, will influence the subjects' weighting of the measures. Specifically, two (or more) measures that are perceived to be related will be weighted less than two measures that are perceived to be unrelated. Thus, when all the measures that are consistently favorable towards one subordinate are included within a single BSC category, these measures will be weighted less than when these measures are not grouped together. Alternatively, when the measures that are grouped together do not consistently favor one subordinate, then Lipe and Salterio (2002) contend that the structure of the BSC will not influence overall performance ratings. The results of two experiments provide support for their predictions.

HYPOTHESES

Performance appraisal researchers (Cardy & Dobbins, 1994; Dipboye, 1985; Lefkowitz, 2000; Murphy & Cleveland, 1991; Turban et al., 1990; Varma, Denisi, & Peters, 1996) recognize that performance evaluation is a judgment involving both affective and cognitive aspects. In a seminal chapter, Zajonc (1980, p. 154) contends that affective reaction "is capable of influencing the ensuing cognitive process to a significant degree." That is, affect influences cognitive processes, which in turn, influence performance ratings (Antonioni & Park, 2001). For example, affect may impact what and how a manager observes the work of a subordinate, and consequently, the manager's memories about the subordinate. Evidence consistent with this view is presented by Murphy, Gannett, Herr, and Chen (1986). They report that evaluators' recall of behavioral information tends to be consistent with their general impressions of a subordinate. Additionally, research (Cardy & Dobbins, 1986; Tsui & Barry, 1986; Turban et al., 1990) investigating the role of affect on appraisal judgments generally occurs in a setting without performance criterion (e.g., targets). In such a setting, Baltes and Parker (2000) contend that subordinate affect may induce bias in evaluators' implicit performance targets (e.g., use a relatively lower performance standard for a likeable subordinate and a relatively higher performance standard for an unlikeable subordinate). In the current study, cognitive biases related to observations and recall are designed away and each subordinate has explicit performance targets. Consistent with previous BSC research (Banker et al., 2004; Dilla & Steinbart, 2005; Libby et al., 2004; Lipe & Salterio, 2000, 2002; Roberts et al., 2004), performance is not observed. Instead, specific performance measure targets and outcomes are provided as a basis to assess performance.

Robbins and DeNisi (1994, 1998) integrate affective and cognitive aspects of appraisals by proposing an affect-consistency heuristic. Under an affect-consistency heuristic, evaluators attend to and emphasize information that is consistent with their affective reaction to the subordinate. Thus, based on this heuristic, for a well-liked subordinate, evaluators are expected to seek out and elevate the importance of favorable performance information and discount poor performance information "as not meaningful or an aberration" (Robbins & DeNisi, 1994, p. 343). This tendency reverses for a subordinate who is unlikeable such that evaluators will seek out and elevate the importance of unfavorable performance information and discount favorable performance information. Varma et al. (1996) further hold that evaluators are more likely to apply an affect-consistency heuristic when performance information is ambiguous, as is commonly the case with multidimensional performance measurement. As described, although not recognized by Robbins and DeNisi (1994, 1998), the affect-consistency heuristic may be viewed as a specific application of motivated reasoning

(Kunda, 1990). Kunda (1990, p. 480) proposes that when motivated "to arrive at a particular, directional conclusion" individuals will rely on a "biased set of cognitive processes." Presumably, an evaluator is motivated to evaluate a likeable subordinate favorably and an unlikeable subordinate unfavorably, creating directional conclusions.

Robbins and DeNisi (1994) test their model in the context of undergraduate management students evaluating the performance of their instructors. This task was selected because of its familiarity to undergraduate students. Nine behavioral incidents were developed to reflect performance information for each of the three instructors. Approximately one-third of the incidents represented above-average, average, and below-average levels of performance, respectively. Several weeks earlier, these students had completed a questionnaire to assess interpersonal affect towards each of the three instructors. The results of the study indicate that affect consistency was not associated with the information acquisition stage. However, affect consistency was significantly associated with the weighting stage. As expected, affect consistent incidents were assigned a higher weight than either affect neutral or affect inconsistent incidents. Finally, affect consistency was found to influence overall instructor ratings. Subsequently, Robbins and DeNisi (1998) provide evidence that affect consistency had a stronger influence on the weighing of instructor incidents and instructor ratings than mood consistency. However, Robbins and DeNisi (1994, 1998) did not provide participants with performance benchmarks or targets. Therefore, their results are consistent with an affect-consistency heuristic but are also consistent with participants using differing targets or benchmarks for more or less likeable instructors.

The current study examines the influence of subordinate likeability in an organizational environment where subordinate managers are formally evaluated based on a large number of performance measures. For each performance measure, pre-determined targets and actual outcomes are reported. However, as is commonly the case, evaluators have discretion on how much weight to give to each performance measure. As discussed above, performance evaluation judgments based on a large number of different performance measures represent a complex evaluation task, and consequently, evaluators are likely to rely on heuristics to simplify the task. Our study examines whether the influence of subordinate likeability on evaluators' judgments depends on the organization structure of the performance measurement instrument. Similar to Lipe and Salterio (2002), one performance measurement instrument adopts a BSC framework whereas the other performance measurement instrument presents performance information as one large group (e.g., non-BSC framework).

First, consider a setting in which the performance measurement instrument presents information about the performance of two managers, one likeable and one unlikeable, as one large group. With such an instrument, based on prior appraisal research (Robbins & DeNisi, 1994, 1998) we expect evaluators to simplify their information processing by adopting the affect-consistency heuristic. Under this heuristic evaluators direct their attention and differentially weight those measures that favor a likeable manager. In this context, a performance measure favors a likeable manager when a likeable manager outperforms an unlikeable manager on the performance measure. Thus, evaluators are expected to differentially attend to and place greater weight upon those measures favoring a likeable manager and either ignore or discount those measures that do not favor a likeable manager. By using this heuristic, a likeable manager's performance rating and, consequently, bonus allocation will increase. That is, by selectively weighting performance measures favoring a likeable manager, the likeable manager is expected to receive a higher performance rating, and consequently, a higher bonus allocation. Also, because evaluators are expected to differentially attend to outcome information specifically about a likeable manager relative to the outcome information specifically about an unlikeable manager, evaluators will exhibit stronger conviction in their evaluations about a likeable subordinate under a non-BSC structure.

Alternatively, consider a setting in which the BSC is used as a basis to organize the multiple measures about each of two managers, one likeable and one unlikeable. Given the complexity of the task and structure provided by the BSC, evaluators are expected to invoke a divide and conquer heuristic based on the structure of the BSC to simplify the task and guide their information processing (Lipe & Salterio, 2002). That is, because the format of the BSC naturally disaggregates performance measures among four categories, evaluators are expected to initially evaluate managers within a BSC category. By dividing the overall task into several smaller tasks, evaluators are expected to attend to each of the relatively few performance measures within the category (Lipe & Salterio, 2002), including those that do not favor a likeable manager. Thus, the structure of the BSC, by fostering attention and consideration of all performance measures within and across performance categories, is expected to lessen the influence of manager likeability on performance evaluation judgments. That is, the BSC format is expected to decrease the influence of manager likeability by increasing the likelihood that measures both favorable and unfavorable to a likeable manager are considered by an evaluator. To the extent that the BSC format successfully guides and focuses attention to all performance measures, manager likeability is expected to have a diminished influence on performance-related judgments. This discussion leads to the following hypotheses:

Hypothesis 1(a). The likeability of subordinate managers and performance measure organization will interact to influence overall performance evaluations. Specifically, the influence of subordinate manager likeability on performance ratings will be significantly diminished when a BSC is used compared to when a BSC is not used.

Hypothesis 1(b). The likeability of subordinate managers and performance measure organization will interact to influence strength of conviction in performance evaluations. Specifically, the influence of subordinate manager likeability on strength of conviction will be significantly diminished when a BSC is used compared to when a BSC is not used.

Hypothesis 1(c). The likeability of subordinate managers and performance measure organization will interact to influence bonus allocations. Specifically, the influence of subordinate manager likeability on bonus allocations will be significantly diminished when a BSC is used compared to when a BSC is not used.

METHOD

Overview and Task

The task and experimental materials were based on those developed by Lipe and Salterio (2000, 2002). Participants were presented with a case involving the Women's Clothing Stores, Incorporated (WSC), a retail firm specializing in women's apparel. While WCS is described as having multiple divisions, participants were told that the case focuses on the two largest divisions. In their assigned role as a senior executive of WCS, participants were instructed that their task was to evaluate the performance of the division managers for the two largest divisions. The case indicated that the company has been using multiple performance measures for several years. The case included a discussion of how company strategy and customer attributes were considered in the development of a common set of multiple measures and targets that were established for all divisional managers for the current year. Thus, the strategy, performance measures, and targets were the same for both divisions. This experimental design was used to ensure that the common-measures bias (Lipe & Salterio, 2000) did not confound our results.

Table 1 presents targets and actual performance levels for the multiple performance measures for both division managers being evaluated using the BSC format. As shown, a total of 16 performance measures are included,

Table 1. The Balanced Scorecard Employed in the Experiment.

Measure	Chris Peters – Division A			Taylor Graham – Division B		
	Target	Actual	% Better than target	Target	Actual	% Better than target
Financial						
Return on sales	24%	26%	8.33%	24%	25%	4.17%
New store sales	30%	30%	0.00%	30%	31.2%	4.00%
Sales growth	34%	35.5%	4.41%	34%	36.5%	7.35%
Monetary market share relative to retail space	\$80	\$84.70	5.88%	\$80	\$82.55	3.19%
Customer-related						
Mystery shopper program rating	85	90	5.88%	85	92	8.24%
Repeat sales	25%	27%	8.00%	25%	26%	4.00%
Returns by customers as % of sales	12%	12%	0.00%	12%	11.5%	4.17%
Customer satisfaction rating	84%	86.2%	2.62%	84%	84%	0.00%
Internal business processes						
Returns to suppliers	8%	7.7%	3.75%	8%	7%	12.50%
Average major brand names/store	32	34	6.25%	32	33	3.13%
Average markdowns	20%	18.5%	7.50%	20%	20%	0.00%
Sales from new market leaders	25%	25.6%	2.40%	25%	26.1%	4.40%
Learning and growth						
Stores computerizing	85%	88%	3.53%	85%	86%	1.18%
Hours of employee training/employee	12	13	8.33%	12	12	0.00%
Average tenure of sales personnel	1.4	1.5	7.14%	1.4	1.6	14.29%
Employee suggestions per employee	3.1	3.1	0.00%	3.1	3.2	3.22%

divided equally among the four BSC categories. The case was designed such that each manager outperformed the other manager on two of four measures within each BSC category and that their overall performance based on an equal weighting of each measure was quite similar.

For each division manager, participants made an overall performance rating. Similar to Lipe and Salterio (2000), participants were asked to place an "X" on a 101-point scale anchored by 0 (labeled "reassign") and 100 (labeled "excellent"). Next, participants were asked to indicate their strength of conviction in their overall performance rating. Again, participants were asked to place an "X" on a 101-point scale. For this scale, 0 was labeled "very weak" and 100 was labeled "very strong." After evaluating both managers, participants were asked to allocate a $20,000 bonus between the two divisional managers. The bonus was included to provide evidence on potential economic consequences associated with performance ratings. A similar bonus measure has been included in previous BSC research (Dilla & Steinbart, 2005; Roberts et al., 2004). Lastly, participants completed a debriefing questionnaire that collected demographic and other perceptual information about the task.

Design

The study employed a 2×2 between subjects design. The first between subjects factor manipulated performance measure organization. Under the BSC format condition, the BSC concept was described and four BSC categories were defined. Further, the 16 performance measures were classified into the appropriate BSC category – financial, customer related, internal business processes, and learning and growth. Thus, each BSC category contained four different performance measures. Under the NOFORM condition, the 16 measures were presented without the BSC format. That is, the 16 performance measures were listed under a single grouping. Similar to Lipe and Salterio (2002), under the NOFORM condition, performance measures were listed either alphabetically or based upon a random ordering.[2]

The second between subjects variable manipulated manager likeability. Likeability information about each division manager was provided in addition to a short professional biography describing each division manager's background. A set of five personal attributes was included to convey the likeability of each division manager. These personal attributes were based on work from Anderson (1968).[3] He identified 555 personal attributes that had positive, negative, or no association with the formation of likeability perceptions. The set of five personal attributes used to portray positive likeability included four positive attributes and one neutral attribute from Anderson (1968). The five attributes are cheerful, loyal, tactful, wholesome, and methodical. The last item represents an attribute with neutral likeability implications. The set of five personal attributes used to portray negative likeability included four negative attributes and one neutral attribute from Anderson (1968). The five attributes are boastful, gossipy, self-centered, superficial, and systematic. The last item represents an attribute with neutral likeability implications. Our approach is similar to the approach used by Cardy and Dobbins (1986) and Dobbins and Russell (1986). In each of these studies, subordinate manager likeability was established by providing a set of five personal attributes from Anderson (1968).

Under the "Like Manager A" condition, the set of five positive likeability attributes were used to describe Division Manager A, Chris Peters, and the set of five negative likeability attributes were used to describe Division Manager B, Taylor Graham. Specifically, under this condition, the description read, in part, "In working with Chris over the last two years, you have found him to be loyal, cheerful, methodical, wholesome, and tactful." Also, under this condition, the description read, in part, "In working with Taylor over the last two years, you have found him to be boastful, gossipy, systematic, self-centered, and superficial." Alternatively, under the "Like Manager B" condition, the set of five positive likeability attributes were used to describe Division Manager B and the set of five negative likeability attributes were used to describe Division Manager A.

Dependent Measures

Hypotheses 1(a)–1(c) focus on overall performance rating, strength of conviction, and compensation for the two division managers. To test the effect of subordinate manager likeability on overall performance ratings and strength of conviction in performance ratings a difference score is determined for each judgment and used as a dependent measure. The performance rating difference score is calculated by subtracting the overall performance rating of division manager B from the overall performance rating of division manager A. Thus, a positive value indicates that division manager A was rated higher than division manager B. Alternatively, a negative value indicates that division manager B was rated higher than division manager A. A conviction difference score is calculated by subtracting the strength of conviction score of division manager B from the strength of conviction score of division manager A. Thus, a positive value indicates that the strength of conviction in the rating of division manager A was stronger than the strength of conviction in the rating of division manager B. Alternatively, a negative value indicates that the strength of conviction is stronger in the rating of division manager B than division manager A. Regarding tests of the effect of relative manager likeability on compensation, participants allocated a bonus pool between manager A and manager B. Therefore, the total bonus is constant and manager B's bonus is a function of the bonus allocated to manager A. Consequently, we use the bonus assigned to division manager A as the dependent measure.

Participants

Evening MBA students at a major metropolitan state university were used as participants for the study. The questionnaire was administered after the topic of the BSC had been discussed in the course. One hundred thirty six students enrolled in a managerial accounting course completed the questionnaire.[4] Background information about these participants is presented in Table 2. As shown, the majority of participants were male and had evaluated individuals in the past. The mean age of participants was approximately 29 years, the mean years of work experience was approximately 6 years, and the mean number of accounting classes was approximately four.

RESULTS

Liking Manipulation Check

As part of the debriefing questionnaire, participants rated the likeability of each division manager. Each statement read, in part, "Based on the information presented, please indicate your impression of how likeable ______ is." The first statement referred to the manager of division A and the second statement referred to the manager of division B. The end points on a seven-point scale for each statement were "Completely Unlikeable" (1) and "Completely Likeable" (7).

Table 2. Participant Demographic Data ($N = 136^{a}$).

Gender:	
Male –	105 participants (77.8%)
Female –	30 participants (22.2%)
Any evaluation experience:	
Yes –	93 participants (68.9%)
No –	42 participants (31.1%)
Age:	
Mean = 29.0 years	
Standard deviation = 4.9	
Years of work experience:	
Mean = 6.0 years	
Standard deviation = 5.1	
Number of accounting courses:	
Mean = 3.7 accounting courses	
Standard deviation = 1.9	

[a] Of the 136 participants, one did not respond to some of the demographic information.

A total of 69 participants were assigned to the "Like Manager A" condition. Under the "Like Manager A" condition the mean likeability ratings (standard deviations) of manager A and manager B were 6.0 (0.9) and 3.8 (1.5), respectively. Among the 67 participants assigned to the "Like Manager B" condition, the mean likeability ratings (standard deviations) of manager A and manager B were 3.2 (1.5) and 5.7 (1.0), respectively. A one-way ANOVA was performed using manager likeability, performance measure organization (PMO), and the interaction as independent variables and likeability ratings of manager A as the dependent measure. Manager likeability condition was significant ($F = 176.37$, $p < .0001$) and the means were in the expected direction. Neither PMO nor the interaction were significant. A second one-way analysis of variance (ANOVA) was performed using manager likeability, PMO, and the interaction as independent variables and likeability ratings of manager B as the dependent measure. Manager likeability condition was significant ($F = 71.34$, $p < .0001$) and the means were in the expected direction. Neither PMO nor the interaction was significant. Overall, these results indicate that the manipulation of manager likeability was successful.

Hypothesis Testing

Hypothesis 1(a) predicts that the BSC instrument will lessen the influence of division manager likeability on performance ratings compared to the

NOFORM instrument. A 2×2 analysis of variance (ANOVA) was conducted using the rating difference score as the dependent measure. The independent variables were manager likeability at two levels and performance measure organization at two levels. This hypothesis would be supported by a significant interaction between manager likeability and performance measure organization and the pattern of results were consistent with the BSC format acting as a moderator variable. Descriptive statistics for the rating difference scores are presented in Table 3, Panel A. The ANOVA results (see Table 3, Panel B) indicate that manager likeability ($F = 8.78$, $p < .01$) is significant. As shown in Table 3, Panel A, the rating difference score was higher in the "Like Manager A" condition (Mean = 4.01) than under the "Like Manager B" condition (Mean = –0.58). However, neither the performance measure organization main effect nor the interaction between manager likeability and performance measure organization is significant. Thus, Hypothesis 1(a) is not supported.

Hypothesis 1(b) predicts that the BSC instrument will lessen the influence of division manager likeability on conviction ratings compared to the NOFORM instrument. A 2×2 analysis of variance (ANOVA) was conducted using the conviction difference score as the dependent measure. The independent variables were manager likeability at two levels and performance measure organization at two levels. Descriptive statistics for the conviction difference scores are presented in Table 4, Panel A. The ANOVA results (see Table 4, Panel B) indicate that manager likeability ($F = 11.93$, $p < .001$) is significant. As shown in Table 4, Panel A, the conviction difference score was higher in the "Like Manager A" condition (Mean = 4.49) than under the "Like Manager B" condition (Mean = −1.43). This indicates that strength of conviction is positively associated with manager likeability. However, neither the performance measure organization main effect nor the interaction between manager likeability and performance measure organization is significant. Thus, Hypothesis 1(b) is not supported.

Hypothesis 1(c) predicts that the BSC instrument will lessen the influence of division manager likeability on bonus allocations compared to the NOFORM instrument. A 2×2 analysis of variance (ANOVA) was conducted using the bonus allocated to manager A as the dependent measure. The independent variables were manager likeability at two levels and performance measure organization at two levels. Descriptive statistics for the bonus assigned to manager A are presented in Table 5, Panel A. The ANOVA results (see Table 5, Panel B) indicate that manager likeability is

Table 3. Analysis of Manager Performance Ratings.

Panel A: Rating difference score[a]

Manager Likeability[c]	Performance Measure Organization[b]		
	BSC format	NOFORM format	Overall
Like Manager A	4.77[d]	3.24	4.01
	(9.48)	(9.10)	(9.26)
Like Manager B	−0.15	−1.00	−0.58
	(9.70)	(7.62)	(8.65)
Overall	2.38	1.22	
	(9.83)	(8.60)	

Panel B: ANOVA with rating difference score as the dependent measure[a]

Sources of Variation	SS	Df	MS	F	p[e]
Performance measure organization (PMO)	48.3	1	48.3	0.60	0.44
Manager likeability (ML)	712.6	1	712.6	8.78	<0.01
PMO × ML	4.02	1	4.02	0.05	0.82
Error	10,712.53	132	81.2		

[a]Rating difference score is calculated by subtracting the overall performance rating of division manager B from the overall performance rating of division manager A. A positive value indicates that division manager A was rated higher than division manager B. Alternatively, a negative value indicates that division manager B was rated higher than division manager A.
[b]Performance measure organization is a between subjects factor that was manipulated. In the BSC format condition, the 16 performance measures were classified into the four BSC categories. In the NOFORM format condition, the 16 performance measures were presented without the BSC format.
[c]Manager likeability is a between subjects factor that was manipulated. In the Like Manager A Condition, positive attributes were associated with manager A and negative attributes were associated with manager B. In the Like Manager B condition, negative attributes were associated with manager A and positive attributes were associated with manager B.
[d]Panel values are the mean rating difference scores. Standard deviations are shown in parentheses below the means.
[e]All reported *p*-values are two-tailed.

significant ($F = 12.26$, $p < .001$). As shown in Table 5, Panel A, the bonus assigned to manager A is higher in the "Like Manager A" condition (Mean = $10,684) than in the "Like Manager B" condition (Mean = $9,716). However, neither the main effect for performance measure organization nor the interaction between performance measure organization and manager likeability is significant. Thus, Hypothesis 1(c) is not supported.

Table 4. Analysis of Strength of Conviction in Performance Rating.

Panel A: Conviction difference score[a]

Manager Likeability[c]	Performance Measure Organization[b]		
	BSC format	NOFORM format	Overall
Like Manager A	4.83[d]	4.15	4.49
	(12.42)	(10.73)	(11.54)
Like Manager B	−0.12	−2.71	−1.43
	(7.97)	(7.82)	(7.94)
Overall	2.43	0.42	
	(10.72)	(9.93)	

Panel B: ANOVA with conviction difference score as the dependent measure[a]

Sources of Variation	SS	Df	MS	F	p[e]
Performance measure organization (PMO)	90.6	1	90.6	0.91	0.34
Manager likeability (ML)	1,183.6	1	1,183.6	11.93	<0.01
PMO × ML	30.8	1	30.8	0.31	0.58
Error	13,093.8	132	99.2		

[a]Conviction difference score is calculated by subtracting the strength of conviction in the performance rating of division manager B from the strength of convicion in the performance rating of division manager A. A positive value indicates that the strength of conviction in division manager A's rating was higher than the strength of conviction in division manager B's rating. Alternatively, a negative value indicates that the strength of conviction in division manager B's rating was higher than the strength of conviction in division manager A's rating.
[b]Performance measure organization is a between subjects factor that was manipulated. In the BSC format condition, the 16 performance measures were classified into the four BSC categories. In the NOFORM format condition, the 16 performance measures were presented without the BSC format.
[c]Manager likeability is a between subjects factor that was manipulated. In the Like Manager A Condition, positive attributes were associated with manager A and negative attributes were associated with manager B. In the Like Manager B condition, negative attributes were associated with manager A and positive attributes were associated with manager B.
[d]Panel values are the mean conviction difference score. Standard deviations are shown in parentheses below the means.
[e]All reported p-values are two-tailed.

Supplemental Analysis

In this supplemental analysis, we seek to determine whether rating difference scores mediate the effect of manager likeability on bonus allocations. Mediating variables do not change the relationship between other variables,

Table 5. Analysis of Bonus Allocation.

Panel A: Manager A bonus[a]

Manager Likeability[c]	Performance Measure Organization [b]		
	BSC format	NOFORM format	Overall
Like Manager A	\$10,833[d]	\$10,530	\$10,684
	(1,602)	(1,461)	(1,530)
Like Manager B	\$9,652	\$9,781	\$9,716
	(1,873)	(1,435)	(1,657)
Overall	\$10,260	\$10,161	
	(1,825)	(1,486)	

Panel B: ANOVA with manager A bonus as the dependent measure[a]

Sources of Variation	SS	Df[e]	MS	F	p[f]
Performance measure organization (PMO)	255,293	1	255,293	0.10	0.75
Manager likeability (ML)	31,422,422	1	31,422,422	12.26	<0.01
PMO $\times$ ML	1,583,852	1	1,583,852	0.62	0.43
Error	335,842,389	131	2,563,682		

[a]Participants allocated a \$20,000 bonus between manager A and manager B. Manager A bonus is the amount of the bonus that participants allocated to manager A. Because manager B bonus is a function of manager A bonus, it is not included in these tests.
[b]Performance measure organization is a between subjects factor that was manipulated. In the BSC format condition, the 16 performance measures were classified into the four BSC categories. In the NOFORM format condition, the 16 performance measures were presented without the BSC format.
[c]Manager likeability is a between subjects factor that was manipulated. In the Like Manager A Condition, positive attributes were associated with manager A and negative attributes were associated with manager B. In the Like Manager B condition, negative attributes were associated with manager A and positive attributes were associated with manager B.
[d]Panel values are the mean bonus allocated to manager A. Standard deviations are shown in parentheses below the means.
[e]One participant did not respond to the bonus allocation question. Therefore, only 135 participants are included in this test.
[f]All reported p-values are two-tailed.

but increase one's understanding of the process by which evaluators "transform the predictor or input variables" (Baron & Kenny, 1986, p. 1178). Baron and Kenny (1986) assert that several relationships must hold to demonstrate mediation. In the context of our study, an independent variable (subordinate manager likeability) must be found to have a direct effect on the mediator (rating difference scores) as well as the dependent variable (manager A bonus). Both of these effects have been found. In addition, there must be a direct relationship between the mediator (rating difference score) and the dependent variable (manager A bonus). The correlation between these two variables is positive and highly significant ($r = 0.77$, $p < 0.0001$). Finally, when the mediator (rating difference score) is controlled for, the effect of the independent variable (manager likeability) on the dependent variable (manager A's bonus) should be eliminated or substantially lowered. To test this final aspect of mediation, analysis-of-covariance was conducted with manager A's bonus as the dependent variable and the rating difference score as a covariate in addition to the two independent variables. The results show that the rating difference score is highly significant ($F = 166.4$, $p < 0.001$) and that manager likeability remains marginally significant ($F = 2.95$, $p < 0.09$). The overall pattern of results indicates that rating difference scores partially mediate the relationship between manager likeability and compensation. This suggests that manger likeability has a direct effect on compensation over and above the indirect effect resulting from elevated performance evaluations.

DISCUSSION

The purpose of this study was to examine the influence of subordinate likeability on overall performance ratings and compensation judgments in a setting with a large number of performance measures. While previous appraisal research (Lefkowitz, 2000) has found that subordinate manager likeability influences appraisal-related judgments, research has not explored the existence of this influence in a setting using a large number of performance relevant measures that includes targets. Additionally, previous research has not explored whether the influence of subordinate likeability can be lessened through the use of a more structured performance measurement instrument as provided by the BSC. As expected, we found subordinate likeability influenced evaluators' judgments when the performance measurement instrument was unstructured. Unexpectedly, the influence of subordinate likeability was not lessened when a BSC was used to structure performance measures. Thus, subordinate likeability significantly influenced evaluators' judgments similarly regardless of the structure of the performance measurement instrument.

Before discussing the results of the study, several limitations should be noted. First, the participants in the study were evening MBA students asked to evaluate the overall performance and allocate bonuses for two divisional managers. Generally, participants had substantial professional work experience and the majority had previously evaluated the work of others. Our participants may have limited experience evaluating performance using a large number of performance measures. However, as Lipe and Salterio (2000, p. 296) note, in explaining their decision to employ graduate business students as participants to examine real world phenomenon related to the BSC, "the theory does not suggest an optimal choice," and consequently "it can be difficult to identify appropriate participants for the experiments."

Secondly, the participants in this study did not actually know the subordinates, and thus, subordinate manager likeability was created through the use of descriptions of the subordinate managers. While such an approach has been used by appraisal researchers (Cardy & Dobbins, 1986; Dobbins & Russell, 1986; Robbins & DeNisi, 1994), this may represent a relatively weak "treatment." The advantage of this approach is that it allows one to study the influence of subordinate manager likeability during the evaluation process without being confounded by information an evaluator learned through previous interactions with the subordinate. In addition, by including targets for each performance measure, participants were presumably prevented from creating a favorable benchmark (e.g., creating a relatively low benchmark for a likeable subordinate and creating a relatively high benchmark for an unlikeable subordinate). This feature is in contrast to earlier managerial affect studies (Cardy & Dobbins, 1986; Dobbins & Russell, 1986; Robbins & DeNisi, 1994) in which performance targets were not provided.

Turning to a discussion of the results, first, our findings of significant main effects of subordinate manager likeability on appraisal-related judgments are consistent with previous appraisal research (Cardy & Dobbins, 1986; Tsui & Barry, 1986; Turban et al., 1990). However, our findings extend previous results in several ways. Evaluators in our study always were given a large number of performance measures including targets for each measure. Given the large number of performance measures there is less need to use subordinate likeability to infer missing information because it is available. This result is consistent with an affect-consistency heuristic in which evaluators attend to and weight measures consistent with the subordinate's likeability.

Our findings also demonstrate that subordinate likeability influences evaluators' judgments even when the performance measurement instrument is structured using the BSC. The results did not support our expectation that the structure of the BSC would lessen the influence of subordinate likeability on evaluators' judgments. In considering this unexpected result we speculate that even though evaluators presumably used a divide and conquer heuristic to simplify the evaluation task, and this process was guided by the BSC, evaluators still used an affect-consistency heuristic to direct attention and evaluation among measures within each BSC category. That is, within category, combining measures within the BSC remains a subjective task and subordinate manager likeability may have influenced this weighting process within each category. This suggests that as long as evaluators have discretion to attend to and weight performance measures subjectively, the tendency of subordinate likeability to influence evaluators' judgments is robust and likely to occur.

Our results should be informative for firms using BSCs whose managers must decide whether to assign explicit weights to each measure or to allow subjectivity in weighing each measure. Our results indicate that when evaluators have discretion to subjectively weight each measure, this process is likely to be influenced by subordinate likeability. Given beliefs about the comprehensiveness of the performance measures included in the BSC and beliefs about the relative value of subordinate likeability within the context of the firm's goals and objectives, this type of subjective weighting might be viewed positively or negatively.

NOTES

1. Within accounting, the role of affect on judgment and decision making has been relatively limited (Kadous, 2001; Kida, Moreno, & Smith, 2001; Moreno, Kida, & Smith, 2002). However, Kida and Smith (1995) recognize that affect is likely to influence cognitive processes of decision makers using accounting-based information.

2. Similar to Lipe and Salterio (2002), we included both a random and an alphabetical order to increase the generalizability of the results. For the random order, the order of the measures was chosen at random with the only constraint that adjacent measures should not come from the same BSC category. Within the NOFORM group, no differences were found for participants with the random versus the alphabetical order for any of the outcome measures. Therefore, these two conditions are collapsed together and analyzed as one condition.

3. Previous research investigating the effect of likeability on performance evaluations has also used this source (Cardy & Dobbins, 1986; Dobbins & Russell, 1986). We did not use identical personal attributes as previous research for two reasons. First, these previous articles only gave partial lists of the personal attribute words they used in their manipulations. Second, some of the previous articles used words that, in today's business environment, may have other behavioral implications (e.g., the dislikeable word "greedy" may also invoke unethical implications).

4. One participant did not provide a response to the bonus allocation. Consequently, the analysis of the bonus allocation is based on responses from 135 participants.

REFERENCES

Alexander, E. R., & Wilkins, R. D. (1982). Performance rating validity: The relationship between objective and subjective measures of performance. *Group and Organization Studies*, *7*(4), 485–496.

Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, *9*(3), 272–279.

Antonioni, D., & Park, H. (2001). The relationship between rater affect and three sources of 360-degree feedback ratings. *Journal of Management*, *27*(4), 479–495.

Baddeley, A. (1994). The magical number seven: Still magic after all these years? *Psychological Review*, *101*(2), 353–356.

Baltes, B. B., & Parker, C. P. (2000). Reducing the effects of performance expectations on behavioral ratings. *Organizational Behavior and Human Decision Processes*, *82*(2), 237–267.

Balzer, W. K. (1986). Biases in the recording of performance-related information: The effects of initial impression and centrality of the appraisal task. *Organizational Behavior and Human Decision Processes*, *37*(3), 329–347.

Banker, R. D., Chang, H., & Pizzini, M. J. (2004). The balanced scorecard: Judgmental effects of performance measures linked to strategy. *The Accounting Review*, *79*(1), 1–23.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Bates, R. (2002). Liking and similarity as predictors of multi-source ratings. *Personnel Review*, *31*(5/6), 540–552.

Bonner, S. (1994). A model of the effects of audit task complexity. *Accounting, Organizations and Society*, *19*(3), 213–234.

Cardy, R. L., & Dobbins, G. H. (1986). Affect and appraisal accuracy. Liking as an integral dimension in evaluating performance. *Journal of Applied Psychology*, *71*(4), 672–678.

Cardy, R. L., & Dobbins, G. H. (1994). Performance appraisal: The influence of liking on cognition. *Advances in Managerial Cognition and Organizational Information Processing*, *5*, 115–140.

DeNisi, A. S., Robbins, T. L., & Summers, T. P. (1997). Organization, processing, and use of performance information: A cognitive role for performance instruments. *Journal of Applied Social Psychology*, *27*(21), 1884–1905.

Dilla, W. N., & Steinbart, P. J. (2005). Relative weighting of common and unique balanced scorecard measures by knowledgeable decision makers. *Behavioral Research in Accounting*, *17*, 43–53.

Dipboye, R. L. (1985). Some neglected variables in research on discrimination in appraisal. *Academy of Management Review*, *10*(1), 116–127.

Dobbins, G. H., & Russell, J. M. (1986). The biasing effects of subordinate likeableness on leaders' responses to poor performance: A laboratory and a field study. *Personnel Psychology*, *39*(4), 759–777.

Ittner, C. D., Larcker, D. F., & Meyer, M. W. (2003). Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The Accounting Review*, *78*(3), 725–758.

Kadous, K. (2001). Improving jurors' evaluations of auditors in negligence cases. *Contemporary Accounting Research*, *18*(3), 425–444.

Kaplan, R., & Norton, D. (1992). The balanced scorecard – measures that drive performance. *Harvard Business Review*, *70*(1), 71–79.

Kaplan, R., & Norton, D. (1996a). *The balanced scorecard*. Boston, MA: Harvard Business School Press.

Kaplan, R., & Norton, D. (1996b). Using the balanced scorecard as a strategic management system. *Harvard Business Review*, *74*(1), 75–85.

Kida, T. E., Moreno, K. K., & Smith, J. F. (2001). The influence of affect on managers' capital-budgeting decisions. *Contemporary Accounting Research*, *18*(3), 477–494.

Kida, T. E., & Smith, J. F. (1995). The encoding and retrieval of numerical data for decision making in accounting contexts: Model development. *Accounting, Organizations, and Society*, *20*(7/8), 585–610.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

Lefkowitz, J. (2000). The role of interpersonal affective regard in supervisory performance ratings: A literature review and proposed causal model. *Journal of Occupational and Organizational Psychology*, *73*(1), 67–85.

Libby, T., Salterio, S. E., & Webb, A. (2004). The balanced scorecard: The effects of assurance and process accountability on managerial judgment. *The Accounting Review*, *79*(4), 1075–1094.

Lipe, M. G., & Salterio, S. (2000). The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review*, *75*(3), 283–298.

Lipe, M. G., & Salterio, S. (2002). A note on the judgmental effects of the balanced scorecard's information organization. *Accounting, Organization and Society*, *27*(6), 531–540.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Moreno, K., Kida, T., & Smith, J. F. (2002). The impact of affective reactions on risky decision making in accounting contexts. *Journal of Accounting Research*, *40*(5), 1331–1349.

Murphy, K. R., & Cleveland, J. N. (1991). *Performance appraisal: An organizational perspective*. Boston, MA: Allyn and Bacon.

Murphy, K. R., Gannett, B. A., Herr, B. M., & Chen, J. A. (1986). Effects of subsequent performance on evaluations of previous performance. *Journal of Applied Psychology*, *71*(3), 427–431.

Park, O. S., Sims, H. P., & Motowildo, S. J. (1987). Affect in organizations. In: H. P. Sims & D. A. Gioia (Eds), *The thinking organization* (pp. 215–237). San Francisco: Josey-Bass.

Robbins, T. L., & DeNisi, A. S. (1994). Interpersonal affect and cognitive processing in performance appraisal: Towards closing the gap. *Journal of Applied Psychology*, *79*(3), 341–350.

Robbins, T. L., & DeNisi, A. S. (1998). Mood vs. interpersonal affect: Identifying process and rating distortions in performance appraisal. *Journal of Business and Psychology*, *12*(3), 313–325.

Roberts, M. L., Albright, T. L., & Hibbets, A. R. (2004). Debiasing balanced scorecard evaluations. *Behavioral Research in Accounting*, *16*, 75–88.

Shanteau, J. (1988). Psychological characteristics and strategies of expert decision makers. *Acta Psychologica*, *68*(1–3), 203–215.

Tsui, A. S., & Barry, B. (1986). Interpersonal affect and rating errors. *Academy of Management Journal*, *29*(3), 586–599.

Turban, D. B., Jones, A. P., & Rozelle, R. M. (1990). Influences of supervisor liking of a subordinate and the reward context on the treatment and evaluation of that subordinate. *Motivation and Emotion*, *14*(3), 215–233.

Ullrich, M. J., & Tuttle, B. M. (2004). The effects of comprehensive information reporting systems and economic incentives on managers' time-planning decisions. *Behavioral Research in Accounting*, *16*, 89–105.

Varma, A., Denisi, A. S., & Peters, L. H. (1996). Interpersonal affect and performance appraisal. *Personnel Psychology*, *49*(2), 341–360.

Wong-On-Wing, B., Guo, L., Li, W., & Yang, D. (2007). Reducing conflict in balanced scorecard evaluations. *Accounting, Organizations and Society*, *32*(4–5), 363–377.

Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, *37*(1), 60–82.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151–175.

This page intentionally left blank

THE MODERATING EFFECT OF MANAGER'S ETHICAL JUDGMENT ON THE RELATIONSHIP BETWEEN BUDGET PARTICIPATION AND BUDGET SLACK

Adam S. Maiga and Fred A. Jacobs

ABSTRACT

This study tests the moderating effects of manager's ethical judgment on the relationship between budget participation and budget slack. To this end, we developed and mailed a questionnaire to 251 managers at different divisions of manufacturing firms. Overall, the results suggest that manager's ethical judgment moderates the relationship between budget participation and budget slack.

INTRODUCTION

Subsequent to the failure of companies such as Enron, WorldCom, Global Crossings, HealthSouth, and others, the importance of ethical behavior has again resurfaced in the business community and society in general. Due to the enormous losses suffered by debt and equity markets, employees, and all

Advances in Accounting, Volume 23, 113–145

ISSN: 0882-6110/doi:10.1016/S0882-6110(07)23005-6

other stakeholders, most of the ethical issues being raised are related to external financial reporting and other related disclosures. However, little attention has been given to losses to a company, which are inevitably passed on to all stakeholders, related to suboptimal allocation of resources due, in part, to the misrepresentation of information within the firm through the occurrences of budget slack. Budget slack is created by lower level managers to exploit private information through the introduction of slack, which is the amount by which managers overstate their needs for resources to complete a task or understate their productive capability when given the opportunity to influence the standard against which their performance will be evaluated (Schiff & Lewin, 1970). Similarly, managers who misrepresent private information regarding resource needs or production capacity may receive excess resources that can be diverted to perquisite consumption (Waller & Bishop, 1990). This opportunistic use of private information is commonly cited as an ethical issue because slack creation may be inconsistent with role-related norms and desired virtues of professional managers and accountants. Furthermore, the resource misallocation that results is detrimental to other organizational units, to investors, and other stakeholders (Douglas & Wier, 2000). Thus, creation of budgetary slack is an ethical dilemma, a predicament with a moral component on the part of the decision-maker (Douglas & Wier, 2000, 2005).

Agency theory is based upon the assumptions of economic rationality of all contracting parties within the firm and that resulting behaviors of agents will reflect self-interest. When agents have private information and are able to conceal that information from their superiors, they may misrepresent that information to maximize their own utility functions. Building in budget slack by agents, according to agency theory, is one form of increasing agency costs because decisions regarding resource allocations can become suboptimal because these decisions are based on incorrect information.

Prior studies suggest that ethical considerations may moderate agency effects. Agency theory, however, is one theory of human behavior among many theories that have been posited in extant literature. Experimental studies on slack creation decisions (Young, 1985; Waller, 1988; Chow, Cooper, & Waller, 1988; Chow, Cooper, & Haddad, 1991) and project evaluation decisions (Harrison & Harrell, 1993; Harrell & Harrison, 1994) have found that some subjects violate classical agency predictions. For example, Young (1985) found that social pressure reduced slack creation. The literature also suggests that factors described as ''aversion to lying'' (Chow et al., 1991), ''personal integrity'' (Chow et al., 1988), and ''ethical considerations'' (Noreen, 1988) may potentially mitigate self-interested behavior. Recent accounting studies that specifically incorporate ethical reasoning into an agency design find that moral reasoning attenuates narrow self-interest behavior – in a project escalation decision (Rutledge & Karim, 1999), and in a budgetary slack creation situation (Stevens, 1998). Luft (1997) and Ghosh (2000) argue that in examining accounting-related behavior, it is important to consider ethics in terms of perceptions of equity (or fairness).

The main objective of this research is to determine if business unit[1] managers use their ethical judgment to evaluate budgetary slack creation as being positive or negative. The importance of this question could help to determine how much organizations have to be concerned with managers' ethical behavior regarding budgetary slack creation. To this end, we gathered information on budgetary process from a large sample of managers in their actual organizational setting, using four scenarios. First, we assess managers' ethical judgment using constructs of "moral equity," "contractualism," and "relativism." Next, we investigate the moderating effects of managers' ethical judgment and budget participation on budget slack. Overall, except for the moderating effects of moral equity (scenario A) and relativism (scenario B), the results show that, under each scenario, the ethical judgment measures have a moderating effect on budget participation to significantly reduce budget slack.

The remainder of the chapter is organized into four sections. The next section provides the research background. The third section offers the hypotheses for study. The fourth section discusses the research methods. In the fifth section, the statistical results are presented. The final section discusses the implications of the results.

BACKGROUND

Several studies have examined the relationship between budget participation and budgetary slack. One stream of research has proposed that budget participation can bring about goal congruence. For example, Onsi (1973) proposed that budget participation decreases slack since participation leads to positive communication, making a manager feel he is not under pressure to create slack. Communication can reduce willingness to miscommunicate by reducing "risk." Cammann (1976) reported that budget participation reduced manager's "defensive responses" such as the creation of budgetary slack. Both Onsi (1973) and Merchant (1985) found significant negative correlations between budget participation and slack.

A second stream of research is based on the ''beyond-budgeting'' process which explores ways of using budgets more flexibly and finding ways of overcoming the gaming and creation of slack (Horngren, Datar, & Foster, 2003). After conducting research at several organizations that fully or partly abandoned their budgeting, the Beyond Budgeting Round Table (BBRT) developed a generic model that is based on 12 principles (de Waal, 2005). The first six principles concern creating a flexible organizational structure in which authority is delegated to employees, while principles seven to twelve deal with designing an adaptive management process for a flexible organizational structure (de Waal, 2005). However, a closer look shows that the rationale of an application of the 12 principles is contingent upon numerous implicit premises (Schaffer & Zyder, 2003). For example, Schaffer and Zyder (2003) argue that Hope and Fraser's (2003) assumption of an age ''of discontinuous change, unpredictable competition, and fickle customers'' will not be relevant to the same degree for all companies (and business units).

A third stream of research has explored the issue of budget slack using an agency perspective. If the manager (agent) has private information about local conditions, participative budgets allow the manager to choose how to ''signal'' that information to the principal (Baiman, 1982; Baiman & Evans, 1983; Magee, 1980); and if the manager perceives that organizational rewards are dependent upon attaining the budget, he/she may withhold or misrepresent his private information to gain a more attainable budget and increase the likelihood of favorable performance evaluations (Dunk, 1993; Waller, 1988). Therefore, participation also provides the manager greater opportunity to introduce slack into the budget (Dunk, 1993; Lukka, 1988; Young, 1985).

A fourth stream of research views budgetary slack creation as an ethical issue (Nouri & Parker, 1998).[2] When managers misrepresent their capabilities, they therefore fail to disclose to their superiors all their information and informed insights and, commonly, they actually present a distorted picture of the possibilities that may lead to resource misallocation. Thus, creation of budgetary slack becomes an ethical dilemma (Merchant, 1985).

Whether budget participation increases or decreases budget slack depends upon whether managers choose to primarily pursue self-interest or organizational interests.[3] This chapter argues that while self-interest may be a powerfully motivating force for some managers, another important motivating force for others is organizational interest. Managers with strong ethical judgment may be motivated to pursue organizational interests, even when organizational interests conflict to some degree with self-interest.

Although significant insights into managers' cognitive moral capacity have been realized (e.g., Jeffrey & Weatherholt, 1996; Lampe & Finn, 1992; Ponemon & Gabhart, 1993; Shaub, 1994; Sweeney, 1995), it remains to be determined to what extent managers actually use their cognitive moral capacity in the resolution of ethical dilemmas encountered in the work place. Hence, in the case of participatory budgets, whether managers seek to reduce budgetary slack to aid organizational planning and coordination or create slack is still an open research question.

HYPOTHESES DEVELOPMENT

The objective of this study is to integrate prior research and to offer new empirical evidence on the moderating effects of managers' ethical judgment on the relationship between budget participation and budget slack. The overall theoretical model is illustrated in Fig. 1.

Moral Equity

The philosophy of moral equity is based on the overall concept of fairness and justice and has been very influential in contemporary moral thought. Dees (1992) observes that "society expects conformance to social norms, including honesty, trustworthiness, fairness, justice, a sense of public duty, respect for the autonomy of others, and avoidance of gratuitous harm." Jones (1991) suggests that concerns for ethics are jointly determined by

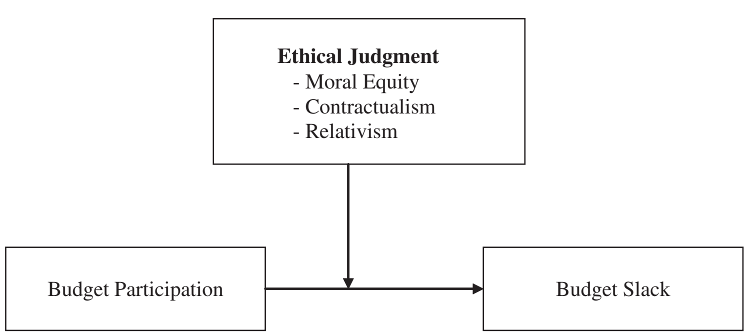


Fig. 1. Model Showing Ethical Judgment Measures as Moderating the Relationship between Budget Participation and Budget Slack.

characteristics of the situation and the individual. Ethical concerns typically arise in situations where self-interest conflicts with a moral duty to others (Bowie & Duska, 1990). To the extent that the subordinate has internalized social norms for honesty and fairness, and budgetary slack is deemed to be inconsistent with these norms, the budgeting task will arouse ethical concerns in the subordinate (Blanchette, Pilot, & Cadieux, 2002). Therefore, opportunistic behavior on the part of managers may be controlled in part by their ethical concerns (see Arrow, 1985; Baiman, 1990).

The above discussion indicates that it is reasonable to assume that managers' judgment will be influenced by ethical considerations, such as their perceived morality and fairness of their actions. Hence, in a participative budget setting, if morality were conceived as internalized social pressure or expectations, then those who perceive an action unfair to the firm and its shareholders would be less likely to take that action because these ethical concerns will motivate the subordinate to conform to the internalized norms, thereby motivating the subordinate to reduce the amount of slack in the budget (Stevens, 1998, p. 9). Therefore, we expect managers' moral equity is to moderate the relationship between their level of budget participation and budget slack. Therefore,

H1. Moral equity moderates the relationship between budget participation and budget slack such that, in higher moral equity, the budget slack effect of budget participation is reduced.

Contractualism

Inherent in the ethical evaluation of an exchange process appears to be the idea of implicit contract and promise (Reidenbach & Robin, 1991). All the company's relationships with its stockholders may be treated as though they were contractual relationships (see, e.g., Donaldson & Dunfee, 1995). However, as Kaptein and Wempe (1998) suggest, in many cases, an unwritten moral contract may be broader than that has been written. The moral obligations of the contracting partners then spring from this implicit moral contract (Kaptein & Wempe, 1998). This argument is supported by the stockholder theory that holds that managers in a corporation have a normative obligation to maximize profits, since this provides the greatest long-term value to the stockholders (Smith, 2002). Therefore, managers are ethically prohibited from investing in initiatives that benefit parties other than the stockholders unless those initiatives are, in the end, the best investments of capital that are available (Bowie & Freeman, 1992). Hence, we posit that manager's ethical action, as measured by contractualism, the implicit moral contract, will moderate the relationship between his level of budget participation and his propensity to create slack. More specifically,

H2. Contractualism moderates the relationship between budget participation and budget slack such that, in higher contractualism, the budget slack effect of budget participation is reduced.

Relativism

Ethical relativism is the thesis that ethical principles or judgments are relative to the individual or culture (Lafollette, 1991). Since relativism bases judgment of the acceptability of an action on cultural or social norms, prior studies suggest that what is traditionally or culturally acceptable appears also to play an evaluative role in the ethical decision-making process (Ferrell & Gresham, 1985). This is in support of Trevino (1986) who acknowledges the impact of culture on the ethical behavior of managers. Additionally, Reidenbach and Robin (1991) acknowledge a strong interaction between culture and tradition and the notion of right and wrong.

Hunt and Vitell's (1986) findings suggest that beliefs about what is culturally and traditionally acceptable play a direct role in the evaluative process. In a non-prescriptive study of professional responsibilities, Gaa (1990) showed that the opportunities for professionals to act in their own self-interest and disregard their responsibilities allow structural instability in the relationship between professionals and society. However, the process of internalizing ethical sanctions may be an opportunity to stabilize the "social contract" between the accounting profession and society referred to by Gaa (1986, 1990). Therefore, ethical concerns can be determined by the individual's value system, which evolves from internalized social norms.

The above discussion suggests that budgetary slack, with its potential to mislead the principal and transfer resources to the subordinate, is likely to raise traditional and socio-cultural concerns within managers. Hence, we expect the proposed effects of ethical judgment based on managers' traditional and cultural norms, as measured by relativism, to moderate the impact of budget participation on budget slack. Therefore,

H3. Relativism moderates the relationship between budget participation and budget slack such that, in higher relativism, the budget slack effect of budget participation is reduced.

RESEARCH METHOD

Sample

A questionnaire was administered to a sample of managers (plant managers, manufacturing managers, operations managers, marketing managers, research managers, distribution managers) from manufacturing companies in U.S.A. We selected a mail survey approach for use in this study (1) because of the ego-involving character of some of the questions and the halo effects that often occur in response to ethics questions, and (2) in the hope that it would substantially reduce the problem of "socially acceptable responses." Providing anonymity to the respondent also tends to dampen these effects (Flory, Phillips, Reidenbach, & Robin, 1992). In using the mail approach, we sacrificed response rate in favor of improving honesty of responses. We anticipated that the nature of the ethical issues investigated could produce more socially acceptable responses. Thus, because of the anonymity associated with the questionnaire response, the mail approach satisfied our concerns about the validity of the responses. Criteria used to select the participants included: (1) each participant should have budget responsibility in the subunit; and (2) each unit would be an investment center. We obtained a mailing list for this study from the *Industry Week* series on manufacturing excellence from which a random sample of 650 names was selected. A cover letter explained the purpose of the study with an exhortation for participation and cooperation. A copy of the questionnaire used in the study appears in the appendix.

In the first three weeks, 167 questionnaires were returned; that was followed by a second mailing resulting in 56 new responses. However, of the total of 223 returned questionnaires, only 193 were usable. In an attempt to increase the number of respondents, a random sample of 150 non-respondents was contacted by telephone; that resulted in a return of 78 questionnaires of which 58 were usable.[4] Overall, this data collection led to 251 usable responses.[5]

Measurement and Validation of Variables

This study uses budgetary slack as the dependent variable, budget participation as the independent variable and three surrogate measurements for ethical judgment as moderator variables: moral equity, contractualism, and relativism. The measures and the scenarios appear in the appendix along with a description of questionnaire construction.

Budgetary slack: Measuring the actual *budgetary slack* within an organization's budget is extremely difficult. Thus, this study used "propensity to create slack," a self-reported measure, as a surrogate measure under the assumption that actual slack and the manager's propensity to create slack are highly correlated. Propensity to create slack is operationalized using the three-item scale found in Kren (1993) and adapted from Merchant (1985). Merchant's original four-item scale was examined by Hughes and Kwon (1990) who suggested deleting one item to improve the scale's reliability. Thus this study uses the three items suggested by Hughes and Kwon (1990). The response scale was a seven-point Likert-type scale ranging from 1 (strongly disagree) to 7 (strongly agree). To examine the extent to which these measures are interrelated, we used principal component analysis with varimax rotation, which produced one factor with total variance of 88.208% and an eigenvalue greater than one. A reliability check for the measures produced a Cronbach alpha of 0.928, indicating that the measures were reliable (Nunnally, 1967).

Budget participation was measured using the Milani's (1975) six-item measure. The response scale was a seven-point Likert-type scale ranging from 1 (strongly disagree) to 7 (strongly agree). A principal component analysis with varimax rotation produced one factor with total variance of 86.154% and an eigenvalue greater than one. A reliability check for the measures produced a Cronbach alpha of 0.905, indicating that the measures were reliable.

Ethical judgment: The questionnaire contains a reduced set of measures developed by Reidenbach and Robin (1991). The measure focuses on the dynamics of decision-making regarding managers' ethical judgment. It consists of eight bipolar scales divided into three dimensions – moral equity, relativism, and contractualism (see appendix). In a subsequent study, Flory et al. (1992) used four of the five scenarios (see appendix) that the IMA Resources Center developed. This provided a useful step in developing a measure of ethical judgment because they portray substantially more involved, realistic situations. Each scenario includes an action statement to assure that all respondents were reacting to the same stimulus. The action statement was particularly necessary with the more complex situations described in this research. Consequently, the four scenarios are used in this study.

Each scenario portrays a different kind of ethical dilemma. Scenarios A and D describe actions that might not be perceived as explicitly ethical or unethical, while scenarios B and C feature what most would label as definitely unethical behavior. Scenario A describes a superior who is making questionable expenditures that he claims meet upper management's approval. The manager, who may find himself in a marketing environment different from his background, is faced with establishing the proper lines of authority in connection with an issue that may not be a violation of company policy.

Scenario B involves a controller who is asked to falsify external financial statements for the purpose of procuring additional working capital. Although this may actually happen in some companies, managers typically agree that falsification of external statements is wrong. This is also true of the specific violations of company policy shown in scenario C. A difference in scenario C, besides the fact that it is an internal situation, is that the manager had previously violated company policy, and now, in an attempt to rectify a resulting failure, decides to violate the policy again. Scenarios A, B, and C all implicitly involve a manager's job security, but in each situation the managers are seemingly concerned with their company's welfare; in contrast, scenario D emphasizes the manager's personal problems. In this scenario, company policy is not clearly delineated, and there could be some uncertainty whether the manager's action is unethical. The additional background information provided in scenario D allows the respondent to empathize with the manager's personal difficulties, although it is unclear whether his personal situation has any bearing on his decision.

To examine the extent to which the ethical judgment measures under each scenario are interrelated, we used factor analysis with principal component analysis and with varimax rotation to determine whether the items for each scenario could be grouped according to studies by Flory et al. (1992) and Reidenbach and Robin (1991) (see Table 1). Three factors with eigenvalues greater than one emerged from the analysis for each scenario, with their corresponding varimax rotation factor solution retaining at least 67.31% of the total variance in the data. The factor solutions for the defined constructs support the construct validity of the survey instrument. Convergent validity is demonstrated by each factor having multiple-question loadings in excess of 0.50. In addition, discriminant validity is supported, since none of the questions in the factor analysis have loadings in excess of 0.40 on more than one factor (see Table 1). Also, the Cronbach alphas shown in Table 2 suggest that the measures from the factor analysis are reliable. Therefore, the loadings are comparable to Flory et al. (1992) and Reidenbach and Robin (1991).

Table 1. Factor Structure for the Four Scenarios.

Dimension	Scenario				Scenario				Scenario			
	A	B	C	D	A	B	C	D	A	B	C	D
Moral equity												
Fair/unfair	**0.859**	**0.873**	**0.799**	**0.884**	−0.054	0.143	0.170	−0.017	0.033	0.019	−0.028	0.088
Just/unjust	**0.925**	**0.920**	**0.886**	**0.851**	0.072	−0.168	0.047	−0.029	−0.049	−0.022	−0.015	0.158
Moraly right/not morally right	**0.903**	**0.885**	**0.774**	**0.799**	0.056	−0.002	−0.073	0.065	0.035	0.026	−0.036	−0.114
Acceptable/unacceptable to famility	**0.680**	**0.620**	**0.706**	**0.686**	−0.109	−0.089	0.039	0.163	−0.187	−0.062	−0.101	−0.137
Contractualism												
Violates/does not violate promise	−0.127	0.026	0.162	0.178	**0.884**	**0.874**	**0.905**	**0.904**	0.057	0.081	−0.146	−0.039
Violates/does not violate contract	−0.268	−0.155	−0.039	−0.097	**0.864**	**0.853**	**0.934**	**0.840**	−0.054	0.116	0.139	0.271
Relativism												
Traditionally acceptable/unacceptable	0.178	0.115	−0.162	0.019	0.008	0.256	0.021	−0.020	**0.803**	**0.711**	**0.802**	**0.858**
Culturally acceptable/unacceptable	0.107	−0.032	0.044	−0.027	−0.071	−0.045	−0.023	0.215	**0.851**	**0.890**	**0.846**	**0.762**

Note: Values in bold indicates appropriate item loadings on corresponding factors.

Table 2. Reliability Coefficients (Coefficient Alpha).

Budget Participation	0.905			
Budget Slack	0.928			
	Scenario			
	A	B	C	D
Moral equity	0.873	0.852	0.808	0.825
Contractualism	0.639	0.715	0.828	0.733
Relativism	0.558	0.512	0.546	0.546

To assess the content validity of the scales, we followed Flory et al. (1992) to test whether our constructs in fact measure manager's ethical judgment. We compared the three dimensions with manager's ethical intention in response to each scenario which was measured by the response opportunities on a seven-point bipolar scale range from 1 (ethical) to 7 (unethical). The results of a common validation procedure, based on regression analysis, in the social sciences appear in Table 3. A high covariation (R^2) between ethical intention and ethical judgment measures suggests that the ethical judgment captures much of what the respondents mean by "ethical." Additionally, the individual beta values for each of the three dimensions, and in each of the scenarios, also helped to define the concept of "ethics" for the respondents. As Table 3 indicates, the three dimensions explain from 56.70% to 80.80% of the variance in what the managers defined as ethical, suggesting that the ethical judgment measures capture much of what the respondents mean by "ethical."

Next, we assessed predictive validity of the scales. The behavioral intention of the respondent in response to each scenario was measured by the statement, "If you were responsible for making the decision described in the scenario, what is the probability that you would make the same decision?" The response opportunities were reported on a seven-point bipolar scale range from highly probable to highly improbable. The relevant R^2s, shown in Table 4, from "predicting" this measure with the multivariate ethics scale range from 0.771 to 0.890, satisfying the expectations for predictive validity.

In addition, control for common method bias was accomplished in two primary ways: the design of the study's procedures (procedural remedies) and statistical controls (statistical remedies). The design of the study's procedures consists of (1) assuring respondents of anonymity (Podsakoff,

Table 3. A Comparison of the Ethical Judgment Measures and the Ethical Intention Measure.

Scenario	Regression Results			
	Overall	Moral equity	Contractualism	Relativism
	R^2	β1	β2	β3
A	0.808	0.649	0.302	0.525
B	0.567	0.497	0.498	0.477
C	0.667	0.407	0.535	0.456
D	0.704	0.529	0.454	0.334

Table 4. The Relationship Between the Ethical Judgment Measures and the Behavioral Intention Measure.

Scenario	Regression Results			
	Overall	Moral equity	Contractualism	Relativism
	R^2	β1	β2	β3
A	0.878	0.557	0.516	0.491
B	0.885	0.527	0.491	0.473
C	0.771	0.539	0.527	0.440
D	0.890	0.520	0.507	0.465

MacKenzie, & Lee, 2003), (2) careful construction of the variable constructs, and (3) counterbalancing the question order (Podsakoff et al., 2003). This approach is known to have the effect of neutralizing some of the method biases that affect the retrieval stage by controlling the retrieval cues prompted by the question context. The statistical control consists of the use of Harman's (1976) one-factor test (discussed in the results section).

We performed Harman's (1976) one-factor test to assess common method variance; this is one of the techniques used most by researchers to address the issue of common method variance (Campbell & Fiske, 1982; Greene & Organ, 1973; Schriesheim, 1979; Organ & Greene, 1981; Fiske, 1982; Anderson & Bateman, 1997; Aulakh & Genctruk, 2000). Under this method, if common method variance were a serious problem in the study, we would expect a single factor to emerge from a factor analysis or one general factor to account for most of the covariance in the variables

(Podsakoff & Organ, 1986). Accordingly, we perform a factor analysis on all items under each scenario, extracting four factors with eigenvalues greater than one, and all items loaded on their theoretical construct and had loadings greater than 0.60, indicating that each item was well reflective of the underlying construct, and that common method variance is not an issue.

Research Model and Testing Procedures

To provide measurements for the hierarchical regression model used to test the hypotheses, we computed the average for the six responses for budget participation, the average for the responses for each of the ethical judgment dimension under each scenario, and the average for the three responses for budget slack. Hypotheses 1, 2, and 3 (above) posit a moderating effect of manager's ethical judgment on the relationship between budget participation and budget slack. To insure that the relationships are significant, a hierarchical regression analysis was used for each scenario. Budget slack was regressed on budget participation and ethical judgment variables in the first step. In the second step, the interaction of each ethical judgment criteria, "moral equity," "contractualism," and "relativism" with budget participation was entered in the regression to determine their impact on the base model containing only the independent variables. Based on this approach, the following regression models were employed to test the hypotheses:

$$\text{B-Slack}_i = \text{BP} + \text{ME}_i + \text{CT}_i + \text{RL}_i + \varepsilon \tag{1}$$

$$\begin{aligned}\text{B-Slack}_i = \text{BP} + \text{ME}_i + \text{CT}_i + \text{RL}_i + (\text{ME}_i \times \text{BP}) \\ + (\text{CT}_i \times \text{BP}) + (\text{RL}_i \times \text{BP}) + \varepsilon\end{aligned} \tag{2}$$

where B-Slack$_i$ is the propensity to create slack, BP the budget participation, ME the moral equity, CT the contractualism, RL the relativism, i the scenarios 1, 2, 3, and 4, and ε the error term.

RESULTS

The results are presented in two parts. First, we present the descriptive statistics. Next, we explain the results from the regression models used to test the hypotheses.

Descriptive Statistics

Table 5 reports the data for the study. The data was obtained from 251 managers, from different companies, in the manufacturing industry with the following job titles: 49 plant managers, 32 manufacturing managers, 48 operations managers, 43 marketing managers, 41 research managers, and 38 distribution managers. Table 5 also provides the profile of the responding companies that constitute a broad spectrum of manufacturers as defined by the 2-digit SIC codes. The classification by the primary 2-digit SIC code place the respondents in the electronic and other electric equipment industry (38), instruments and related products (22), chemical and allied products (48), fabricated metal (37), primary metal industries (39), food and kindred products (28), paper and allied products (18), and apparel and other fabricated textile products (21).

In Table 6, the mean values of the variables used to test the hypotheses denote that many respondents indicated some probability of engaging in the activity specified in the scenarios and their level of budget participation. Additional information on respondents' characteristics is provided in Table 6. The respondents to the question regarding number of years with the division had a mean of 9.14 years in their current position. To the number of years in management question, respondents indicated a mean of 13.12 years. The results also show that the average number of employees equals 241. For the 194 divisions that provided sales figures, the mean was \$5.324 million.

Hypotheses Tests

We constructed a hierarchical regression model for the dependent variable. In the first step, we entered only the independent variables. In the second step, the interaction terms were added. Standard scores[6] are used for the independent variables in order to provide a clearer basis to interpret signs of the interaction coefficients in the second step (Brownell & Hirst, 1986) and to minimize multicolinearity between main and cross-product effects (Cronbach, 1987). Tolerance greater than 0.10 was achieved. Variance inflation factor values from the regression analyses conducted for all the variables were less than 2, which is lower than the guideline of 10 (Hair, Anderson, Tatham, & Black, 1995). Hence, multicolinearity does not appear to be a problem. Statistical interpretation of the results followed the approach adopted by Lau, Low, and Eggleton (1995) and Jaccard, Turrisi,

Table 5. Characteristics of Respondents.

Industry	Plant Manager	Manufacturing Manager	Operating Manager	Marketing Manager	Research Manager	Distribution Manager	Total
Electronic and other electric equipment	10	7	5	3	9	4	38
Instruments and related products	3	4	7	6	0	2	22
Chemical and allied products	12	6	7	11	7	5	48
Fabricated metal	9	4	8	7	2	7	37
Primary metal industries	3	5	9	9	10	3	39
Food and kindred products	7	3	4	4	6	4	28
Paper and allied products	0	2	5	3	3	5	18
Apparel and other fabricated textile products	5	1	3	0	4	8	21
Total	49	32	48	43	41	38	251

Table 6. Variable Means and Standard Deviations.

	Mean	Standard Deviation
Size (Number of employees)	240.841	148.897
Years at division	9.143	8.999
Years in management position	13.116	9.035
Net sales (millions – $)	5.524	1.339
Budget participation	3.890	1.529
Budget slack	2.685	1.392
Ethical judgment		
Scenario A		
Moral equity	4.651	1.582
Contractualism	3.414	1.321
Relativism	4.295	1.250
Scenario B		
Moral equity	4.786	1.527
Contractualism	4.803	1.414
Relativism	4.823	1.274
Scenario C		
Moral equity	4.950	1.421
Contractualism	4.572	1.444
Relativism	4.357	1.282
Scenario D		
Moral equity	4.880	1.428
Contractualism	4.645	1.406
Relativism	4.761	1.26
Ethical measure and behavioral intention measure		
Scenario A		
Ethical intention measure	5.332	0.889
Behavioral intention measure	5.489	0.844
Scenario B		
Ethical intention measure	5.772	0.790
Behavioral intention measure	5.812	0.810
Scenario C		
Ethical intention measure	5.539	0.863
Behavioral intention measure	5.687	0.765
Scenario D		
Ethical intention measure	5.577	0.905
Behavioral intention measure	5.720	0.806

and Wan (1990).The moderating effects of the three ethical judgment measures (i.e., moral equity, contractualism, and relativism) on the relationship between budget participation and budget slack are presented in Table 7.

Table 7. Regression Analyses.

	Eq (1) Standardized Beta	*t*	Sig.	Eq (2) Standardized Beta	*t*	Sig.	Eq (1) Standardized Beta	*t*	Sig.	Eq (2) Standardized Beta	*t*	Sig.
Scenario A							*Scenario B*					
ME	−0.092	−1.570	0.118	−0.104	−1.863	0.064	−0.027	−0.449	0.653	−0.008	−0.141	0.888
CT	0.288	4.964	0.000	0.273	4.802	0.000	0.195	3.178	0.002	0.214	3.666	0.000
RL	0.163	2.869	0.004	0.107	1.972	0.050	0.083	1.356	0.176	0.098	1.671	0.096
BP	−0.338	−5.806	0.000	−0.335	−5.750	0.000	−0.295	−4.900	0.000	−0.272	−4.445	0.000
ME × BP				−0.093	−1.611	0.108				−0.170	−2.847	0.005
CT × BP				−0.231	−4.176	0.000				−0.229	−3.929	0.000
RL × BP				−0.190	−3.436	0.001				−0.072	−1.243	0.215
R^2			0.210			0.309			0.154			0.246
R^2-change						0.099						0.092
F-value					15.535	<.0001					11.342	<.0001
Scenario C							*Scenario D*					
ME	−0.016	−0.257	0.797	0.006	0.097	0.923	−0.024	−0.412	0.681	−0.046	−0.847	0.398
CT	0.177	2.957	0.003	0.183	3.213	0.001	0.237	4.138	0.000	0.270	4.930	0.000
RL	0.007	0.118	0.906	−0.007	−0.122	0.903	0.244	4.287	0.000	0.228	4.252	0.000
BP	−0.317	−5.212	0.000	−0.280	−4.567	0.000	−0.319	−5.573	0.000	−0.302	−5.248	0.000
ME × BP				−0.163	−2.706	0.007				−0.119	−2.131	0.034
CT × BP				−0.183	−3.209	0.002				−0.254	−4.712	0.000
RL × BP				−0.194	−3.386	0.001				−0.125	−2.369	0.019
R^2			0.132			0.234			0.241			0.342
R^2-change						0.102						0.101
F-value					10.610	<.0001					18.049	<.0001

Notes: ME, Moral equity; CT, Contractualism; RL, Relativism; BP, Budget participation.

Scenario A in Table 7 assesses the moderating effect of the ethical judgment variables on the relationship between manager's budget participation and budget slack. Results in Table 7, Eq. (2), show that, overall the interactions between ethical judgment variables and budget participation are significant and negative ($F = 15.535$, $p < 0.0001$, R^2-change $= 0.099$). The model explains 30.90% of the variance. The regression results also indicate that budget slack is a significant negative function of the interaction between moral equity and budget participation (ME $\times$ BP) ($t = -1.611$, $p = 0.108$), contractualism and budget particiapation (CT $\times$ BP) ($t = -4.176$, $p = 0.000$), and relativism and budget participation (RL $\times$ BP) ($t = -3.436$, $p = 0.001$).

Next, scenario B is examined. Results in Table 7, Eq. (2), show that, overall the interactions between ethical judgment variables and budget participation are significant and negative ($F = 11.342$, $p < 0.0001$, R^2-change $= 0.092$). The model explains 24.6% of the variance. The regression results also indicate that budget slack is a significant negative function of the interaction between moral equity and budget participation (ME $\times$ BP) ($t = -2.847$, $p = 0.005$), contractualism and budget particiapation (CT $\times$ BP) ($t = -3.929$, $p = 0.000$). However, although the interaction effects of relativism and budget participation (RL $\times$ BP) is negative, it was not found significant ($t = -1.243$, $p = 0.215$).

Scenario C also assesses the moderating effect of the ethical judgment variables on the relationship between manager's budget participation and budget slack. Results in Table 7, Eq. (2), show that, overall the interactions between ethical judgment variables and budget participation are significant and negative ($F = 10.610$, $p < 0.0001$, R^2-change $= 0.102$). The model explains 23.40% of the variance. The regression results also indicate that budget slack is a significant negative function of the interaction between moral equity and budget participation (ME $\times$ BP) ($t = -2.706$, $p = 0.007$), contractualism and budget participation (CT $\times$ BP) ($t = -3.209$, $p = 0.002$), and relativism and budget participation (RL $\times$ BP) ($t = -3.386$, $p < 0.001$).

Finally, scenario D in Table 7 shows that, overall the interactions between ethical judgment variables and budget participation are significant and negative ($F = 18.049$, $p = 0.0001$, R^2-change $= 0.101$). The model explains 34.20% of the variance. The regression results also indicate that budget slack is a significant negative function of the interaction between moral equity and budget participation (ME $\times$ BP) ($t = -2.131$, $p = 0.034$), contractualism and budget participation (CT $\times$ BP) ($t = -4.712$, $p = 0.000$), and relativism and budget participation (RL $\times$ BP) ($t = -2.369$, $p = 0.019$).

Overall, as reported in Table 7, except for the moderating effects of relativism in scenario B, the results show that, under each scenario, the ethical judgment measures moderate the relationship between budget participation and budget slack, with significant negative impacts. Therefore, both H1 and H2 are supported, while we have found substantial support for H3 in three of the four scenarios.

SUMMARY AND CONCLUSION

There is a growing body of research that suggests that considerations of managers' ethical judgment may play a role in budgetary setting (Luft, 1997; Douglas & Wier, 2000). The research presented in this chapter tests the moderating effect of multidimensional measures of manager's ethical judgment on the relationship between budget participation and budget slack. Experienced managers responded to four hypothetical scenarios related to their ethical judgments that are expected to moderate the relationship between budget participation and budget slack. Overall, the results suggest that the relationship between budgetary participation and budgetary slack is moderated by manager's ethical judgment. However, it was surprising that, although they are in the predicted direction, the moderating effects of moral equity (scenario A) and relativism (scenario B) were not significant. These counter-intuitive findings suggest that our understanding of the interactive effects of moral equity and budget participation (scenario A) and of relativism and budget participation (scenario B) are poorly understood or may in part be an artifact of our design, and deserves further investigation.

The findings have several potential implications for research and for organizations that seek to minimize budgetary slack. For example, organizations might consider strategies that are designed to increase both the ethical judgment and budget participation of their employees. The multidimensional measure is a potentially useful tool for directing research on why managers make certain ethical judgment. Knowing the bases on which managers tend to make ethical judgment can be useful in practical ways; the information can be used to establish ethical norms within an organization (e.g., through ethical training) and to identify managers whose behavior may present potential problems.

In light of prior studies (e.g., Luft, 1997), the results of this study suggest that organizational goals may take precedence over self-interest for those employees ethically sensitive to organizational goals and values. That is, high level of ethical judgment or more effective ways to sensitize managers to the moral problems of budgeting may reduce some of the agency problems encountered in budgeting. Overall, the results of this study are consistent with the findings that greater perceived accountability pressure are associated with more ethical response in the business setting (Weigold & Schlenker, 1991; Tetlock, 1985; Tetlock, Skitka, & Boettger, 1989). The results are also consistent with an increase in perceptions of personal responsibility for actions as a means to avoid moral disengagement (Ashton, Kleinmuntz, Sullivan, & Tomassini, 1988), and with the growing body of evidence that suggests that equity and fairness may play a role in utility functions, and may attenuate some agency theory predictions (Luft, 1997).

However, continued application of the scale used in this study to other business situations with different groups with different characteristics (e.g., age, gender, education, and race), and extensions with other methodologies may improve our understanding of ethical judgment (Flory et al., 1992). Also, there are many budgeting avenues, on the scale of traditional budgeting to beyond budgeting, to modernize the budgeting process (de Waal, 2005). Therefore, situations other than those used in this study may elicit different and perhaps undiscovered dimensions, which may stimulate future research. For example, research has shown that the more beyond-budgeting principles an organization implements, the better it performs (Fraser & de Waal, 2001). However, empirical research using U.S. samples is lacking. Therefore, research is warranted in the context of U.S. firms.

Despite its limitations, the current study demonstrates that ethical judgment is an important variable in the relationship between budget participation and budgetary slack. Ethical judgment may be an important factor to be considered in future research involving budget behavior or organizational control.

NOTES

1. The term business unit is used to refer to a self-contained sub-unit (e.g., division) of a larger corporation.
2. This ethical issue, although suggested in agency theory is not included in the agency model.
3. It may be that behaving according to one's sense of ethical conduct is in the self-interest of the manager. This form of "utility" to "self" is excluded from the agency model, which assumes that "economic" rewards are uniquely in the self-interest of managers. Nevertheless, it is included in classical "utility theory."
4. Because of contravening company policy, some preferred not to participate.
5. We used discriminant analysis to compare the groups of respondents, the early and late respondents (Fowler, 1993). Results revealed that the two groups did not differ significantly in either the level of the variables or in the relationship between the variables at the 0.05 level. This suggests that non-response bias may not be a problem.

6. The independent variables were transformed into new measurement variables with a mean of 0 and standard deviation of 1.

ACKNOWLEDGMENTS

We benefited greatly from the helpful guidance and comments from Govind Iyer (associate editor), and two anonymous reviewers.

REFERENCES

Anderson, L. M., & Bateman, T. S. (1997). Cynicism in the workplace: Some causes and effects. *Journal of Organizational Behavior*, *18*, 449–469.

Arrow, K. J. (1985). The economics of agency. In: J. W. Pratt & R. J. Zeckhauser (Eds), *Principals and agents: The structure of business* (pp. 37–51). Boston, MA: Harvard Business School Press.

Ashton, R., Kleinmuntz, D., Sullivan, J., & Tomassini, L. (1988). Audit decision making. In: A. R. Abdel-khalid & I. Solomon (Eds), *Research opportunities in auditing: The second decade* (pp. 95–132). Sarasota, FL: The American Accounting Association.

Aulakh, P. S., & Genctruk, E. F. (2000). International principal-agent relationships – control, governance and performance. *Industrial Marketing Management*, *29*, 521–538.

Baiman, S. (1982). Agency research in managerial accounting: A survey. *Journal of Accounting Literature*, 154–213.

Baiman, S. (1990). Agency research in managerial accounting: A second look. *Accounting, Organizations and Society*, *15*(4), 341–371.

Baiman, S., & Evans, J. H., III. (1983). Pre-decision information and participative management control systems. *Journal of Accounting Research*, 371–395.

Blanchette, D., Pilot, C., & Cadieux, J. (2002). *Managers' evaluation of budgetary slack creation*. Working paper.

Bowie, N. E., & Duska, R. F. (1990). *Business ethics* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Bowie, N. E., & Freeman, R. E. (1992). *Ethics and agency theory: An introduction*. New York, NY: Oxford University Press.

Brownell, P., & Hirst, M. (1986). Reliance on accounting information, budgetary participation, and task uncertainty: Tests of a three-way interaction. *Journal of Accounting Research*, *24*, 241–249.

Cammann, C. (1976). Effects of the use of control systems. *Accounting, Organizations and Society*, *1*, 301–313.

Campbell, D. T., & Fiske, D. (1982). Convergent and discriminant validation by the multitrait-multimethod matrice. *Psychological Bulletin*, *56*, 81–105.

Chow, C., Cooper, J., & Waller, W. (1988). Participative budgeting: Effects of a truth-inducing pay scheme and information asymmetry on slack and performance. *The Accounting Review*, 111–122.

Chow, C. W., Cooper, J. C., & Haddad, K. (1991). The effects of pay schemes and ratchets on budgetary slack and performance: A multiperiod experiment. *Accounting, Organizations and Society*, *15*, 47–60.

Cronbach, L. J. (1987). Statistical tests for moderator variables: Flaws in analyses recently proposed. *Psychological Bulletin*, *102*, 414–417.

Dees, J. (1992). Principals, agents, and ethics. In: N. E. Bowie & R. E. Freeman (Eds), *Ethics and agency theory: An introduction* (pp. 25–58). New York, NY: Oxford University Press.

Donaldson, T., & Dunfee, T. W. (1995). Contractarian business ethics: Current status and next steps. *Business Ethics Quarterly*, *5*(2), 173–186.

Douglas, P. C., & Wier, B. (2000). Integrating ethical dimensions into a model of budgetary slack creation. *Journal of Business Ethics*, *28*(3), 267–277.

Douglas, P. C., & Wier, B. (2005). Cultural and ethical effects in budgeting systems: A comparison of US and Chinese managers. *Journal of Business Ethics*, *60*(2), 159–174.

Dunk, A. (1993). The effect of budget emphasis and information asymmetry on the relation between budgetary participation and slack. *The Accounting Review*, 400–410.

Ferrell, O. C., & Gresham, L. G. (1985). A contingency framework for understanding ethical decision making in marketing. *Journal of Marketing*, *49*(3), 87–96.

Fiske, D. W. (1982). Convergent-discriminant validation in measurements and research strategies. In: D. Brinberg & L. H. Kidder (Eds), *Forms of validity in research* (pp. 77–92). San Francisco: Jossey-Bass.

Flory, S., Phillips, T., Reidenbach, E., & Robin, D. (1992). A multidimensional analysis of selected ethical issues in accounting. *The Accounting Review*, *67*, 284–302.

Fowler, F. J., Jr. (1993). *Survey research methods*. Newbury Park, CA: Sage.

Fraser, R., & de Waal, A. A. (2001). *BBRT benchmarking project: Report on an exploratory research*. CAM-I's Beyond Budgeting Round Table, London.

Gaa, J. (1986). User primacy in corporate financial reporting: A social contract approach. *The Accounting Review*, *61*, 435–454.

Gaa, J. (1990). A Game-theoretic analysis of professional rights and responsibilities. *Journal of Business Ethics*, *9*, 159–169.

Ghosh, D. (2000). Organizational design and manipulative behavior: Evidence form a negotiated transfer pricing experiment. *Behavioral Research in Accounting*, *12*, 1–30.

Greene, C. N., & Organ, D. W. (1973). An evaluation of causal models linking the received role with job satisfaction. *Administrative Science Quarterly*, *18*, 95–103.

Hair, J. F., Jr., Anderson, R. E., Tatham, R. J., & Black, W. D. (1995). *Multivariate data analysis and readings*. Englewood Cliffs, NJ: Prentice Hall.

Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago: University of Chicago Press.

Harrell, A., & Harrison, P. (1994). An incentive to shirk, privately held information, and managers' project evaluation decisions. *Accounting, Organizations and Society*, *19*(7), 569–577.

Harrison, P., & Harrell, A. (1993). The impact of adverse selection on project continuation decisions. *Academy of Management Journal*, *36*(3), 635–643.

Hope, J., & Fraser, R. (2003). *Beyond budgeting: How manager can break free from the Annual performance trap*. Cambridge, MA: Harvard Business Review Press.

Horngren, C. T., Datar, SM., & Foster, G. (2003). *Cost accounting: A managerial emphasis* (11th ed.). Prentice Hall.

Hughes, M. A., & Kwon, S. Y. (1990). An integrative framework for theory construction and testing. *Accounting, Organizations and Society*, *15*(3), 179–191.

Hunt, S. D., & Vitell, S. J. (1986). A general theory of marketing ethics. *Journal of Macromarketing*, *6*, 5–16.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression*. CA: Sage.

Jeffrey, C., & Weatherholt, N. (1996). Ethical development, professional commitment, and rule observance attitudes: A study of CPAs and corporate accountants. *Behavioral Research in Accounting*, 8–31.

Jones, T. (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of Management Review*, *16*, 366–395.

Kaptein, M., & Wempe, J. (1998). Twelve Gordian knots when developing an organizational code of ethics. *Journal of Business Ethics*, *17*(8), 853–869.

Kren, L. (1993). Control system effect on budget slack. *Advances in Management Accounting*, *2*, 109–118.

LaFollette, H. (1991). The truth in ethical relativism. *Journal of Social Philosophy*, 146–154.

Lampe, J., & Finn, D. (1992). A model of auditors' ethical decision processes. *Auditing: A Journal of Practice & Theory*, *11*, 33–59.

Lau, C. M., Low, L. C., & Eggleton, R. C. (1995). The impact of reliance on accounting measures on job related tension and managerial performance: Additional evidence. *Accounting, Organizations and Society*, *19*(2), 359–381.

Luft, J. (1997). Fairness, ethics and the effect of management accounting on transaction costs. *Journal of Management Accounting Research*, *9*, 199–216.

Lukka, K. (1988). Budgetary biasing in organizations: Theoretical framework and empirical evidence. *Accounting, Organization and Society*, *13*, 281–301.

Magee, R. P. (1980). Equilibria in budget participation. *Journal of Accounting Research*, 551–573.

Merchant, K. (1985). Budgeting and the propensity to create budgetary slack. *Accounting, Organizations and Society*, *10*, 201–210.

Milani, K. (1975). The relationship of participation in budget-setting to industrial supervisor performance and attitudes: A field study. *The Accounting Review*, *50*(2), 274–284.

Noreen, E. (1988). The economics of ethics: A new perspective on agency theory. *Accounting, Organizations, and Society*, *13*, 359–369.

Nouri, H., & Parker, R. J. (1998). The relationship between participation and job performance: The roles of budget adequacy and organizational commitment. *Accounting, Organizations and Society*, *25*(5–6), 74–91.

Nunnally, J. (1967). *Psychometric methods*. New York: McGraw-Hill.

Onsi, M. (1973). Factor analysis of behavioral variables affecting budgetary slack. *The Accounting Review*, 535–548.

Organ, D. W., & Greene, C. N. (1981). The effects of formalization on professional involvement: A compensatory process approach. *Administrative Science Quarterly*, *26*, 237–252.

Podsakoff, P. M., MacKenzie, S. B., & Lee, J. Y. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, *12*(4), 531–544.

Ponemon, L., & Gabhart, D. (1993). *Ethical reasoning in accounting and auditing*. Vancouver, BC: Canadian General Accountants' Research Foundation.

Reidenbach, R., & Robin, D. P. (1991). A conceptual model of corporate moral development. *Journal of Business Ethics*, *10*(4), 273–284.

Rutledge, R., & Karim, K. (1999). The influence of self-interest and ethical considerations on managers' evaluation judgments. *Accounting, Organizations and Society*, *24*, 173–184.
Schaffer, U., & Zyder, M. (2003). *Beyond budgeting: A new management fashion*. Research paper No. 5.
Schiff, M., & Lewin, A. Y. (1970). The impact of people on budgets. *The Accounting Review*, 259–268.
Schriesheim, C. A. (1979). The similarity of individual-directed and group-directed leader behavior descriptions. *Academy of Management Journal*, *22*, 345–355.
Shaub, M. (1994). An analysis of factors affecting the cognitive moral development of auditors and auditing students. *Journal of Accounting Education*, 1–26.
Smith, J. H. (2002). Ethics and information systems: Resolving the quandaries. *ACM SIGMIS Database*, *33*(3), 8–22.
Stevens, D. E. (1998). *Determinants of budgetary slack in the laboratory: An application and extension of agency theory*. Unpublished working paper, University of New Hampshire.
Sweeney, J. (1995). The moral expertise of auditors: An exploratory analysis. In: L. Ponemon (Ed.), *Research on accounting ethics*. Greenwich, CT: JAI Press.
Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly*, *48*, 227–236.
Tetlock, P. E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability: Conformity, complexity, and bolstering. *Journal of Personality and Social Psychology*, *57*, 632–640.
Trevino, L. K. (1986). Ethical decision making in organizations: A person-situation interactionits model. *Academy of Management Review*, *11*(3), 601–617.
de Waal, A. A. (2005). Insights from practice: Is your organization ready for beyond budgeting? *Measuring Business Excellence*, *9*(2), 56–67.
Waller, W. S. (1988). Slack in participating budgeting: The joint effect of a truth-inducing pay scheme and risk preferences. *Accounting, Organizations and Society*, *13*, 87–98.
Waller, W. S., & Bishop, R. A. (1990). An experimental study of incentive pay schemes, communication and intrafirm resource allocation. *The Accounting Review*, *65*(4), 812–836.
Weigold, M. F., & Schlenker, B. R. (1991). Accountability and risk-taking. *Personality and Social Psychology Bulletin*, *17*, 25–29.
Young, S. M. (1985). Participative budgeting: The effects of risk aversion and asymmetric information on budgetary slack. *Journal of Accounting Research*, 829–842.

APPENDIX

PART I

SCENARIOS (IMA Resources Center, Reidenbach & Robin, 1991; Flory et al., 1992)

The following four scenarios were used in this study. Each scenario appeared on a separate page followed by brief instructions, a randomized presentation of the scales, the univariate ethics measure, and the behavioral intention measure as shown below.

Scenario A

Tom Waterman is a young management accountant at a large, diversified company. After some experience in accounting at headquarters, he has been transferred to one of the company's recently acquired divisions run by its previous owner and president, Howard Heller. Howard has been retained as vice president of this new division, and Tom is his accountant. With a marketing background and a practice of calling his own shots Howard seems to play by a different set of rules than those to which Tom is accustomed. So far it is working, as earnings are up and sales projections are high.

The main area of concern to Tom is Howard's expense reports. Howard's boss, the division president, approves the expense reports without review, and expects Tom to check the details and work out any discrepancies with Howard. After a series of large and questionable expense reports, Tom challenges Howard directly about charges to the company for typing that Howard's wife did at home. Although company policy prohibits such charges, Howard's boss again signed off on the expense. Tom feels uncomfortable with this and tells Howard that he is considering taking the matter to the Board Audit Committee for review. Howard reacts sharply, reminding Tom that "the Board will back me anyway" and that Tom's position in the company would be in jeopardy.

ACTION: Tom decides not to report the expense charge to the Audit Committee. Please evaluate this action of Tom Waterman.

MORAL EQUITY DIMENSION

Fair	___	___	___	___	___	___	___	Unfair
Just	___	___	___	___	___	___	___	Unjust
Morally right	___	___	___	___	___	___	___	Not morally right
Acceptable to my family	___	___	___	___	___	___	___	Unacceptable to my family

RELATIVISM DIMENSION

Culturally acceptable	___	___	___	___	___	___	___	Culturally unacceptable
Traditionally acceptable	___	___	___	___	___	___	___	Traditionally unacceptable

CONTRACTUALISM DIMENSION

Does not violate an unwritten contract	___ ___ ___ ___ ___ ___ ___	Violates an unwritten contract
Violates an unspoken promise	___ ___ ___ ___ ___ ___ ___	Does not violate an unspoken promise

ETHICAL INTENTION MEASURE

If you were responsible for making the decision described in the scenario, how would you judge the decision?

ethical ___ ___ ___ ___ ___ ___ ___ unethical

BEHAVIORAL INTENTION MEASURE

If you were responsible for making the decision described in the scenario, what is the probability that you would make the same decision?

highly probable ___ ___ ___ ___ ___ ___ ___ highly improbable

Scenario B

Anne Devereaux, company controller, is told by the chief financial officer that in an executive committee meeting the CEO told them that the company "has to meet its earnings forecast, is in need of working capital and that's final." Unfortunately, Anne does not see how additional working capital can be raised even through increased borrowing, since income is well below the forecast sent to the bank. Seth suggests that Anne review bad debt expense for possible reduction and holding sales open longer at the end of the month. He also brushes off the management letter request from the outside auditors to write down the spare parts inventory to reflect its "true value."

At home on the weekend, Anne discusses the situation with her husband, Larry, a senior manager of another company in town. "They're asking me to manipulate the books," she says. "On the one hand," she complains, "I'm supposed to be the conscience of the company and on the other, I'm supposed to be absolutely loyal." Larry tells her that companies do this all the time, and when business picks up again she'll be covered. He reminds her how important her salary is to help maintain their comfortable lifestyle, and that she shouldn't do anything drastic that might cause her to lose her job.

ACTION: Anne decides to go along with the suggestions proposed by her boss. Please evaluate this action of Anne Devereaux.

MORAL EQUITY DIMENSION

Fair	___ ___ ___ ___ ___ ___ ___	Unfair
Just	___ ___ ___ ___ ___ ___ ___	Unjust
Morally right	___ ___ ___ ___ ___ ___ ___	Not morally right
Acceptable to my family	___ ___ ___ ___ ___ ___ ___	Unacceptable to my family

RELATIVISM DIMENSION

Culturally acceptable	___ ___ ___ ___ ___ ___ ___	Culturally unacceptable
Traditionally acceptable	___ ___ ___ ___ ___ ___ ___	Traditionally unacceptable

CONTRACTUALISM DIMENSION

Does not violate an	___ ___ ___ ___ ___ ___ ___	Violates an
unwritten contract	___ ___ ___ ___ ___ ___ ___	unwritten contract
Violates an	___ ___ ___ ___ ___ ___ ___	Does not violate an
unspoken promise	___ ___ ___ ___ ___ ___ ___	unspoken promise

ETHICAL INTENTION MEASURE

If you were responsible for making the decision described in the scenario, how would you judge the decision?

Ethical ___ ___ ___ ___ ___ ___ ___ Unethical

BEHAVIORAL INTENTION MEASURE

If you were responsible for making the decision described in the scenario, what is the probability that you would make the same decision?

Highly probable ___ ___ ___ ___ ___ ___ ___ Highly improbable

Scenario C

Drew Isler, the plant's chief accountant, is having a friendly conversation with Leo Sullivan, operations manager an old college buddy, and Fred LaPlante, the sales manager. Leo tells Drew that the plant needs a new computer system to increase operating efficiency. Fred interjects that with the increased efficiency and decreased late deliveries their plant will be the top plant next year.

However, Leo wants to bypass the company policy which requires that items greater than $5,000 receive prior Board approval and be capitalized. Leo would prefer to generate purchase orders for each component part of the system, each being under the $5,000 limit, and thereby avoid the approval "hassle." Drew knows this is clearly wrong from a company and an accounting standpoint, and he says so. Nevertheless, he eventually says that he will go along.

Six months later the new computer system has not lived up to its expectations. Drew indicates to Fred that he is really worried about the problems with the computer, and the auditors will disclose how the purchase was handled in the upcoming visit. Fred acknowledges the situation by saying that production and sales are down and his sales representatives are also upset. Leo wants to correct the problems by upgrading the system (and increasing the expenses), and urges Drew to "hang in there."

ACTION: Feeling certain that the system will fail without the upgrade, Drew agrees to approve the additional expense. Please evaluate this action of Drew Isler.

MORAL EQUITY DIMENSION

Fair ___ ___ ___ ___ ___ ___ ___ Unfair
Just ___ ___ ___ ___ ___ ___ ___ Unjust
Morally right ___ ___ ___ ___ ___ ___ ___ Not morally right
Acceptable to my family ___ ___ ___ ___ ___ ___ ___ Unacceptable to my family

RELATIVISM DIMENSION

Culturally acceptable ___ ___ ___ ___ ___ ___ ___ Culturally unacceptable
Traditionally acceptable ___ ___ ___ ___ ___ ___ ___ Traditionally unacceptable

CONTRACTUALISM DIMENSION

Does not violate an	___ ___ ___ ___ ___ ___ ___	Violates an
unwritten contract	___ ___ ___ ___ ___ ___ ___	unwritten contract
Violates an	___ ___ ___ ___ ___ ___ ___	Does not violate an
unspoken promise	___ ___ ___ ___ ___ ___ ___	unspoken promise

ETHICAL INTENTION MEASURE

If you were responsible for making the decision described in the scenario, how would you judge the decision?

Ethical ___ ___ ___ ___ ___ ___ ___ Unethical

BEHAVIORAL INTENTION MEASURE

If you were responsible for making the decision described in the scenario, what is the probability that you would make the same decision?

Highly probable ___ ___ ___ ___ ___ ___ ___ Highly improbable

Scenario D

Paul Tate is the assistant controller at Stern Electronics, a medium-sized manufacturer of electrical equipment. Paul is in his late fifties and plans to retire soon. His daughter has been accepted into medical school, and financial concerns are weighing heavily on his mind. Paul's boss is out of the office recuperating from health problems, and in his absence Paul is making all decisions for the department.

Paul receives a phone call from an old friend requesting a sizable amount of equipment on credit for his new business. Paul is sympathetic but cognizant of the risk of extending credit to a new company, especially under Stern's strict credit policy for such transactions. When Paul mentions this conversation to Warren, the general manager, he is immediately interested. Warren notes that the company needs an additional $250,000 in sales to meet the quarterly budget and, thus, ensure bonuses for management, including Paul.

ACTION: Paul decides to make the sale to his friend's new business. Please evaluate this action of Paul Tate.

MORAL EQUITY DIMENSION

Fair	___ ___ ___ ___ ___ ___ ___	Unfair
Just	___ ___ ___ ___ ___ ___ ___	Unjust
Morally right	___ ___ ___ ___ ___ ___ ___	Not morally right
Acceptable to my family	___ ___ ___ ___ ___ ___ ___	Unacceptable to my family

RELATIVISM DIMENSION

Culturally acceptable	___ ___ ___ ___ ___ ___ ___	Culturally unacceptable
Traditionally acceptable	___ ___ ___ ___ ___ ___ ___	Traditionally unacceptable

CONTRACTUALISM DIMENSION

Violates an	___ ___ ___ ___ ___ ___ ___	Does not violate an
unwritten contract	___ ___ ___ ___ ___ ___ ___	unwritten contract
Violates an	___ ___ ___ ___ ___ ___ ___	Does not violate an
unspoken promise	___ ___ ___ ___ ___ ___ ___	unspoken promise

ETHICAL INTENTION MEASURE

If you were responsible for making the decision described in the scenario, how would you judge the decision?

Ethical ___ ___ ___ ___ ___ ___ ___ Unethical

BEHAVIORAL INTENTION MEASURE

If you were responsible for making the decision described in the scenario, what is the probability that you would make the same decision?

Highly probable ___ ___ ___ ___ ___ ___ ___ Highly improbable

PART II

Please answer the following:

1. What is the number of employees at your company?_________
2. What is your approximate dollar volume of sales?_________
3. Please provide your 2-digit SIC-code_________

PART III

Please provide the following information for the person completing the questionnaire:

1. What is your present job title?_________
2. Number of years at this position?_________
3. Number of years in management_________

PART IV

1. Is your division an investment center?______Yes______No
2. Do you have a budget responsibility in your division? ______Yes ______No

PART V

If you answer to both 1 and 2 in Part IV is **yes**, please answer the remaining parts of the questionnaire, otherwise stop at Part IV and return the questionnaire.

PARTICIPATION
(response anchors: 1 = strongly disagree, 2 = moderately disagree, 3 = mildly disagree, 4 = neutral, 5 = mildly agree, 6 = moderately agree, 7 = strongly agree)
1. I am involved in setting all of my budget
2. My superior clearly explains budget revisions
3. I have frequent budget-related discussions with my superior
4. I have a great deal of influence on my final budget

5. My contribution to the budget is very important
6. My superior initiates frequent budget discussions when the budget is being prepared

PROPENSITY TO CREATE SLACK

(response anchors: 1 = strongly disagree, 2 = moderately disagree, 3 = mildly disagree, 4 = neutral, 5 = mildly agree, 6 = moderately agree, 7 = strongly agree)

1. To protect himself, a manager submits a budget that can safely be attained
2. In good business times, your superior is willing to accept a reasonable level of slack in the budget
3. Slack in the budget is good to do things that cannot be officially approved

This page intentionally left blank

AN EXAMINATION OF FIRST CALL'S COMPANY ISSUED GUIDANCE DATABASE

Lynn Rees and Rebecca Wynalda

ABSTRACT

This paper examines the accuracy and consistency of coding variables in the Company Issued Guidance (CIG) database issued by First Call. Specifically, we examine "CIG Code" – a variable used by First Call to describe the type of guidance issued by management (point, range, or qualitative), and "CIG Description Code" – a variable indicating the direction of the earnings guidance (positive, negative, or neutral). Our examination reveals extensive misclassification errors for these variables. In particular, the errors for CIG description code are of such a magnitude that we recommend researchers not to rely on this variable when employing the First Call's CIG database.

1. INTRODUCTION

In this paper, we provide descriptive evidence on the accuracy and consistency of certain coding variables provided in the Company Issued

Advances in Accounting, Volume 23, 147–177

ISSN: 0882-6110/doi:10.1016/S0882-6110(07)23006-8

Guidance (CIG) database issued by First Call. The CIG database makes available in a machine-readable format a rich sample of management earnings forecasts and/or company earnings preannouncements and is used extensively in empirical research. Since the database was made available in the late 1990s, the accounting literature has seen a proliferation of studies that examine the value-relevance and information content of management earnings disclosures that precede formal earnings announcements.

The specific variables we examine as listed on the CIG database are "CIG Code" and "CIG Description Code". CIG Code is a variable used by First Call to describe the type of guidance issued by management. For example, the management forecast could be a point or range estimate, or several other types of qualitative guidance. CIG Description Code is used to signify the nature of the information contained in the management guidance based on how it compares to the existing analysts' consensus forecast at the time of the guidance. A management earnings forecast that is higher (lower) than the consensus forecast from analysts qualifies as positive (negative) news for the company; whereas, if the management forecast is equal to the consensus analyst forecast, then the news is classified as neutral.

In our analysis, we find that approximately 1 percent of the total observations in the CIG database are misclassified with respect to the type of management guidance as indicated by CIG Code. These misclassifications are sometimes merely the result of inconsistently classifying specific terminology. For example, CIG Code is assigned a specific value when management uses the terminology that earnings will "meet or exceed expectations", but this terminology is sometimes assigned a value corresponding with the terminology "may exceed" a specified amount (in which the amount might not be equal to earnings expectations). Although this type of misclassification might not present serious consequences for researchers, it nevertheless represents an inconsistent coding of management terminology in the CIG database.

Other CIG Code misclassifications are more blatant. For example, First Call may mistakenly classify a qualitative forecast that uses the terminology "at least" as a point estimate. These types of misclassifications can completely change the nature of the guidance provided from a negative announcement for the company to, for example, a neutral or even positive announcement. Nevertheless, in spite of the CIG Code classification errors documented, we believe that a 1 percent error rate for this variable does not present a serious threat to empirical research. We conclude that empirical results should not be substantially affected by a researcher relying upon the CIG Code variable to classify company guidance as point, range, and qualitative forecasts. Nevertheless, when employing a relatively small sample size, the benefits to verifying the accuracy of CIG Code (using other information available in the CIG database) could exceed the costs.

With respect to the variable CIG Description Code, our conclusion is not as favorable. Our analysis suggests a substantial error rate that could affect inferences made in studies that rely on this classification variable to assess reactions to positive, negative, and/or neutral management guidance. Determining the exact error rate of this variable is difficult, since the nature of the information from the management guidance will depend upon the benchmark used to compare the management forecast. We use various analyst forecast benchmarks to derive our own classification code on the nature of the information contained in the management guidance. We then assess the market reactions to positive, negative, and neutral news based on our classification scheme and compare these results to those obtained using CIG Description Code. We find that when our classification scheme conflicts with CIG Description Code, the market reaction is generally consistent with the nature of the information based on our classification code. These results suggest that researchers would do well to derive their own analyst forecast benchmark to determine the nature of the information contained in management guidance rather than relying on the CIG Description Code provided by First Call.

This paper is the first that we are aware of that carefully scrutinizes the accuracy and consistency of the First Call CIG database, which is becoming increasingly important in accounting and management research. Several academic studies employ this database to examine the value-relevance of and incentives behind management voluntary disclosures (Soffer, Thiagarajan, & Walther, 2000; Clement, Frankel, & Miller, 2003; Hutton, 2005; Rogers & Stocken, 2005). While our findings do not distract from the importance and usefulness of the database, we do document results that strongly suggest researchers should consider developing their own metric in determining the nature of the news provided in the company disclosure. Our findings should be of interest to researchers and other users of this database.

The remainder of the paper is organized as follows. In the next section, we provide more description of the First Call database and explain our procedures in assessing the consistency and accuracy of the variables of interest. In Section 3, we present our empirical results. The paper concludes with a summary of our findings and a discussion on the implications of our results.

2. DESCRIPTION OF FIRST CALL DATABASE AND RESEARCH METHODOLOGY

Many studies rely upon First Call's CIG database as a source for management earnings forecasts and earnings preannouncements.[1] For example, Clement et al. (2003) use the CIG database to extract confirming management forecasts issued between 1993 and 1997. Their evidence reveals that confirming forecasts reduce uncertainty about future earnings and that investors price this reduction in uncertainty. In addition, Rogers and Stocken (2005) use First Call management forecasts to determine that managers who are more likely to face litigation release less-optimistic forecasts. In particular, their study makes use of the CIG Code variable, which we specifically analyze in this paper. The First Call CIG database is also utilized in Soffer, Thiagarajan, and Walther's (2000) overall analysis of the factors influencing the decision of a firm to voluntarily provide an earnings preannouncement.

In the following paragraphs, we explain the format of the database and the research methodology employed in this study.

2.1. Description of the First Call CIG Database

The First Call CIG database consists of management earnings forecasts and earnings preannouncements. Although coverage began in 1990, the number of observations in the early periods was sparse and did not exceed 100 for any given fiscal period until the year 1995. Coverage has continued to expand over time until the number of observations exceeds 5,000 for the fiscal period ended December 2004. As of the end of 2004, the database is approaching 70,000 total observations, which represents a substantial increase in the availability of management earnings guidance disclosures when compared to what was used in research studies prior to the issuance of the CIG database.[2] Accordingly, the CIG database allows for more powerful and detailed tests while affording significant time savings by providing data in a machine-readable format.

The CIG database is organized in a similar manner as databases that contain analysts' forecasts. A row of data contains (among other variables) a company identifier, the forecast announcement date, the fiscal period for which the disclosure pertains, and whether the earnings guidance is for quarterly or annual data. Since the earnings guidance can come in the form of a (1) point estimate, (2) range estimate, or (3) qualitative disclosure, two variables (CIG1 and CIG2) are used to indicate the magnitude of the management forecast. That is, for a point estimate, the variable CIG1 is set equal to the forecast and CIG2 is coded as missing; for a range estimate, CIG1 (CIG2) is set equal to the lower (upper) bound of the range; and for a qualitative disclosure, both CIG1 and CIG2 are coded as missing and additional variables are available to provide more information about the type of earnings guidance.

2.2. Research Methodology

The specific variables we examine in this study are "CIG Code" and "CIG Description Code". CIG Code is a variable that provides detailed information about the type of guidance issued by management. This variable has 29 different values for earnings guidance,[3] with the most prominent value being "B", which signifies that management provided an earnings forecast that is somewhere between two numbers (i.e., a range estimate). The CIG Code value of "A", which signifies a point estimate was made by management, also has a high frequency of occurrence. Most of the remaining 27 values for CIG Code correspond with qualitative earnings guidance that management might provide (see Table 1 for a complete listing of the CIG Code values and their meaning). For example, the letter "O" signifies that management disclosed they are "okay with expectations" and the letter "E" signifies management guidance indicating that earnings will be "at least" a specified amount.

To assess the accuracy and consistency of the CIG Code classification, we compare the value assigned to CIG Code by First Call with a comment field contained in the CIG database. The comment field provides a brief text that describes the terminology used in the management earnings guidance disclosure. Thus, for example, suppose the CIG Code has a value of "A" and CIG1 equals \$2.32, then the comment field should contain text indicating that the forecast is "about \$2.32". If, on the other hand, the comment field indicates that the forecast is "at least \$2.32" or "below \$2.32", this would suggest that the CIG Code has been misclassified since the management guidance is not a point estimate but rather, qualitative guidance provided by management. In the next section, we provide detailed data as to the frequency of misclassifications we find for each of the 29 different CIG Code values.

The other variable of interest, CIG Description Code, is used to signify the nature of the information contained in the management guidance based on how it compares to the existing analysts' consensus forecast at the time of the guidance. When the management earnings forecast is higher (lower) than the consensus forecast from analysts, this qualifies as positive (negative)

Table 1. Frequency and Definitions of CIG Codes and Misclassifications.

CIG Codes and Definitions	Frequency in First Call	Misclassified	% Misclassified	Most Frequent Group Reclassifications (N)
1 – May be below	69	4	5.8	K(2)
2 – Not comfortable with	6	2	33.3	F(1), N(1)
3 – Significantly more than	82	0	0.0	
4 – Significantly less than	139	2	1.4	1(1), 8(1)
5 – Meets or exceeds expectations	175	9	5.1	C(4), E(2)
6 – May not meet earnings of between	12	6	50.0	1(2), L(2)
7 – Slightly more than	121	16	13.2	E(12)
8 – Slightly less than	54	2	3.7	L(2)
A – About	12,283	127	1.0	B(58), N(29), E(19)
B – Between	39,394	83	0.2	E(17), F(13), H(10)
C – May exceed	182	19	10.4	E(9), M(4), U(3)
D – Below expectations	1,433	39	2.7	K(14), N(13), B(4)
E – At least	1,152	8	0.7	5(3), A(2)
F – Comfortable with	2,625	10	0.4	O(3), B(2), E(2), N(2)
G – Low end of	369	13	3.5	B(9)
H – High end of	405	26	6.4	E(11), M(9), B(6)
J – May not meet expectations	25	4	16.0	N(3)
K – May be below expectations	1	0	0.0	
L – Less than	1,332	43	3.2	4(19), 8(16)
M – More than	1,227	37	3.0	7(22), 3(9)
N – None	120	56	46.7	B(15), X(7), A(4)
O – Okay with expectations	2,263	11	0.5	H(5), N(4)
P – Above expectations	374	10	2.7	5(3), E(2), H(2), M(2)

Table 1. (*Continued*)

CIG Codes and Definitions	Frequency in First Call	Misclassified	% Misclassified	Most Frequent Group Reclassifications (N)
U – At or below	105	11	10.5	E(8), A(3)
V – As low as	57	2	3.5	B(2)
W – As high as	87	2	2.3	A(1), N(1)
X – Expects loss	975	9	0.9	Y(2)
Y – Expects profit	725	9	1.2	B(4), A(2)
Z – Breakeven	363	42	11.6	B(27), A(6), 7(4)
Total	66,155	601	0.9	

Note: The sample is comprised of all available earnings guidance observations from the First Call CIG database. Misclassifications are determined by comparing the CIG Code and its definition (as published by First Call) to the comment field within the CIG database that provides a brief description of the terminology used within the company announcement.

news for the company; whereas, if the management forecast is equal to the consensus analyst forecast, then the news is classified as neutral. In many cases, the nature of the news contained in the management guidance cannot be unambiguously determined. For example, suppose management discloses that current period earnings should be "at least \$2.00 per share" and the existing consensus analyst forecast is \$2.10 per share. The classification of this guidance as good, bad, or neutral news is debatable and depends on other contextual features within the announcement. In these cases, the CIG Description Code should be coded as "A", which signifies that the specific announcement is not coded as either positive, negative, or neutral.

To assess the accuracy and consistency of CIG Description Code as provided by First Call, we develop our own classification scheme based on various analyst forecast metrics that exist at the time of the management guidance. First Call stipulates that the direction of the earnings guidance (i.e., positive, neutral, or negative) as coded by CIG Description Code is based on the consensus analyst forecast that exists at the time of the guidance. However, the consensus analyst forecast can differ depending on the individual analysts' forecasts that are included in the consensus. Therefore, we use a variety of analyst forecast benchmarks.

The first benchmark we use is the consensus analyst forecast at the time of the guidance as reported by the First Call CIG database. This consensus estimate is often included in the CIG comment field (as described above). Our classification scheme codes the management guidance as, respectively, positive, negative, or neutral when the management point forecast is greater than, less than, or equal to the consensus analyst forecast. For range estimates, the mid-point of the range is used to compare with the consensus analyst forecast unless the company guidance indicates management expects actual earnings to be at the high or low end of the range. In these cases, the high point or the low point of the range is used to compare with the consensus analyst forecast.

For qualitative forecasts, it depends on the context of the earnings guidance with respect to the existing analyst forecast. To provide some examples, if the management guidance indicates that management is ''okay with expectations'', this announcement would be classified as neutral. If the qualitative guidance indicates that management expects to report earnings greater than \$2.00 and the current consensus forecast is \$1.90, this announcement would be classified as positive; however, if for the same earnings guidance the consensus forecast was \$2.05, this announcement would be coded ''CBD'' (cannot be determined), since greater than \$2.00 could be negative guidance ($<$\$2.05), neutral guidance ($=$\$2.05) or positive guidance ($>$\$2.05).

This consensus analyst forecast as provided by the CIG database is the most appealing benchmark to use when assessing the accuracy of the CIG Description Code since it is this consensus forecast that is most likely used by First Call to determine the direction of the earnings guidance. However, this consensus forecast is missing for over half of the database population. To provide further evidence on the accuracy of the CIG Description Code, we calculate two additional analyst forecast benchmarks for firm i in period t to compare with the management forecast. Using the detailed analyst forecast database published by First Call, we take all the individual analysts' forecasts for a given fiscal period that exist prior to the management earnings guidance announcement and calculate the mean of these forecasts. If the same analyst provides two different forecasts for the same firm, we include only the most recent forecast made by that analyst. Based on the calculated mean consensus forecasts, we develop our own determination of the direction of management guidance using the same rules as described above.[4]

Finally, our last analyst forecast benchmark is a ''combined'' measure, which is defined as the consensus forecast provided by the CIG database when available. Otherwise, it is the calculated analyst forecast using the individual forecasts obtained from the First Call detailed analyst forecast database. In this manner, we maximize the number of observations available in our analysis while retaining the benefits of using the consensus forecast as provided by the CIG database, when available.

3. EMPIRICAL RESULTS

The sample for this study is comprised of all available observations from the CIG database as of February 2005, which results in 66,155 earnings guidance observations. The earliest observation occurred in 1990 and the voluntary management disclosures include guidance for both annual and quarterly earnings. The number of unique company identifiers is 6,519.

3.1. CIG Codes

Our initial task is to compare the CIG Code classification as presented in the CIG database with the text in the comment field within the same database. Table 1 provides a listing of the CIG Code values used by First Call and their definitions. Also, Table 1 presents the frequency of each CIG Code value as classified within the CIG database along with the percentage and type of misclassification from our analysis.

The CIG Code value with the most frequent occurrence is "B", which is a type of range estimate and represents approximately 60 percent (39,394/66,155) of the total observations. Other types of range estimates are assigned values of "G" and "H", where management indicates earnings will fall within the low end and high end of a specific range, respectively (1.2 percent of the sample).

Point estimates also represent a substantial percentage of the total sample. The CIG Code "A" indicates that management disclosed that actual earnings will be about a specific amount and consists of approximately 18.5 percent of the sample. The CIG Codes "F" and "Z" also represent types of point estimates where management discloses, respectively, they are "comfortable with" a specific amount and they expect to "breakeven" (4.5 percent of the sample).

The remaining CIG Codes represent different types of qualitative guidance that management can provide. Some of these codes indicate that management believes actual earnings will exceed or fall below a specified point or range. Other codes indicate that the management guidance is relative to existing expectations as defined by analysts' forecasts. In total, the qualitative guidance observations represent approximately 16.2 percent of the total sample (10,716/66,155).[5]

The CIG Code value "N" deserves special consideration. This code value is reserved by First Call to indicate that no guidance is provided. In looking at the comment field for these observations, it appears that the most common type of disclosure represented by this code is one in which management withdraws guidance that was previously issued. In some cases, this code is used when management forecasts a catastrophic loss or provides an earnings growth forecast.

3.1.1. Misclassification of CIG Codes

To assess the accuracy and consistency of the CIG Code classification as provided by the CIG database, we examine for every observation the value of the CIG Code and compare it with the comment field within the same database. As presented in Table 1, our analysis revealed 601 cases where there is a discrepancy between the CIG Code classification and the terminology used by management in their guidance announcement as indicated by the CIG comment field. This level of discrepancy represents an error rate of slightly less than 1 percent for the entire database.

In examining the error rate across classifications, we observe the highest occurrence of errors for the CIG Code value "A", which represents a point estimate. We find that the earnings guidance for 127 observations with this classification is something other than a point estimate. The final column in Table 1 indicates the most frequent type of reclassification that would occur for the misclassified observations given the information in the CIG comment field. Thus, there are 58 observations where the CIG Code has a value of "A", indicating a point estimate, but the comment field indicates the earnings guidance is in the form of a range and therefore, the CIG Code value should be "B". Further, given the information in the comment field, 29 observations would be reclassified as "N" (e.g., the disclosure is for forecasted earnings growth or management withdrew guidance that was provided on an earlier date), and 19 observations would be reclassified as "E" (management's forecast indicated that actual earnings would be "at least" a specified amount).

The CIG Code value of "B", indicating a range forecast, has the next highest occurrence of errors with 83. However, given that this classification code also has the highest number of observations, this error rate is a very low 0.2 percent. The most frequent misclassification for this group is where the comment field indicates that for 17 observations, management forecasts actual earnings that are "at least" a specified amount. For 13 observations, the earnings guidance is actually a point estimate. The next most frequent error is where the earnings guidance is a range but management indicated actual earnings will fall at the high end of the range and therefore, the guidance should be assigned a CIG Code value of "H".

The highest error rate occurs for the CIG Code value "N". The definition for this CIG Code value is "none", which appears to correspond with management withdrawing previously issued guidance, an earnings growth forecast, or earnings guidance related to a catastrophic loss. However, in examining this code, we find 56 observations that are incorrectly coded, which represents an error rate of almost 47 percent. The most frequent type of earnings guidance for these misclassified observations is in the form of a range estimate (15 observations).

The CIG Code value "Z", representing management guidance indicating the company will breakeven, also has a relatively high error rate of 11.6 percent. Of the 42 observations that are misclassified as "Z", most should be classified as a range estimate (27 observations).

Some of the 601 misclassifications that we observe do not represent a major error. For example, the CIG Code value of "7" indicates that management believes earnings will be "slightly more than" a specified amount. Of the 16 observations that are misclassified in this group, 12 observations should be classified as "E", which indicates management believes actual earnings will be "at least" a specified amount. The difference in meaning between the categories "7" and "E" seems small. Nevertheless, we include these observations as errors since they represent an inconsistency in the use of the different CIG Codes.

Table 2 provides a more summarized description of the error rates after pooling CIG Codes into five different groups. Group 1 represents guidance where management discloses that actual earnings may or will fall below a specified point or range. Group 2 consists of guidance where management believes earnings may or will exceed a specified point or range. Groups 3 and 4 comprise point and range estimates, respectively. Finally, Group 5 consists of all earnings guidance announcements where the guidance is relative to existing expectations as defined by analysts' forecasts.

Within each group, we separate all observations according to whether the guidance relates to annual or quarterly observations. We then determine the number of observations that are misclassified by First Call within the corresponding CIG Codes and also, the number of observations that should be classified in the corresponding CIG Codes. The last column in Table 2 presents a revised number of observations within each group after making the appropriate adjustments. An overall accuracy score, labeled percentage turnover, is computed as the number of misclassifications divided by the number of observations that fall within each group as coded by First Call.

From Table 2, there does not appear to be systematic differences in error rates between annual and quarterly data. CIG Codes for range estimates appear to be the most accurate (consistent with the CIG Code "B" having a very small error rate as disclosed in Table 1) with a turnover percentage of

Table 2. Frequencies of Coding Errors on Type of Management Guidance in First Call's CIG Database.

Type of Management Guidance	CIG Codes and Definitions	Period	Observations as Classified by First Call	Misclassified Observations			Revised Observations
				Less: Should be classified as something else	Add: Erroneously classified as something else	% Turnover (misclassified/First Call observations)	
Group 1 – Earnings may or will fall below a specified point or range	1 – May be below 2 – Not comfortable with 4 – Significantly less than 6 – May not meet earnings of between 8 – Slightly less than L – Less than U – At or below W – As high as X – Expects a loss	Quarter	2,259	21	20	1.8	2,258
		Annual	520	10	17	5.2	527
Group 2 – Earnings may or will exceed a specified point or range	3 – Significantly more than 7 – Slightly more than C – May exceed E – At least M – More than V – As low as Y – Expects profit	Quarter	2,046	23	74	4.7	2,097
		Annual	1,499	8	35	2.9	1,526

Group 3 – Point estimate	A – About F – Comfortable with	Quarter	8,876	118	26	1.6	8,784
	Z – Breakeven	Annual	6,395	53	21	1.2	6,363
Group 4 – Range estimate	B – Between G – Low end of	Quarter	21,225	54	95	0.7	21,266
	H – High end of	Annual	18,943	39	43	0.4	18,947
Group 5 – Guidance is relative to expectations	5 – Meets or exceeds expectations D – Below expectations	Quarter	2,714	33	17	1.8	2,698
	J – May not meet expectations K – May be below expectations O – Okay with expectations P – Above expectations	Annual	1,557	22	17	2.5	1,552

0.7 percent and 0.4 percent for annual and quarterly data, respectively. CIG Code classifications for qualitative earnings guidance are the least accurate, especially for the annual observations in Group 1 with a turnover percentage of 5.2 percent.

In summary, there is a relatively large number of CIG Codes in the CIG database that are coded incorrectly. However, the overall error rate is only about 1 percent. A larger error rate occurs where the earnings guidance is qualitative, but many of these errors appear to be minor since the First Call classification appears to have only a small difference in meaning from what the classification should have been. Overall, we believe it is unlikely that the errors we discovered would cause significant changes in inferences when the CIG database is employed. Nevertheless, researchers should be aware that errors in the CIG Code exist. Accordingly, when the CIG database is employed and the sample size is small, the researcher might do well to examine the comment field in the database to ensure that the CIG Code is classified properly since the costs would be relatively small and the effects of errors in small samples could be nontrivial.

3.1.2. Examination of Source Documents

The above procedures to examine the accuracy of CIG Code critically hinge on our assumption that First Call's comment field accurately reflects the terminology provided by management in the original press release. If errors exist in the comment field, then we could be overstating the error rate in the CIG Code when it reflects the actual press release but the comment field is inaccurate. Furthermore, our analysis could understate the error rate in the CIG Code when the CIG Code accurately reflects an incorrect comment field.

To assess the accuracy of the comment field, we randomly select two sets of observations to examine their original press releases. The first subsample is comprised of observations where the CIG Code is consistent with the comment field. From this subsample, we randomly select 50 observations evenly distributed over our sample period where we find a press release for the company on the same date as indicated in the First Call database.[6] Of these 50 press releases, we are unable to find a specific management forecast within three of them.[7] For the remaining 47 observations where we do locate a management forecast, we find that the comment field is accurate. Thus, we conclude that when the CIG Code and comment field are consistent, the comment field accurately reflects management's terminology in the press release.

For the second set of observations, we follow the same procedures as above except that the subsample used is those firms in which the CIG Code and comment field are inconsistent with each other. We randomly select 50 observations evenly distributed over the sample period where we find a press release that occurred on the same date as indicated by First Call.[8] We do not find a management forecast in 2 of the 50 press releases. Of the remaining 48, we find one case where the comment field, while technically correct, excludes some important details. Specifically, for this observation, the comment field indicates that management provided qualitative guidance that earnings would fall below analyst's expectations, which would be consistent with a CIG code of "D" and is what we use in assessing the accuracy of the CIG Code. In fact, management did use this terminology, but management also provided a forecasted range, which is consistent with a CIG Code of "B" and is what First Call reports. In this case, First Call could have used either the CIG Code "D" or "B". The code that First Call reports is one that provides more quantitative information ("B"), but is not included in the comment field. Thus, our procedure of relying upon the comment field results in 1 case (out of 48) where the CIG Code is correct but we categorized it as being misclassified.

Overall, our analysis suggests that the comment field within First Call is accurate and our reliance on the comment field to assess the accuracy of the CIG Code is valid. The one case where we find that our categorization of the CIG Code is in error due to relying on the comment field suggests that our misclassification rate, as reported in Table 1 is slightly overstated. However, this result only reinforces our overall conclusion that the error rate in the CIG Code is most likely not extensive enough to cause erroneous inferences in large samples.

3.2. CIG Description Code

3.2.1. Misclassification of CIG Description Code

In this section, we examine the accuracy of the CIG Description Code, which provides information about the direction of the earnings guidance, i.e., positive, negative, or neutral. The first column of Table 3 lists the various categories provided by First Call. The value "A" is used by First Call when the observation is not coded. This could be a result of ambiguity in the guidance announcement. As an example, when management announces that earnings should be at least \$2.00 and the existing consensus analyst forecast is \$2.05, whether this guidance represents negative, positive, or neutral news is difficult to ascertain and will probably depend on other contextual information within the announcement.

Table 3. Frequencies of Coding Errors on Direction of Management Guidance in First Call's CIG Database.

Code for Direction of Earnings Guidance from First Call	Observations	Direction of Earnings Guidance when Management Forecast is Compared to:														
		Analyst forecast from First Call CIG database					Mean analyst forecast from First Call EST database					Combination: Forecast from First Call EST database when forecast from CIG database is unavailable				
		U	N	D	CBD	NA	U	N	D	CBD	NA	U	N	D	CBD	NA
A – Not coded	7,483	443	1,487	1,040	281	4,232	673	1,521	1,446	510	3,333	750	1,535	1,546	632	3,020
D – Negative surprise	17,921	61	7	6,002	111	11,740	623	92	12,905	410	3,891	493	63	15,127	453	1,785
E – Positive surprise	7,602	3,578	9	30	11	3,974	5,160	56	190	126	2,070	6,592	42	186	130	652
G – If "greater than number" is < = mean, qualifies as a positive surprise	4	3	0	0	0	1	3	0	1	0	0	3	0	1	0	0
L – If "less than number" is < = mean, qualifies as negative surprise	1	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0
M – Does not qualify as a surprise	32,675	5,949	3,324	5,811	233	17,358	9,825	4,523	8,568	896	8,863	11,875	5,678	11,195	960	2,967
N – If mean is positive, qualifies as a negative surprise	298	35	45	122	0	96	41	46	128	2	81	41	46	129	2	80
P – If mean is negative, qualifies as a positive surprise	2	0	0	1	0	1	1	0	1	0	0	1	0	1	0	0
S – Compare single figure forecast to mean using same algorithm as for reported earnings	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1
Missing code	55	8	4	20	1	22	3	3	13	0	36	10	4	25	1	15
Total		10,077	4,876	13,019	637	37,426	16,330	6,241	23,252	1,944	18,251	16,485	6,379	22,949	1,954	18,251

U, Up; N, neutral; D, down; CBD, cannot be determined based on the open-ended management guidance and the available analyst forecast; NA, analyst forecast is unavailable.

The values G, L, N, P, and S are used sparingly and conditional upon the actual management forecast. It is curious as to why these codes are used at all since the direction of the guidance could be determined by comparing the management forecast with the analyst consensus forecast, which is, presumably, how the guidance direction was determined for all other observations.

Table 3 reveals that almost one-half of the observations (32,675) have a CIG Description Code of "M", indicating neutral guidance. Of the observations that are coded and represent a surprise to the market, many more observations represent a negative surprise (17,921) compared to the number of observations that represent a positive surprise (7,602). In the remaining columns of Table 3, we compare the CIG Description Code to the direction of the earnings guidance news based on various consensus analysts' forecasts.

The first earnings benchmark we use is the consensus forecast as obtained from the comment field in the CIG database. There are two disadvantages to using this earnings benchmark. First, this variable is unpopulated for over half of the observations (37,426 observations coded as NA). The second disadvantage is one of convenience; the variable is not easily extracted since it is embedded within the text comment field. However, the significant advantage to this benchmark is that presumably, it would be the same consensus forecast that First Call uses to determine the direction of the earnings news as coded in the CIG database.

As indicated in Table 3, large differences exist between the CIG Description Code and the direction of the earnings guidance when compared with the consensus analyst forecast as obtained from the CIG database. In fact, of the 10,077 observations that were classified as having positive earnings guidance based on the consensus forecast published in the CIG database, only 3,578 observations are consistently classified using the CIG Description Code. Most of the observations classified as positive news using the published analyst forecast in the CIG database were classified as neutral according to the CIG Description Code (5,949 observations). Similarly, approximately 45 percent (5,811 observations) of the negative news observations based on the published analyst forecast were classified as neutral according to the CIG Description Code.

A better matching is found for the neutral guidance news. That is, of the 4,876 observations that were classified as neutral according to the published analyst forecast, 68 percent were also classified as neutral according to the CIG Description Code (3,324 observations). However, we find a large portion of the neutral observations based on the published analyst forecast were not coded according to the CIG Description Code (i.e., had a value of "A").

Given the large number of observations where a consensus analyst forecast is unavailable in the CIG database, we calculated a consensus forecast using available individual analysts' forecasts from the First Call analyst forecast database. The second set of results in Table 3 corresponds to the calculated mean consensus forecast. The number of observations where a consensus forecast is not available (coded as NA) drops substantially when we employ this method. However, the disadvantage of this approach is that our calculated consensus could be different from the consensus that is used to determine the CIG Description Code.[9] Nevertheless, the results for our calculated benchmarks are quite similar to the published consensus forecast in the CIG database. Specifically, we find many observations that have positive and negative earnings guidance, based on our calculated consensus, but are classified by the CIG Description Code as neutral guidance. Also, when the CIG Description Code is not coded (i.e., has a value of "A"), we are able to determine the direction of the earnings news from our calculated analyst forecast.

Finally, to maximize the number of observations in our tests, we combine the consensus forecast as provided by the CIG database with the calculated consensus. That is, we use the calculated consensus when the CIG database does not provide a consensus. In this manner, we delete from the analysis only those observations where both benchmarks are not available. The third set of results presented in Table 3 corresponds with the combined benchmark. The results are qualitatively similar, in that we find several observations where the CIG Description Code appears to be coded incorrectly in the CIG database.

3.2.2. When Does First Call Classify an Observation as a Surprise?

The most striking aspect within Table 3 is probably the large proportion of observations within the First Call CIG database that are classified as "M – does not qualify as a surprise" even though we are able to classify many of these observations as either upward or downward guidance when compared with the consensus analyst forecast. We examine two possible reasons for this result.

First, the CIG database potentially imposes a materiality constraint before it classifies guidance as either upward or downward. We assess this possibility by examining the distributions of the magnitude of the guidance deviation from the consensus forecast for neutral, negative, and positive guidance as classified by First Call (classifications M, D, and E, respectively). The guidance deviation is defined as the difference between the CIG and the existing consensus analyst forecast as reported by the First Call CIG database. The distributions are presented as bar graphs in Fig. 1.[10]

An examination of the bar graphs suggests the possibility of a materiality constraint imposed by First Call. The highest frequency of guidance deviation for negative (positive) surprises occurs at the −\$0.04 (\$0.03) cents. A guidance deviation of one penny for the positive and negative surprise groups is relatively less frequent when compared with a frequency of two or three pennies guidance deviations. Similarly, for the neutral CIG Description Code, there are many observations classified as neutral where the guidance deviation is only ± 1 penny.

However, allowing for a materiality constraint before First Call classifies guidance as a positive or negative surprise only mitigates, but does not eliminate, the apparent inconsistencies with the CIG Description Code. Note that there are 2,409 observations classified as neutral within First Call (i.e., an "M" classification) that have a guidance deviation of ± 2 pennies. Even at a guidance deviation of three pennies, there are close to 1,500 observations classified as neutral. Furthermore, if we allow for a materiality constraint, then many observations that are classified as positive or negative surprises should then be re-classified as neutral. For example, there are 314 observations with a guidance deviation of −1 penny that are classified as a negative surprise by First Call, but with a 1-penny materiality constraint, these observations should be classified as neutral. In our opinion, the materiality constraint does not seem to be a viable explanation for the apparent widespread inconsistencies in the CIG Description Code.

Another possible explanation is that a large proportion of the guidance observations is in the form of a range estimate and First Call might classify all observations as neutral if the consensus analyst forecast falls anywhere within the range. Recall that our procedure in determining the nature of the guidance is to take the midpoint of the range and compare it with the consensus analyst forecast, unless the guidance specifically indicates that management expects actual earnings to be at the high or low end of the range, in which case the high or low point of the range is compared with the consensus analyst forecast. Our procedure is consistent with prior research that takes the midpoint of the range estimate to measure the level of news disclosed by management at the time of the guidance (Rogers & Stocken, 2005; Clement et al., 2003; Baginski & Hassell, 1990).

To assess the degree to which range estimates affect the neutral classification of the CIG Description Code, we repeat our analysis in Table 3, after classifying as neutral all observations where the existing consensus analyst forecast falls anywhere within the range estimate. The results from

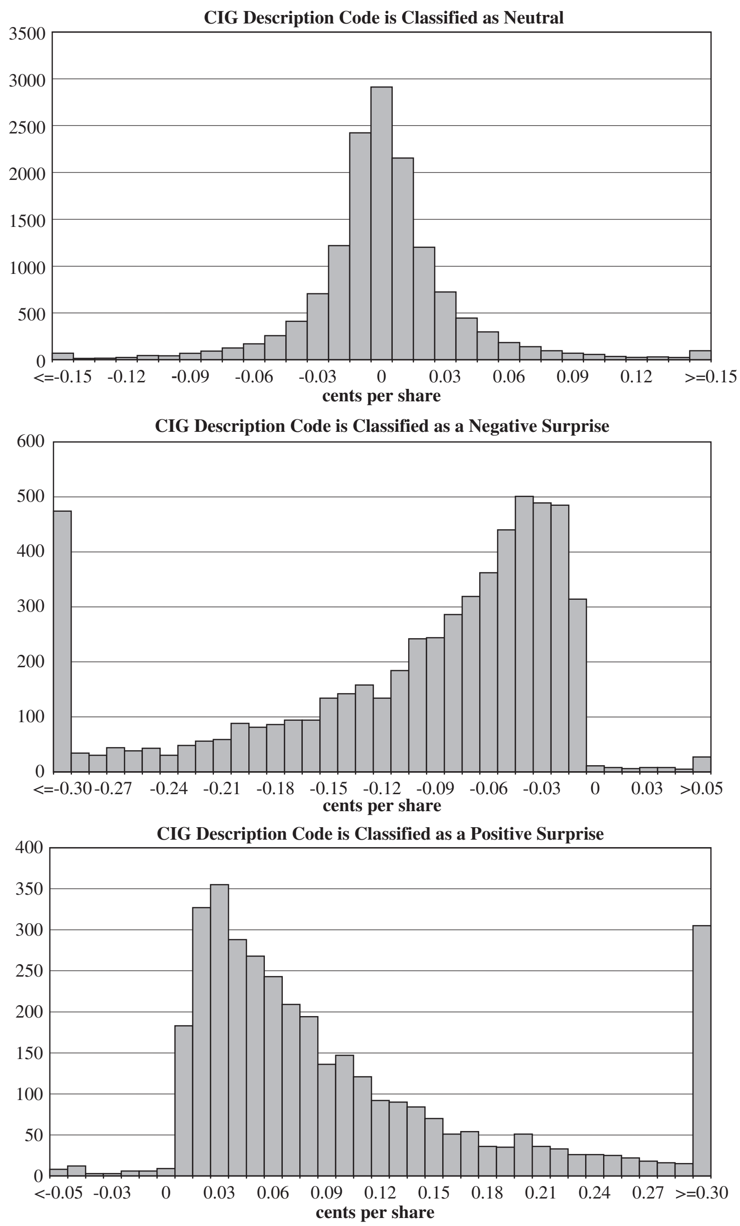


Fig. 1. Frequency Distribution of Guidance Deviations from Consensus Analysts' Forecasts Across First Call's CIG Description Codes. *Note:* Guidance deviation is defined as the difference between the company issued guidance and the existing consensus analyst forecast as reported by the First Call CIG database.

this repeat analysis are presented in Table 4. As the table reveals, the number of observations classified as neutral increases no matter which consensus analyst forecast measure we use. Focusing on the consensus forecast as provided by First Call, the number of neutral observations increases from 4,876 to 13,301. More importantly, the number of observations classified as neutral where the CIG Description Code also represents no surprise ("M") increases from 3,324 to 10,515.

Despite the substantial increase in consistency for observations classified as "no surprise" by the CIG Description Code, we observe a *decrease* in consistency for the negative and positive surprise groups. That is, for both the negative and positive groups (CIG Description Codes "D" and "E", respectively), there are fewer observations that are consistently classified as upward and downward guidance when compared to the consensus analyst forecast. Instead, a greater number of these observations are now classified as neutral.

In summary, we examine two possibilities that could potentially explain the large number of observations classified by First Call as not qualifying as a surprise: (1) a materiality constraint imposed by First Call before classifying observations as a positive or negative surprise and (2) all observations where the consensus falls anywhere within a range estimate classified as neutral. While both of these potential explanations help to mitigate the misclassification of neutral observations, neither one completely eliminates the problem. Moreover, both procedures exacerbate inconsistencies for negative and positive surprise classifications.

3.2.3. Analysis of Stock Price Response to Company Issued Guidance

The analyses in Tables 3 and 4 clearly show large discrepancies between the CIG Description Code and the direction of the earnings guidance based on published and calculated consensus analysts' forecasts. To provide further evidence on this issue, we compute for each observation in our sample a 3-day market adjusted return surrounding the date of the earnings guidance.[11] Table 5 presents the mean and median abnormal returns for various subgroups of our sample.

Panel A of Table 5 compares the CIG Description Code with the direction of the earnings news based on the consensus forecast from the CIG database.[12] Panels B and C of Table 4 are similarly organized, but the earnings benchmarks are different. The earnings benchmark in Panel B is the mean consensus analyst forecast as calculated using the First Call analysts' forecast database. In Panel C, we combine the two analyst forecast sources by using the consensus forecast provided in the CIG database when

Table 4. Repeat Analyses of Coding Errors on CIG Description Code after Reclassifying as Neutral Observations where Analyst Forecast Falls Anywhere within Range Estimate Issued by Management.

Code for Direction of Earnings Guidance from First Call	Observations	Direction of Earnings Guidance when Management Forecast is Compared to:														
		Analyst forecast from First Call CIG database					Mean analyst forecast from First Call EST database					Combination: Forecast from First Call EST database when forecast from CIG database is unavailable				
		U	N	D	CBD	NA	U	N	D	CBD	NA	U	N	D	CBD	NA
A – Not coded	7,483	401	1,571	998	281	4,232	589	1,691	1,360	510	3,333	649	1,739	1,443	632	3,020
D – Negative surprise	17,921	49	717	5,304	111	11,740	509	1,038	12,073	410	3,891	453	1,371	13,909	453	1,785
E – Positive surprise	7,602	3,156	440	21	11	3,974	4,756	495	155	126	2,070	5,931	736	153	130	652
G – If "greater than number" is < = mean, qualifies as a positive surprise	4	0	3	0	0	1	0	3	1	0	0	0	3	1	0	0
L – If "less than number" is < = mean, qualifies as negative surprise	1	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0
M – Does not qualify as a surprise	32,675	2,536	10,515	2,033	233	17,358	6,565	10,992	5,359	896	8,863	6,732	16,263	5,753	960	2,967
N – If mean is positive, qualifies as a negative surprise	298	33	51	118	0	96	39	53	123	2	81	37	56	123	2	80
P – If mean is negative, qualifies as a positive surprise	2	0	0	1	0	1	1	0	1	0	0	1	0	1	0	0
S – Compare single figure forecast to mean using same algorithm as for rep. Earnings	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1
Missing code	55	4	16	12	1	22	3	4	12	0	36	6	17	16	1	15
Total		6,183	13,301	8,488	637	37,426	12,463	14,275	19,085	1,944	18,251	12,505	14,227	18,964	1,954	18,251

U, Up; N, neutral; D, down; CBD, cannot be determined based on the open-ended management guidance and the available analyst forecast; NA, analyst forecast is unavailable.

Table 5. Short-Window Market-Adjusted Equity Returns Surrounding Earnings Guidance.

Panel A: Analyst forecast from CIG database

Direction of Guidance Per First Call	Direction of earnings guidance when compared to existing consensus analyst forecast as obtained from CIG																	
	U			N			D			CBD			NA			TOTAL		
	Observations	Mean	Median	Observations	Mean	Median	Observations	Mean	Median	Observations	Mean	Median	Observations	Mean	Median	Observations	Mean	Median
Positive surprise	3,323	**4.9**	**4.0**	7	0.8	−0.1	27	**−10.0**	−4.0	11	1.0	−1.4	3,004	**5.3**	**4.5**	6,372	**5.0**	**4.2**
Neutral	5,674	**1.7**	**1.4**	2,915	**0.6**	**0.5**	5,516	**−1.2**	**−0.5**	220	−0.2	0.1	13,203	0.0	**0.2**	27,528	**0.2**	**0.3**
Negative surprise	57	**3.1**	1.6	6	−3.8	−1.4	5,406	**−6.5**	**−4.3**	107	**−3.2**	−1.6	8,036	**−9.4**	**−6.3**	13,612	**−8.1**	**−5.3**
Not coded	408	**4.3**	**3.2**	1,050	−0.7	−0.3	809	**−9.7**	**−6.1**	255	**−3.2**	**−2.6**	2,940	0.2	−0.1	6,023	**−1.94**	**−1.00**
Total	9,462	**3.0**	**2.1**	3,978	0.3	0.3	11,758	**−4.2**	**−2.0**	593	**−2.0**	**−0.9**	27,183	**−2.2**	**−0.7**			

Panel B: Mean analyst forecast from First Call EST database

Direction of Guidance Per First Call	Direction of earnings guidance when compared to existing mean consensus analyst forecast as obtained from First Call																	
	U			N			D			CBD			NA			TOTAL		
	Observations	Mean	Median	Observations	Mean	Median	Observations	Mean	Median	Observations	Mean	Median	Observations	Mean	Median	Observations	Mean	Median
Positive surprise	4,414	**5.6**	**4.8**	46	4.0	4.3	151	−1.0	0.1	90	1.3	2.0	1,671	**4.2**	**3.1**	6,372	**5.0**	**4.2**
Neutral	8,615	**1.6**	**1.1**	3,607	0.2	0.2	7,198	**−1.6**	**−0.7**	669	−0.4	−0.0	7,439	**0.4**	**0.4**	27,528	**0.2**	**0.3**
Negative surprise	540	**−2.7**	**−1.6**	73	−2.0	−1.4	9,742	**−9.2**	**−6.2**	291	**−7.7**	**−4.8**	2,966	**−6.0**	**−3.5**	13,612	**−8.1**	**−5.3**
Not coded	582	**3.7**	**3.2**	1,078	−0.7	−0.3	1,055	**−9.3**	**−6.1**	374	**−3.3**	**−1.8**	2,373	**1.1**	0.2	6,023	**−1.94**	**−1.00**
Total	14,151	**2.8**	**2.0**	4,804	−0.0	0.1	18,146	**−6.1**	**−3.3**	1,424	**−2.5**	**−0.9**	14,449	**−0.4**	0.1			

Table 5. (*Continued*)

Panel C: Combination- Forecast from First Call EST database when forecast from CIG database is unavailable

Direction of Guidance Per First Call	Direction of earnings guidance when compared to existing combined analyst forecast benchmark																	
	U			N			D			CBD			NA			TOTAL		
	Observations	Mean	Median	Observations	Mean	Median	Observations	Mean	Median	Observations	Mean	Median	Observations	Mean	Median	Observations	Mean	Median
Positive surprise	5,741	**5.2**	**4.4**	34	3.1	0.5	146	−1.6	0.1	94	1.2	1.7	357	**4.9**	**3.4**	6,372	**5.0**	**4.2**
Neutral	10,566	**1.6**	**1.3**	4,698	0.3	0.2	9,733	**−1.4**	**−0.5**	728	−0.5	−0.1	1,803	0.1	0.0	27,528	**0.2**	**0.3**
Negative surprise	431	**−2.3**	**−1.0**	46	−2.9	−1.4	11,821	**−8.3**	**−5.5**	333	**−7.0**	**−4.1**	981	**−8.8**	**−5.5**	13,612	**−8.1**	**-5.3**
Not coded	661	**3.8**	**3.2**	1,093	−0.7	−0.4	1,159	**−8.8**	**−5.6**	488	**−3.3**	**−2.3**	2,061	**1.5**	0.3	6,023	**−1.94**	**-1.00**
Total	17,399	**2.8**	**2.1**	5,871	0.1	0.1	22,859	**−5.4**	**−2.7**	1,643	**−2.6**	**−1.0**	5,202	**−0.7**	**−0.3**			

U, Up; N, neutral; D, down; CBD, cannot be determined based on the open-ended management guidance and the available analyst forecast; NA, analyst forecast is unavailable.

Bold indicates significance at .01 level.

available and otherwise, we use the calculated consensus from the First Call detailed analyst forecast database.

As expected, when the classifications are consistently positive, the mean 3-day stock price response is significantly positive in all three panels (4.9, 5.6, and 5.2 percent, respectively). Similarly, when the classification of the direction of the earnings guidance is consistently negative, the mean 3-day stock price response is significantly negative (−6.5, −9.2, and −8.3, percent, respectively). For neutral news, when the classifications are consistent, Panel A shows a stock price response that is relatively small but significantly positive (0.6 percent). This positive response is consistent with the finding in Clement et al. (2003) that confirming guidance by management results in a positive stock price response. However, we do not document a similar result for neutral news in Panels B and C. Rather, the stock price response is not significantly different from zero in these panels when the classifications are consistently neutral.

The primary results of interest in Table 5 are when discrepancies exist between the CIG Description Code and the classification based on the published and/or calculated consensus analysts' forecasts. We consider each of these cases individually. First, focusing on Panel A, we consider when the earnings guidance is positive based on the published analyst forecast but is classified as either neutral, negative, or not coded in the CIG Description Code. In each of these cases, the mean 3-day abnormal return is significantly positive, which is consistent with there being error in the CIG Description Code. Panels B and C provide similar results except when CIG Description Code is classified as negative. In this case, the abnormal return is significantly negative, consistent with the CIG Description Code more accurately reflecting the direction of the earnings news compared to the calculated consensus forecast.

Next, we investigate abnormal returns when the earnings guidance is classified as neutral based on the published consensus analyst forecast, but the CIG Description Code indicates a positive surprise, negative surprise, or is not coded. In every one of these cases for all panels of Table 4, the abnormal return is not significantly different from zero. This is consistent with the CIG Description Code containing error and implies that our approach of determining the direction of the earnings news based on published or calculated consensus forecasts is more accurate.

Next, we examine the cells where the direction of the earnings guidance is negative based on the published consensus forecast but the CIG Description Code indicates something else. In almost every cell across the panels, the abnormal return is significantly negative even when the CIG Description

Code indicates something other than a negative surprise. The only exceptions are in Panels B and C where the CIG Description Code indicates a positive surprise and the abnormal returns are not significantly different from zero. Again, the combined evidence is consistent with there being error in the CIG Description Code.

The columns labeled CBD present results where the direction of the earnings guidance could not be determined given the type of guidance and the existing consensus forecast. Also, the columns labeled NA present results for observations in the CIG database for which we were unable to obtain consensus forecasts. The results in these columns do not represent a discrepancy between CIG Description Code and the available consensus forecast but rather, provide an indication of the reliability of the CIG Description Code when a consensus analyst forecast may not be available to the researcher. An examination of these results reflect favorably on the CIG Description Code since a positive (negative) surprise generally results in a significantly positive (negative) abnormal return and neutral news results in an abnormal return that is close to zero or not significantly different from zero.

To summarize, the stock return analysis supports the results in Tables 3 and 4 that there is substantial error in the CIG Description Code concerning the nature of the guidance news. Based on this evidence, we conclude that a researcher would obtain more reliable information with respect to the direction of the earnings guidance by comparing the guidance information with the available consensus forecast in the CIG database or by calculating an earnings benchmark from available analysts' forecasts on First Call's detailed analyst forecast database (or some other analyst forecast database). However, when analysts' forecasts are not available for a specific company, the abnormal returns are generally consistent with the CIG Description Code.

3.2.4. Intertemporal Error Rates

Our final analysis investigates how the error rate in the CIG Description Code has varied over time. This analysis can provide insight as to whether the measurement error is persistent over time or is isolated to a particular time period. One might expect the error rate to diminish over time as the data collection procedures followed by First Call become more standardized.

In calculating the error rates, we focus only on those observations where there is clearly a discrepancy between what is coded in First Call and our determination based on the consensus analyst forecast as provided in the CIG database. Thus, we delete observations that were not coded in First Call or where we were unable to determine the nature of the information due either to the ambiguity of the guidance as compared to the analyst forecast (CBD) or an analyst forecast was unavailable (NA).

Table 6 presents the error rates by year across the negative (D), positive (E), and neutral (M) CIG Description Codes and the total error rate. Two sets of results are presented. In the first panel, we follow our normal procedure of calculating a midpoint when a range estimate is provided by

Table 6. Error Rates Over Time for CIG Description Code.

Year	CIG Description Code				
	N	D – Negative surprise (%)	E – Positive surprise (%)	M – Does not qualify as a surprise (%)	Total (%)
Panel A: Results when midpoint of guidance range is used to compare with analyst forecast to determine nature of guidance					
1995	21	0	0	0	0
1996	81	0	0	12	5
1997	274	2	0	13	8
1998	382	0	17	25	16
1999	110	0	25	2	2
2000	240	0	0	27	25
2001	290	0	11	26	24
2002	7,135	2	2	76	46
2003	5,391	0.2	0	82	52
2004	10,847	1	1	85	52
Panel B: Observation classified as neutral when analyst forecast is anywhere within guidance range					
1995	21	0	0	0	0
1996	81	0	0	11	5
1997	274	2	0	13	8
1998	382	0	17	25	16
1999	110	0	25	2	2
2000	240	0	0	27	25
2001	290	0	33	23	22
2002	7,135	12	12	31	23
2003	5,391	14	13	31	25
2004	10,847	15	13	31	24

Note: The error rate for each CIG Description Code by year is calculated as the number of observations that were coded consistently when compared with the consensus analyst forecast as provided in the First Call CIG database divided by the sum of all observations classified as having upward, downward, or neutral guidance when compared with the same consensus analyst forecast. Observations were not included in this analysis when they were not coded by First Call, did not have an analyst forecast available, or were unable to be coded when compared with the consensus analyst forecast.

management and comparing that midpoint to the consensus forecast. When management indicates they expect actual earnings to fall at the high (low) end of the range, instead of calculating a midpoint, we compare the high (low) point of the range with the analyst forecast. In the second panel, we follow the procedure of coding all observations as neutral if the consensus analyst forecast falls anywhere within the estimated range.

The results in Table 6 clearly show an increasing trend in the total error rate over time as the total number of observations within the CIG database increases. The table also shows that most of the errors occur when First Call codes the nature of the news as neutral, consistent with information provided in Table 3. Thus, the nature of the news classified as neutral by First Call can be determined more accurately when the guidance is compared with existing analysts' forecasts. In Panel B of Table 6, although we effectively allow for more observations to be coded as neutral, we continue to document large error rates in the most recent years. In addition, although the error rate for neutral news decreases substantially in panel B, the error rates for positive and negative surprises increase. Overall, we conclude that the error rate in the CIG Description Code has not been eliminated over time and, in fact, there appears to be an increasing trend in the measurement error, especially for guidance that is classified as neutral by First Call.

4. SUMMARY

In this study, we examine the accuracy and consistency of two variables contained in the CIG database issued by First Call: CIG Code and CIG Description Code. CIG Code provides a description of the type of earnings guidance provided by management. For example, the guidance could be a point estimate, a range estimate, or other qualitative guidance. CIG Description Code indicates the direction of the earnings guidance news relative to existing expectations.

From our analysis of over 65,000 observations on the CIG database, we find 601 observations where the CIG Code was not classified consistently. Some of these misclassifications are blatant errors, such as, a point estimate that is classified as a range estimate; or, a qualitative estimate where management discloses earnings will be "at least" a specified amount but is classified as a point estimate in the CIG database. However, other misclassifications represent minor errors, such as, a qualitative estimate indicating that earnings will be "slightly more than" a specified amount that should have been classified as earnings being ''at least'' a specified amount. Although the error rate is more than we would expect from a reputable and established forecast provider, we doubt that inferences from research that uses the CIG database would be substantially affected by using the CIG Code. Nevertheless, we recommend that for small sample sizes in studies that employ the CIG Code, the researcher verify the accuracy of this variable using information contained in the comment field since the costs are relatively low and the effects of errors on small sample sizes could be significant.

Our conclusions with respect to the CIG Description Code, however, are not as benign. We develop our own scheme to determine the direction of the earnings guidance by comparing the earnings guidance with consensus analysts' forecasts that we either obtain from the CIG database or we calculate from the First Call analyst forecast database. Our analysis finds large discrepancies between our classification schemes and the CIG Description Code, especially for guidance that is coded as neutral by First Call. Measurement error in the CIG Description Code persists even when we relax the parameters that indicate whether guidance is neutral. Analysis of abnormal stock returns surrounding the guidance date corroborates our conclusion. Further, the error rate tends to increase over time as First Call has expanded their coverage of management guidance announcements.

We conclude that researchers would do well to develop their own earnings benchmark when determining the direction of earnings guidance rather than relying on the CIG Description Code alone. However, we also find for the portion of our sample that did not have available analysts' forecasts, that the abnormal returns were generally consistent with the CIG Description Code. Thus, a combination of using published analysts' forecasts and the CIG Description Code might be optimal, in that it increases power by maximizing sample size. That is, we recommend that researchers should use published analysts' forecasts to determine the direction of the earnings guidance, but where sufficient analyst forecast data are not available, the CIG Description Code could be employed.

NOTES

1. A ''preannouncement'' is a voluntary disclosure of a tentative earnings amount made shortly before the formal earnings announcement date (Soffer, Thiagarajan, & Walther, 2000).

2. Early research on earnings guidance required hand collection of voluntary management disclosures and includes studies like Ajinkya and Gift (1984), Waymire (1984), Hassell and Jennings (1986), and McNichols (1989), among several others. Although screening criteria were used to identify the experimental sample (e.g., qualitative forecasts were usually deleted), the sample size in these studies typically ranged from 100 to 1,000 observations. The largest sample size we are aware of within a hand-collected management forecast study is the 3,420 management disclosures (including some qualitative forecasts) used by Lev and Penman (1990).

3. Although not as well populated, the First Call's CIG database also has company issued revenues guidance. There are additional codes for revenues guidance beyond the 29 we examine within this paper.

4. The management earnings forecasts from the CIG database are not adjusted for subsequent stock splits/dividends; whereas, the analyst forecast data provided by First Call are adjusted for stock splits. Accordingly, when comparing management forecasts with analysts' forecasts to determine the direction of the earnings guidance, the management forecasts are manually adjusted for stock splits/dividends using a cumulative adjustment file provided by First Call.

5. In addition to the CIG Code "Z", which could potentially be classified as "A" without apparent loss of information, the CIG Codes "K" and "J" also appear to be redundant to each other. Interestingly, only one observation in the entire CIG database has a CIG Code with the value "K". Other CIG Codes have a very similar meaning but are not exactly equivalent, such as the codes "8" and "L", or the codes "E" and "M".

6. Our source for the press releases is Lexis–Nexis. We attempted but were unable to find a press release on the announcement date indicated in First Call for an additional 27 observations, most of which occurred in the early years of the sample period.

7. In two of the press releases, the comment field appears to reflect management's discussion of *actual* earnings in the prior period. This would suggest that some observations in First Call's Company Issued Guidance database should not be included, as they do not relate to guidance about future earnings. However, it is possible that our search simply failed to find the specific press release on that date where management provided an earnings forecast.

8. For an additional 20 observations, we were unable to find from Lexis–Nexis a press release on the announcement date indicated in First Call. As before, these observations occurred in the early years of the sample period.

9. For firms with both analysts' forecast benchmarks available – a consensus forecast available in the CIG database and our calculated consensus – the correlation between them is 93 percent, indicating that our method of calculating analysts' forecasts closely parallels what is done by First Call.

10. The total number of observations employed in the distribution analysis is less than what is used in Table 3, since the distribution analysis requires quantitative guidance (i.e., qualitative guidance observations are deleted). Also, the number of observations is sensitive to the availability of consensus analyst forecasts. We performed similar analyses using the other consensus forecast measures with qualitatively similar results.

11. The 3-day return window is defined as day -1 to day $+1$ relative to the date of the earnings guidance. The CRSP equally weighted market return is used as the market adjustment.

12. Note that the total number of observations in Table 4 has declined since sufficient price data to compute the 3-day abnormal returns were not available for the entire sample.

REFERENCES

Ajinkya, B. B., & Gift, M. J. (1984). Corporate managers' earnings forecasts and symmetrical adjustments of market expectations. *Journal of Accounting Research*, *22*, 425–444.

Baginski, S., & Hassell, J. (1990). The market interpretation of management earnings forecasts as a predictor of subsequent financial analyst forecast revision. *The Accounting Review*, *68*(January), 175–190.

Clement, M., Frankel, R., & Miller, J. (2003). Confirming management earnings forecasts, earnings uncertainty, and stock returns. *Journal of Accounting Research*, *41*, 653–679.

Hassell, J. M., & Jennings, R. H. (1986). Relative forecast accuracy and the timing of earnings forecast announcements. *The Accounting Review*, *61*(January), 58–75.

Hutton, A. (2005). Determinants of managerial earnings guidance prior to regulation fair disclosure and bias in analysts' earnings forecasts. *Contemporary Accounting Research*, *22*(Winter), 867–914.

Lev, B., & Penman, S. H. (1990). Voluntary forecast disclosure, nondisclosure, and stock prices. *Journal of Accounting Research*, *28*(Spring), 49–76.

McNichols, M. (1989). Evidence on informational asymmetries from management earnings forecasts and stock returns. *The Accounting Review*, *64*, 1–27.

Rogers, J., & Stocken, P. (2005). Credibility of management forecasts. *The Accounting Review*, *80*(October), 1233–1260.

Soffer, L. C., Thiagarajan, S. R., & Walther, B. R. (2000). Earnings preannouncement strategies. *Review of Accounting Studies*, *5*(March), 5–26.

Waymire, G. (1984). Additional evidence on the information content of management earnings forecasts. *Journal of Accounting Research*, *22*, 703–718.

This page intentionally left blank

THE INFORMATION CONTENT OF REVERSE STOCK SPLITS

Dahlia Robinson

ABSTRACT

This chapter investigates whether abnormal reverse split announcement returns are related to information about earnings. I find that abnormal announcement returns are negative on average and significantly correlated with unexpected earnings in the years prior to, and subsequent to the reverse split event. I also find that analyst earnings forecasts are revised downwards after reverse split announcements and that these forecast revisions are correlated with abnormal announcement returns. Finally, I document a significant decrease in earnings response coefficients (ERC) after the reverse split announcements. These results suggest that reverse splits provide information about the permanence of past and future earnings performance.

1. INTRODUCTION

Reverse splits are unusual events in which a corporation substitutes one new share for multiple outstanding shares, which proportionately reduces the number of shares outstanding, proportionately increases the par value of the shares,[1] and increases the market price of the shares by approximately the same factor. Reverse splits are generally associated with firms whose

Advances in Accounting, Volume 23, 179–205

ISSN: 0882-6110/doi:10.1016/S0882-6110(07)23007-X

stock has been trading at low prices for an extended period. Managers of firms proposing reverse splits often claim that reverse splits will reduce transactions costs, enhance the firms' image among institutional investors, reduce shareholder servicing costs, and also enable firms to forestall delisting from exchanges that require a minimum share price to maintain listing. Some of these claims have empirical support in the literature. For example, Han (1995) documents increased liquidity, Masse, Hanrahan, and Kushner (1997) document increased marginability,[2] and Dravid (1987), Lamoureux and Poon (1987), and Peterson and Peterson (1992) document decreased risk following a reverse split. These are all value-enhancing benefits that should ostensibly elicit a positive market reaction to the news of a reverse split. However, previous studies have consistently documented a negative market reaction to news of impending reverse splits (Radcliffe & Gillespie, 1979; Woolridge & Chambers, 1983; Spudeck & Moyer, 1985; Peterson & Peterson, 1992; Han, 1995).[3]

This study examines whether reverse split abnormal announcement returns are synonymous with continued poor price and earnings performance. My analysis is based on the conjecture that managers expecting future earnings improvements would not need to incur the additional costs associated with reverse splits to increase stock price.[4] Thus, reverse splits implicitly communicate that managers do not expect earnings increases and associated stock price increases in the near future. Several studies suggest that reverse splits portend bad news about future earnings prospects, thus generating a negative market reaction (West & Brouilette, 1970; Radcliffe & Gillespie, 1979; Spudeck & Moyer, 1985), but to date, there is no direct evidence concerning this potential explanation for the negative market reaction to reverse split announcements. The purpose of this study is to provide such evidence.

I investigate whether reverse split announcements provide information about the persistence of past earnings performance and the likelihood of future earnings declines by examining whether pre- and post-split abnormal earnings are positively correlated with reverse split announcement returns. I estimate a regression of cumulative abnormal stock returns over the four days centered on reverse split announcements on earnings changes in the six years surrounding the reverse split. Second, I examine whether security analysts revise their earnings forecasts after these events. If reverse splits signal unfavorable future earnings, then security analysts monitoring these firms would be expected to revise their earnings forecasts downwards after split announcements. I examine the mean and median long-term forecast revisions centered on these reverse split events, the number of positive to negative forecast revisions, as well as whether the forecast revisions are correlated with the abnormal returns. Finally, I examine whether reverse split firms have lower earnings response coefficients (hereafter ERCs) after the reverse split. I hypothesize that the reverse split (a pre-announcement disclosure), provides information about future earnings changes, (information which is highly correlated with the information in future earnings), which likely preempts some of the information in subsequent earnings releases. Thus, reverse split announcements reduce investors' reliance on the firm's earnings announcements, which in turn reduces the ERC after the reverse split.[5] To investigate whether the price–earnings relation declines after a reverse split I regress annual abnormal stock returns on contemporaneous earnings changes two years before and after the reverse split.

I find that reverse splitting firms on average, experience significant losses in the three years before the announcement year, losses which appear to be non-transitory since they continue up to two years after the reverse split. Consistent with research expectations, I find a significantly positive relation between reverse split announcement returns and earnings changes in the years surrounding the reverse split. Thus, reverse split announcements communicate that pre-split losses are not transitory and future losses are probable. This interpretation is further supported by evidence that the mean and median long-term analyst forecast revision is significantly negative and positively correlated with reverse split announcement returns. Thus, these more sophisticated investors also view reverse splits as signals of future earnings declines. Finally, I find evidence that the price–earnings relation is weaker in the post-split years (smaller post-split ERCs) implying that reverse splits are associated with less (incrementally) informative subsequent earnings releases. This suggests reverse split information preempts the information in subsequent earnings announcements.

This study makes several important contributions. First, the study increases our understanding of reverse splits and documents that the market's perception of the reverse split is conditioned on the splitting firms' financial performance in the years surrounding the event.[6] On average, firms with a more favorable market reaction to the split announcement generally perform better pre- and post-split (earnings, return on assets, and price) than firms with less favorable market reaction. These firms are also more likely to maintain exchange listing in the years following the reverse split. Thus, market participants were able to determine at the reverse split date which firms were more likely to improve performance post-split. This evidence contributes to the literature by formally documenting that these non-earnings events convey information about firms' future performance.

Second, the evidence is consistent with investors decreasing the weight they place on earnings subsequent to the reverse splits suggesting that investors use reverse split information about earnings to update their expectations about future cash flows. Thus, reverse split information appears to improve earnings expectations formed by investors so that there is less ''surprise'' in subsequent earnings releases.

The remainder of this chapter is organized as follows. Section 2 develops the research hypotheses in the context of prior research. Section 3 describes the sample selection procedure, data sources, variable specification, and testing methodology. Section 4 provides descriptive statistics for the sample firms and discusses the results of empirical tests. Section 5 presents the conclusions.

2. LITERATURE REVIEW AND HYPOTHESES DEVELOPMENT

2.1. Why do Firms Undertake Reverse Split?

The number of firms initiating reverse stock splits has increased dramatically over the past ten years with the 1990s alone accounting for 1093 or 68% of the total reverse split from 1962 to 2000.[7] Managerial incentives for using reverse splits, and investors' reactions to reverse split announcements are not well understood. Empirical evidence suggests that reverse splits impart some tangible benefits to firms. Han (1995) finds evidence using a sample of 136 reverse splits over 1963–1990 that bid-ask spreads and the number of non-trading days decrease and trading volume increases following a reverse split. Masse et al. (1997) document improved marginability (ability of the stock to be purchased on margin) after a reverse split. Finally, Peterson and Peterson (1992) report decreased risk following a reverse split. These findings support managers' claims that they use reverse splits to improve liquidity and marginability of the shares and to reduce trading costs, thus making them more attractive to the investing public.[8]

The use of a reverse split may have both psychological and practical implications. There is strong anecdotal and institutional evidence that low stock prices are sub-optimal for many firms. Managers of firms with a low-priced stock are attempting to manage the institutional perception that all low-priced stocks (''penny stocks'') are speculative.[9] Several institutional investors have guidelines that disfavor stocks trading below \$5.00 per share.[10] These investors may avoid low-priced stocks for practical reasons such as increased transaction costs. This sentiment is consistent with the findings of Bhushan (1994) who documents that institutional investors shun low-priced stocks citing indirect costs such as price pressure effects and delays in filling an order as important determinants of their trading and arbitrage activities.

Additionally, the structure of trading commissions tends to have an adverse effect on holders of low-priced securities because the brokerage commission on the sale of a low-priced stock generally represents a higher percentage on the sales price. Bhardwaj and Brooks (1992) present evidence that proportional bid-ask spreads (difference between bid and ask prices divided by the average of the two) vary inversely with the level of the stock price. For example, the median estimated bid-ask spread for firms with share prices up to \$5 is 5.1%, and for firms with prices greater than \$20 it is only 0.8%. Bhardwaj and Brooks (1992) also report median brokerage commission rate for individual investors of 7.4% for share price up to \$5 and 1.3% for share price greater than \$20. Thus, transaction costs are likely to decrease following reverse splits making these firms more attractive to institutional investors.

A low stock price may also affect the ability of investors to purchase stocks on margin. Many brokerage firms will not make margin loans on stocks trading below \$5 (long position) and stocks trading below \$10 (short position). Thus, fund managers state that they typically do not consider investing in low-priced stocks. Finally, a low share price can also hinder a firm's visibility in the financial press. A number of financial publications limit their published quotations of listed securities to lists comprised of higher priced securities. For example, the Wall Street Journal (WSJ) imposes a \$3.00 requirement for reporting a company's common stock on the "National" listing of NASDAQ System stocks. This potentially reduces the visibility and thus the liquidity of the company's stock.

As a practical matter, managers may use reverse splits to increase current stock prices even though they anticipate future earnings increases, in order to maintain listing on an exchange. Some exchanges have a minimum price requirement for initial and continued listing and usually recommend that a firm reverse split if its stock trades at a low price for extended periods. The New York Stock Exchange (NYSE) does not have a minimum required share price but its continued listing rules include a provision to consider delisting in the case of an "unusually low price" (*NYSE Company Guide*, Section 802).[11] The American Stock Exchange (AMEX) requires a minimum price of $3 per share (AMEX Company Guide, Section 102, paragraph (b)) while the National Association of Securities Dealers (NASDAQ) exchange requires that the bid price be at least $5 per share for initial listing and $1 for continued listing (National Association of Securities Dealers (1992)). If the stock trades for less than the minimum price for ten consecutive business days then NASDAQ enforces this policy for low-priced securities by recommending a reverse split. The firm is given ninety days in which to comply with the exchange requirements. Failure to comply with this requirement results in delisting from the NASDAQ system, limiting trading in the company's shares to the NASD's over-the-counter (OTC) Bulletin Board (pink sheets), generally considered a less efficient market than the NASDAQ. Thus, firms may use a reverse split to maintain exchange listing.

2.2. Market Perception of Reverse Splits

Within the investment community, there is a lack of consensus on what information reverse splits convey. Some analysts suggest that reverse splits remove the stigma of a low price, thus raising interest in a company's stock among brokers and investors. However, most question the value of a reverse split. Bill Montague of USA Today suggests that because companies that reverse splits often have deeply troubled balance sheets, the ploy of increasing the stock price through a reverse split often does not change investor perception.[12] John Lewis IV, president of Gardner Lewis Asset Management, echoes what many analysts believe, "… reverse splits might lengthen a company's runway, but if it isn't accompanied … by a credible recovery plan it carries little weight with investors."[13] This view is consistent with the negative market response to reverse split announcements.

2.3. Prior Research on Reverse Splits

Empirical research on the relation between reverse splits and firms' earnings has been relatively sparse. West and Brouilette (1970) find that more than half of the reverse splitting firms in their sample report losses prior to and subsequent to the reverse split, and had actual stock prices that were lower than predicted after the reverse split. Thus, they conclude "that the stock market's reaction to a reverse split tends to be closely related to a company's earnings history and, more importantly, its earnings outlook."

Radcliffe and Gillespie (1979) document a mean cumulative abnormal return of –10.59% in the reverse split month but find no difference in the price reaction for dividend-increasing versus dividend-constant firms and dividend-paying versus non-dividend-paying firms. They conclude that the price impact of reverse splits cannot be explained on the basis of their implications for dividend policy. Woolridge and Chambers (1983) find that relative earnings performance influences shareholder returns on the announcement dates, with better (poorer) performing firms yielding smaller (larger) absolute negative returns. In summary, these studies document that reverse split announcements are generally accompanied by a negative market reaction, suggesting that the announcements convey unfavorable information to the market.

More recently, the evidence suggests that reverse splits impart some tangible benefits such as increased liquidity, increased marginability, lower bid-ask spreads, reduced number of non-trading days and increased trading volume (Han, 1995; Masse et al., 1997), and decreased risk (Peterson & Peterson, 1992). In light of these positive benefits, the documented average negative abnormal announcement returns suggest that reverse splits convey negative information to the market. I argue that managers implicitly convey information that pre-split losses are not transitory and that future earnings declines are likely. Rationally, managers expecting future earnings and stock price improvements would not incur the costs associated with reverse splits. Thus, reverse splits signal unfavorable information about future earnings performance. Reverse-splitting managers may have no explicit desire to signal and may simply wish to return the stock to what they believe is a better trading range, or to maintain listing on the current exchange. However, investors implicitly infer that the reverse split signal managers' beliefs that earnings and stock prices will continue to decrease. Grinblatt, Masulis, and Titman (1984) argue that (splits) may be a less costly means of communicating unfavorable earnings prospects than an explicit negative management earnings forecast since it is more ambiguous and thus likely to subject mangers to less litigation risks if the information inferred by the market turns out to be incorrect.

If reverse splits signal that past and current earnings and price performance are likely to be permanent and that additional declines are likely in the future,[14] then the market reaction to the announcement of a proposed reverse split should be correlated with earnings performance before and after the split. Therefore, I examine the following testable hypothesis (in alternative form):

> **H1.** Abnormal returns around reverse splits announcement dates are positively related to pre-split and post-split earnings changes.

Hypothesis 2 states that if reverse split announcements signal unfavorable future earnings performance, then security analysts monitoring these firms would be expected to revise their earnings forecasts downwards on the news of an impending reverse split. Therefore, I examine the following hypothesis (in alternative form):

> **H2.** Security analysts revise their earnings forecasts downwards for firms that announce reverse stock splits and the forecast revision is positively correlated to the abnormal announcement returns.

2.4. Reverse Splits and the Returns–Earnings Relation

I also examine whether the information conveyed by reverse splits reduces investor reliance on earnings information after the reverse split. In the Holthausen and Verrecchia (1988) (HV) model, informed investors rely on both financial and non-financial information in making investment decisions. In the context of earnings announcements, investors most likely weigh the value of the earnings information conditional on the value of any pre-disclosed (financial and non-financial) information. The sensitivity of the market response to earnings information is determined by the relative precision of the earnings information provided by the firm and the pre-disclosed non-financial information signal, as well as investors' fundamental uncertainty about firm value. HV posits that a pre-disclosure information signal ($\tilde{y}_1$) provides information to investors about earnings and future cash flows. Thus, investors weigh the precision of the earnings signal ($\tilde{y}_2$) relative to that of the non-financial, pre-disclosed signal ($\tilde{y}_1$) in order to evaluate the incremental informational relevance of earnings. In this study, I use the earnings response coefficient (ERC) to measure the usefulness of earnings information. Prior research has used the earnings response coefficient (the effect of \$1 of unexpected earnings on security returns) to measure the response of market participants to earnings announcements and thereby infer the information content of accounting earnings. The HV model predicts that the ERC is increasing in the uncertainty or noise in pre-disclosure information (n_1), and decreasing in (c), the correlation between the earnings signal ($\tilde{y}_2$), and the non-financial pre-disclosed signal ($\tilde{y}_1$). Using the theoretical arguments from the HV model I posit that a reverse split provides investors with some information about future cash flows that is highly correlated with information in earnings (high c). This information effectively preempts subsequent earnings announcements and therefore causes a decrease in the informational relevance of firms' earnings following the reverse split event. Thus, I test the following hypothesis stated in alternative form:

H3. The correlation between annual abnormal stock return and contemporaneous earnings change is stronger in the pre-split years and weaker in the years subsequent to the reverse split.

3. RESEARCH METHODOLOGY

3.1. Testing the Relation between Split Announcement Returns and Earnings Changes

To investigate whether reverse split announcements are associated with earnings, I test the following model:

$$\begin{aligned}\mathrm{CAR}_{it} = \beta_0 &+ \beta_1 \mathrm{SUEPS}_{it-2} + \beta_2 \mathrm{SUEPS}_{it-1} + \beta_3 \mathrm{SUEPS}_{it0} \\ &+ \beta_4 \mathrm{SUEPS}_{it+1} + \beta_5 \mathrm{SUEPS}_{it+2} + \beta_6 \mathrm{SUEPS}_{it+3} + \varepsilon_{it} \qquad (1)\end{aligned}$$

where: CAR_{it} is the market-adjusted return for firm i for the four days (–1, +2) surrounding the announcement. The market-adjusted return is the cumulative abnormal return calculated as the sum of the difference between the raw return R_{it}, the return on firm i's security on day t, and the returns for the equally weighted market index. The market index is firm-specific so that the CRSP NYSE-AMEX index file is used for NYSE- and AMEX-traded firms and the NASDAQ index for NASDAQ-traded firms.

$$\mathrm{SUEPS}_{i,t} = \frac{\mathrm{EPS}_{i,t} - \mathrm{EPS}_{i,t-1}}{\mathrm{P}_t}$$

is the standardized (by closing stock price in event year) earnings change in year t. Consistent with prior research, earnings changes proxy for unexpected earnings. Earnings are annual earnings per share before extraordinary items and discontinued operations (restated) obtained from the Annual Compustat tapes.[15] EPS is income before extraordinary items divided by common stock outstanding.[16] This approach is consistent with

Asquith, Healy, and Palepu (1989) study of stock splits. I re-estimate Eq. (1) including control variables such as firm size and split factor as additional explanatory variables. Firm size is measured as the market value of equity (MVE) in the year before the reverse split. Firm size proxies for pre-split information and is predicted to be negatively related to announcement returns. Split factor is a measure of the size of the reverse split and is measured as:

$$\text{FACTOR} = \left(\frac{\text{the number of shares outstanding after the split}}{\text{the number of shares outstanding before split}}\right) - 1$$

Thus, as factor decreases, the size of the reverse split increases. I predict a negative relation between factor and announcement returns since there is a negative relationship between the factor and the pre-split price.

3.2. Testing Analysts Forecast Revisions Surrounding Reverse Split Announcements

To investigate whether security analysts revise their earnings forecasts after reverse split announcements, I compare pre-split earnings forecasts and post-split I/B/E/S earnings forecasts. Consistent with Doran (1994), analyst forecast revision (REVISE%) is:

$$\text{REVISE\%}_{j,t} = \left[\frac{\text{Post-Split Forecast-Pre-Split Forecast}}{\text{Pre-Split Forecast}}\right]$$

I calculate the mean and median REVISE% as well as the number of negative/positive REVISE%. I also examine the correlation between REVISE% and CAR_{jt}.

$$\text{CAR}_{jt} = \lambda_0 + \lambda_1 \text{REVISE\%}_{jt} + \varepsilon_{jt} \tag{2}$$

3.3. Testing the Relation between Returns and Earnings Changes Around Reverse Splits

To investigate whether the price–earnings relation (ERC) weakens in the post-split period, I examine the relation between annual abnormal returns and contemporaneous earnings changes for years centered on the stock split using the following regression:

$$\mathrm{CAR}_{jt} = \lambda_0 + \lambda_1 \Delta\mathrm{EPS}_{jt} + \lambda_2 D_1 \Delta\mathrm{EPS}_{jt} + \lambda_3 D_2 \Delta\mathrm{EPS}_{jt} + \varepsilon_{jt} \qquad (3)$$

where $\mathrm{CAR}_{j,t}$ is the annual market-adjusted return for firm j in year t defined as the difference between the firm's return and the return on the equally weighted market index. Consistent with Asquith et al. (1989), annual returns for year –1 is the cumulative return for the year ending two weeks after the annual earnings announcement immediately prior the reverse split. I obtain earnings announcement dates from quarterly EPS Compustat data. ΔE_{jt} is the change in EPS for firm j in year t deflated by closing stock price in the year prior to the reverse stock split. D_1 is an indicator variable equal to one in years 1 and 2 (post-split years), and zero in all other years and D_2 is an indicator variable taking on a value of one in year –2 through year 2 when EPS_{jt} is negative, and zero otherwise.

4. DATA DESCRIPTION AND EMPIRICAL RESULTS

The initial sample consists of all reverse splits reported on CRSP during the 1962 through 1998 period. Sample observations must have the split announcement recorded in the WSJ, and have no other events reported over the four-day event window. Returns and earnings data must be available from CRSP and Compustat databases, respectively. These criteria result in a final sample of 360 firms. Sample firms include firms listed on the NYSE, AMEX, and NASDAQ exchanges.

Table 1 reports summary descriptive statistics for the final sample. Panel A of Table 2 documents that reverse splits are an increasing phenomenon. Even though the sample period spans 4 decades, 77.0% of the 360 sample reverse splits occurred between 1990 and 1999 alone. Panel B of Table 1 reports data on the magnitude of the reverse split factors. Approximately 62% of the reverse splits have a one-for-five or lower split factor, with more extreme reverse splits (one-for-six and above) making up 38% of the sample. One-for-ten reverse splits are the most common split factor followed by one-for-five and one-for-four factors, combining to make up approximately 51% of the sample. This result is consistent with the view that reverse split firms are trading at very low prices and therefore require larger split factors to achieve relatively higher prices after the reverse split.

Panels C and D of Table 1 provide support for anecdotal evidence that reverse splits are primarily undertaken by small firms trading on over the counter markets. Approximately 64% of the reverse splitting firms trade on the NASDAQ exchange, with only 35% of firms trading on the NYSE and AMEX exchanges. Panel D also documents that approximately 62% of the sample is attributable to firms with MVE in the lower half of the MVE distribution for all CRSP firms. Panel E of Table 1 reports industry distribution of the reverse split sample. Although sample reverse stock splits represent eight broad industries, two industries (manufacturing and services) account for 59% of the total number of reverse splits.

In Table 2 Panel A reports that the average reverse splitting firm has assets of \$793 million in the year of the reverse split. The median asset size of \$29 million, however, suggests that the majority of firms are extremely small. The mean (median) MVE of \$257 (\$14.3) million in the reverse split year is much larger than that reported in prior studies such as Radcliffe and Gillespie (1979) and Han (1995), which use much smaller samples of reverse splits. The firms in this sample typically have relatively low cash holdings with a median of \$1.9 million in the event year. Panel B of Table 2 reports earnings data for sample firms over the period surrounding the reverse split. On average reverse split firms suffered losses over the three years before the reverse split. Earnings in these years are each significantly less than zero ($p<.0001$, two-tail test). Earnings performance improved in year $+3$ after the reverse split. Industry-adjusted earnings are all significantly negative, showing that sample firms had earnings in the pre- and post-split period significantly below the industry average over this period. In summary, reverse split sample firms experience poor earnings performance. These results are consistent with the findings of West and Brouilette (1970) and Woolridge and Chambers (1983). Finally, Panel C of Table 2 reports dividend-payout behavior of the sample firms. Approximately 12% (42) of the 360 firms paid a dividend in the event year similar to the 10% reported by Woolridge and Chambers (1983). Both the number of firms paying dividends and the number of firms increasing dividends decreased in the years leading up to the reverse split, but this pattern reversed after the reverse split. The mean dividend in the event year was \$0.43.

Price data reported in Panel B of Table 3 suggest that median-adjusted closing stock prices were significantly lower for reverse split firms than for firms in the same size decile in the years before and just after the reverse split, but were not significantly different by the third year after the reverse split. Data in Table 2 reveal that on average, reverse split firms experience a monotonic decrease in the fiscal year end closing prices in the years leading up to the reverse split, followed by increases in fiscal year end closing prices in the years after the reverse split. The mean (median) fiscal year end closing

Table 1. Description of the Sample of Reverse Splits.

Panel A: Incidence of reverse splits over time

Time Period	Number of Reverse Splits	Percentage of Sample
1964–1969	5	1.4
1970–1979	22	6.1
1980–1989	56	15.6
1990–1998	277	76.9
Total	360	100.0

Panel B: Incidence of reverse splits by split factor

Split Factor	Number of Reverse Splits	Percentage of Sample
1 for 5 and lower	222	61.7
>1 for 5	158	38.3
Total	360	100.0

Panel C: Distribution of reverse splits by exchange listing

Stock Exchange	Number of Reverse Splits	Percentage of Sample
New York	65	18.0
AMEX	62	17.2
NASDAQ	231	64.2
Other	2	0.5
Total	360	100.0

Panel D: Size decile classification for final sample

Size Decile Classification	Number of Reverse Splits	Percentage of Sample
1–5	224	62.2
6–10	136	37.8
Total	360	100.00

Panel E: Distribution of reverse splits by industry classification

Broad Classification	Two-Digit SIC Code	Industry	Number of Firms	Percentage of Sample
Mining	10–16	Oil & gas extraction	52	14.4
Manufacturing	20–38	Manufacturing	141	39.2
Transportation, communication utilities	45–49	Transportation, communication utilities	16	4.4

Table 1. *(Continued)*

Panel E: Distribution of reverse splits by industry classification

Broad Classification	Two-Digit SIC Code	Industry	Number of Firms	Percentage of Sample
Wholesale trade	50–51	Durable & nondurable goods, wholesale	26	7.2
Retail trade	56–59	Apparel & accessory stores/jewellers/etc	14	3.9
Finance, ins. & real estate	61–67	Lending institurions/ health services/ insurance carriers	38	10.6
Services	70–87	Health and retirement communities	73	20.3
		Total	360	100

price in the event year of $3.69 ($1.50) is higher than the Radcliffe and Gillespie (1979) result of $2.86. Subsequent to the reverse split, the mean stock price increased to $8.82 by the end of the third year. Finally, Panel A of Table 3 presents evidence that some reverse split firms were able to successfully move their stock prices to higher price intervals after the reverse split, that is, to stock prices more consistent with similar-sized firms.

In Table 4 Panel A reports market-adjusted returns (MAR) on the reverse split announcement date for the full sample of 360 reverse splits. The MAR over the four days surrounding the reverse split announcement is –4.89% ($p<.0001$, two-tail test). These results are similar to those of Han (1995) with an average abnormal return of –4.69% for a sample of 136 reverse splits, and to the 4.40% abnormal return reported by Asquith et al. (1989) for 121 stock splits. Results are similar when mean-adjusted and size-adjusted abnormal returns are used.[17] Thus, these results confirm the findings from prior research that the market reacts negatively to reverse split announcements.

In Table 4 Panels B and C present results of tests of the market reaction to the reverse split announcement based on whether the firm did/did not pay a dividend and whether the firm sustained a loss or a profit in the year before the split. The results suggest that the firm's earnings performance in the pre-split year as well as dividend-payout behavior significantly affected

Table 2. Descriptive Statistics for Years Surrounding Announcements of Reverse Stock Splits for Full Sample.

Panel A: Descriptive statistic for sample

Year Relative to Reverse Split	Number of Firms	Mean Asset[a] (Median) Millions $	Mean Closing[b] Price (Median) $	Mean Cash (Median) Millions $	Market Value of Equity (Median) Millions $	Mean Market-To Book (Median)
−3	360	743.6***	4.59***	81.1**	251.0***	3.6***
		(23.5)	(2.13)	(1.93)	(17.3)	(1.4)
−2	360	775.4***	4.11***	81.2**	242.8***	2.7***
		(24.1)	(1.63)	(1.77)	(15.9)	(1.3)
−1	360	760.2***	3.60***	78.13**	243.2***	2.8***
		(26.6)	(1.56)	(1.79)	(15.3)	(1.4)
0	360	792.6***	3.70***	87.7**	256.9***	3.7***
		(29.1)	(1.50)	(1.87)	(14.3)	(1.1)
1	360	752.1***	6.59***	74.9**	237.0***	9.1
		(34.8)	(3.60)	(2.29)	(12.9)	(1.0)
2	360	752.9***	8.45***	71.8**	284.0***	2.0***
		(38.9)	(4.81)	(2.45)	(18.3)	(1.2)
3	360	803.6***	8.82***	73.7**	338.2***	1.5***
		(43.1)	(4.88)	(2.93)	(20.4)	(1.2)

Table 2. (*Continued*)

Panel B: Earnings changes

Year Relative to Reverse Split	Number of Firms	Mean EPS[a] (Median) $	Median Industry-Adjusted EPS[b] $	Mean Standardized Earnings Changes
−3	360	−0.271***	−0.637***	0.059
		(−0.035)	(−0.32)	(0)
−2	360	−0.316***	−0.665***	−0.200
		(−0.035)	(−0.34)	(0)
−1	360	−0.419***	−0.765***	−0.124
		(−0.05)	(−0.34)	(0)
0	360	−0.292***	−0.648***	0.214**
		(−0.03)	(−0.27)	(0.026)
1	360	−0.644***	−1.023***	−3.082**
		(−0.02)	(−0.29)	(0.011)
2	360	−0.012	−0.391***	3.090*
		(0.06)	(−0.30)	(0.053)
3	360	0.038	−0.365***	0.353*
		(0.110)	(−0.22)	(0.058)

Panel C: Summary statistics on dividend-paying behavior in years surrounding announcements of reverse stock splits [a,b]

Year Relative to Reverse Split	Number of Dividend Paying Firms	Mean (Median) Dividend $	Number of Dividend-Increasing Firms	Number of Dividend-Decreasing Firms
−3	55	0.5219***	27	61
		(0.3595)		
−2	48	0.5803***	26	33
		(0.3843)		

−1	47	0.3752^{***} (0.3000)	23	34
0	42	0.4334^{***} (0.2982)	24	28
1	46	0.8118^{***} (0.6062)	36	17
2	41	0.5187^{***} (0.4275)	28	25
3	46	0.5561^{***} (0.4764)	37	14

Note to Panel C: Student *t*-test statistics – mean earnings changes are not significantly different from zero.
The probability levels are for two-tailed tests of significance.
*Significant at 10%.
**Significant at 5%.
***Significant at <1%.
[a]The reverse split sample comprises 360 firms that announced reverse splits over the 1964 to 1999 period. Industry-adjusted earnings for a given firm in year *t* are defined as the difference between the raw earnings for the reverse split firm in year *t* and the median industry earnings in the same year for the same 2-digit SIC codes.
[b]Earnings are annual earnings per share before extraordinary items and discontinued operations (restated) obtained from the Annual Compustat tapes. Annual earnings announcement dates obtained from the quarterly tapes are used to center the earnings changes relative to the reverse split announcement. Changes in earnings per share before extraordinary items and discontinued operations are standardized by the closing stock price in the year of the reverse split.

Table 3. Descriptive Statistics for Final Sample of Reverse Stock Splits over 1964–1998.

Panel A: Number of firms in final sample in various price categories before and after event

Trading Range of Stock Price	Number of Firms in Price Intervals						
	P<$1	1<P<2	2<P<3	3<P<4	4<P<5	P>5	Total
YEAR BEFORE REVERSE SPLIT	127	93	42	27	17	54	360
YEAR AFTER REVERSE SPLIT	37	48	40	33	28	174	360
% Change	−71%	48%	−5%	+22%	+65%	+222%	

Panel B: Median decile-adjusted price before and after reverse split event

Year Relative to Reverse Split	Number of Firms	Median Size Decile-Adjusted Price $	*t*-Value	Probability
−3	360	−4.35	−9.42	<.0001
−2	360	−4.66	−9.26	<.0001
−1	360	−5.21	−10.96	<.0001
0	360	−5.24	−10.05	<.0001
1	360	−2.56	−6.34	<.0001
2	360	−0.81	−1.31	0.191
3	360	−0.49	−0.96	0.336

Note: Median size-decile adjusted prices are the firm's stock price for a particular year minus the median stock price for all firms in the same size decile.

the market's interpretation of the reverse split. The market's reaction is significantly different for profit and loss firms as well as for dividend-paying and non-dividend-paying firms. In fact, the abnormal announcement returns are most negative for firms with losses that pay no dividends and then least negative for firms with positive income in the year before that pay a dividend. Thus, the market appears to use earnings performance as well as dividend payout behavior to interpret the signal implied in the reverse split. This is consistent with the predictions of H1.

In Table 5 Panel A reports the results from regressing reverse split announcement returns on pre- and post-split earnings changes for the sample of 360 firms. The estimated coefficients for the pre- and post-split earnings changes for years –2 through +3 are significantly different from zero at greater than the 1% level. These estimates indicate that, all else being constant, as standardized earnings increase (decrease) MAR increase (decrease) in the four days surrounding the reverse split announcement. These results are consistent with Hypothesis 1. Thus, the information released by a reverse split is positively correlated with earnings changes two

Table 4. Summary Statistics on Returns for Days Surrounding Announcements of Reverse Stock Splits for Full Sample[a,b].

Panel A: Cumulative abnormal returns (CAR)

Days Relative to Reverse Split	Number of Firms[c]	Mean	Student *t* Probability	First Quartile	Median	Third Quartile
CAR0 (0)	360	−0.035	−6.02***	−0.073	−0.025	0.014
CAR2 (−1, 0)	360	−0.032	−4.91***	−0.073	−0.016	0.022
CAR3 (0, +2)	360	−0.049	−6.69***	−0.113	−0.034	0.031
CAR4 (−1, +2)	360	−0.046	−5.83***	−0.115	−0.032	0.031

Panel B: Cumulative abnormal returns (CAR) conditional on firm profitability and dividend-paying behavior

Firm Profitability/Dividend Behavior	*N*	Median CAR	
Loss/no dividend	186	−0.051****	Lowest
Loss	197	−0.048***	
No dividend	318	−0.040****	
Profit/no dividend	132	−0.027**	
Profit	163	−0.024**	
Loss/dividend	11	−0.015	
Dividend	42	−0.011*	
Profit/dividend	31	−0.011	Highest

Panel C: Probability for one-sided tests of differences

Firm Profitability/Dividend Behavior	*N*	Probability *T* of Differences in Means
Loss: Profit	197:163	0.020**
Dividend: No dividend	42:318	0.004**
Profit/dividend: Profit/no dividend	31:132	0.355
Loss/dividend: Loss/no dividend	11:186	0.003**

Notes:

[a]The reverse split sample comprises 360 firms that announced reverse splits from 1964 to 1998 period. Market-adjusted returns are the raw returns for firm *j* over the event period minus the returns for the CRSP equally weighted market index for Amex and NYSE firms and the NASDAQ index for Nasdaq-traded firms.

[b]The holding periods are for days relative to the CRSP record date of the reverse split.

[c]Panel A report results of tests that the mean abnormal retruns are not different from zero. Panel B reports tests of differences in medians. The statistics are based on nonparametric tests such as Wilcoxon and Krushkal–Wallis tests. Panel C reports the probability of *t*-test of differences in means of the various subgroups.

Probability levels are based on two-tailed tests of significance.

*Significance at 10% level.

**Significance at 5% level.

***Significance at 1% level.

Table 5. Tests of the Relation between Reverse Split Announcement Returns and Earnings Changes Surrounding the Reverse Split.

Model:

$$\mathrm{CAR}_{it} = \beta_0 + \beta_1 \mathrm{SUEPS}_{it-2} + \beta_2 \mathrm{SUEPS}_{it-1} + \beta_3 \mathrm{SUEPS}_{it0} + \beta_4 \mathrm{SUEPS}_{it+1} + \beta_5 \mathrm{SUEPS}_{it+2} + \beta_6 \mathrm{SUEPS}_{it+3} + \varepsilon_{it} \quad (1)$$

Panel A: Coefficient estimates, t-statistics, and R^2

Parameter	Coefficient	t-Statistic	Probability
β_0	−0.048	−6.77	0.000
β_1	0.012	4.12	0.000
β_2	0.020	4.54	0.000
β_3	0.011	3.41	0.001
β_4	0.009	3.97	0.000
β_5	0.009	3.58	0.000
β_6	0.009	3.23	0.001
R^2	0.102		
Ad. R^2	0.087		
N	360		

Model:

$$\mathrm{CAR}_{it} = \beta_0 + \beta_1 \mathrm{SUEPS}_{it-2} + \beta_2 \mathrm{SUEPS}_{it-1} + \beta_3 \mathrm{SUEPS}_{\mathrm{it}0} + \beta_4 \mathrm{SUEPS}_{it+1} + \beta_5 \mathrm{SUEPS}_{it+2} + \beta_6 \mathrm{SUEPS}_{it+3} + \beta_7 \mathrm{SIZE}_i + \beta_8 \mathrm{FACTOR}_i + \varepsilon_{it} \quad (2)$$

Panel B: Coefficient estimates, t-statistics, and R^2

Parameter	Coefficient	t-Statistic	Probability
β_0	0.019	0.43	0.665
β_1	0.011	3.89	0.000
β_2	0.019	4.35	0.000
β_3	0.012	3.58	0.000
β_4	0.009	4.03	0.000
β_5	0.009	3.66	0.000
β_6	0.007	3.12	0.002
β_7	0.004	0.91	0.364
β_8	0.102	2.08	0.038
R^2	0.120		
Ad. R^2	0.100		
N	360		

Notes: CAR_{it} is the cumulative market-adjusted return for firm i during the four-day measurement period ending two days after the reverse split announcement date (−1 to +2). Market-adjusted returns are the raw returns for firm j over the event period minus the returns on a CRSP equally-weighted market index.

N = the number of firms used to estimate the parameters of the model.

Firm size = market value of equity in the year before the reverse split. Split factor is a measure of the size of the reverse split and is measured as:

$$\mathrm{FACTOR} = \left(\frac{\text{the number of shares outstanding after the split}}{\text{the number of shares outstanding before split}}\right) - 1$$

years before to the reverse split, in the year of the reverse split, and three years after the reverse split. These findings suggest that reverse split announcements convey information to investors about the permanence of pre-split earnings changes, and some information about future earnings performance. The significance on the pre-split earnings variables are consistent with the findings of Asquith et al. (1989), while the significance on the post-split variables are consistent with Pilotte and Manuel (1996), Pilotte (1997), Rankine, and Stice (1997), and Ikenberry, Rankine, and Stice (1996) for forward splits. Table 5 Panel B re-estimates Eq. (1) with firm size and split factor as controls. The split factor variable is significant and positive at the .05 level of significance suggesting that as the size of the reverse split increase the associated abnormal announcement returns decrease. MVE is not significant in explaining announcement returns.

In Table 6 Panels A–D report the results for those reverse split firms with analysts' forecast data in the period around the reverse split. The mean (median) pre-split forecast of EPS is \$1.32 (0.95), while the mean (median) post-split forecast is \$0.68 (0.55) all significant at $<.0001\%$. The mean (median) forecast revision for 111 firms is –\$0.28 (–0.18), which is also significant at $<.0001\%$. Mean (median) forecast revisions for one-year and two-year ahead forecasts are –0.23 (–0.14), and –0.32 (–0.25), respectively, all highly significant. These results are similar to those reported by Doran and Nachtmann (1988) and Doran (1994) for stock splits. Further, the number of negative/positive or zero forecast revisions (21–90) is significant at <0.0001 (χ^2 statistic $=42.89$). Finally, similar to Klein and Peterson (1989) and McNichols and Dravid (1990) for stock splits, the median analyst forecast revision is positively correlated with abnormal announcement returns (at the .08% level). This evidence is therefore consistent with security analysts inferring that reverse split announcements are signals of future earnings declines, thus supporting H2.

Table 7 presents results from the estimation of Eq. (2) when the indicator variable takes on the value of 0 for years –2, and –1, (pre-split period) and 1 for years 1–2 (post-split period). The estimate for λ_1, the average price to earnings elasticity for years –2 to –1, is positive and significantly different from zero at the 1% level (two-tail test) indicating that on average there is a significant association between price and earnings changes.[18] The estimate for λ_1 of 0.05 implies that a 5% decrease in earnings is accompanied by a 1% decrease in stock price. The estimate for λ_2, the *change* in price to earnings elasticity in years 1–2 relative to years –2 to –1, is negative and significant at the 1% level (two-tail test) suggesting that in the two years after a reverse split, the relation between earnings and price changes is significantly lower

Table 6. Analyst Forecast Revisions for Some Reverse Splitting Firms.

Panel A: Pre-split and post-split forecasts and forecast revisions centered on reverse split dates

Variable	N	Mean	LQ	Median	UQ	t	Pr<|t|
Pre-split forecast	111	1.324	0.250	0.948	1.85	9.32	0.000
Post-split forecast	111	0.678	0.050	0.555	1.35	8.71	0.000
Revise%	111	−0.258	−0.519	−0.136	0	−5.04	0.000

Panel B: Analysts long-term forecast revisions centered on reverse split dates

Median one-year ahead forecast revision	61	−0.232	−0.474	−0.143	−0.0004	−4.41	0.000
Median two-year ahead forecast revision	54	−0.321	−0.800	−0.251	−0.0185	−3.19	0.002

Panel C: Sign of forecast revisions centered on reverse split dates

Sign of Forecast Revision	Number
Positive/no change	21
Negative	90
$\chi^2=42.89$ Prob. < .0001	

Panel D: Tests of the relation between reverse split announcement returns and analysts forecast revisions.

Model:

$$\mathrm{CAR}_{jt} = \lambda_0 + \lambda_1 \mathrm{REVISE\%} + \varepsilon_{jt}$$

Parameter	Coefficient	t-Statistic	Probability
λ	−0.032	−2.46	0.017
λ_1	0.038	1.39	0.085
R^2	0.032		
Ad. R^2	0.015		
N	111		

$$\mathrm{REVISE\%}_{j,t} = \left[\frac{\text{Post-Split Forecast} - \text{Pre-Split Forecast}}{\text{Pre-Split Forecast}}\right]$$

Notes: CAR_{it} is the cumulative market-adjusted return for firm i during the four-day measurement period ending two days after the reverse split announcement date (–1 to +2). Market-adjusted returns are the raw returns for firm j over the event period minus the returns on a CRSP equally-weighted market index.

N = the number of firms used to estimate the parameters of the model.

Table 7. Tests of the Relation between Annual Market-Adjusted Returns and Earnings Changes in the Years Surrounding Reverse Splits.

Model:

$$\mathrm{CAR}_{jt} = \lambda_0 + \lambda_1 \Delta\mathrm{EPS}_{jt} + \lambda_2 D_1 \Delta\mathrm{EPS}_{jt} + \lambda D_2 \Delta\mathrm{EPS}_{jt} + \varepsilon_{jt}$$

Panel A: Earnings changes and market adjusted returns

Parameter	Coefficient	t-Statistic	Probability
λ_0	0.128	6.24	<0.000
λ_1	0.056	2.59	0.010
λ_2	−1.141	−4.21	<0.000
λ_3	−0.062	−2.31	0.021
R^2	0.013		
Ad. R^2	0.012		
N	1440		

Notes: Specification tests are conducted to assess whether the residuals are homoskedastic and cross-sectionally independent. These tests (including the White's test) fail to reject the null hypothesis that the residuals are homoskedastic and cross-sectionally independent at the 0.05 level.

$\mathrm{MAR}_{j,t}$ is the market-adjusted return for firm j for year t.

$\Delta E_{j,t}$ is the change in earnings for firm j in year t standardized by the firm's stock price at the beginning of the year.

D_1 is an indicator variable taking on a value of one in years 1 and 2 and zero otherwise.

D_2 is an indicator variable taking on a value of one in year –2 through year 2 when EPS_{jt} is negative, and zero otherwise.

N is the number of firms used to estimate the parameters of the model.

than in the pre-split years. This result is consistent with investors reducing their reliance on subsequent earnings releases after a reverse split. Because losses may decrease the association between price and earnings following a reverse split due to the documented reduced association in the price–earnings relation for loss firms (Hayn, 1995), I include an additional indicator variable D_2, which takes on a value of one in year –2 through year 2 when EPS_{jt} is negative, and zero otherwise. λ_3 the coefficient estimate for D_2 is negative and significant consistent with losses reducing the price–earnings relation. However, these findings suggest that the observed decrease in the price reaction to earnings after a reverse split is likely consistent with investors reducing their reliance on earnings information after a reverse split, consistent with H3.

5. SUMMARY AND CONCLUSIONS

The negative market reaction to reverse split announcements is puzzling to researchers since reverse splits are associated with increased liquidity, marketability, and marginability of firms' stocks. Prior studies conjecture that reverse splits convey unfavorable information about future earnings prospects (West & Brouilette, 1970; Radcliffe & Gillespie, 1979; Spudeck & Moyer, 1985) but do not provide definitive evidence of this link.

I examine the relation between reverse split announcement returns and earnings changes in the five years surrounding a reverse split. I find a significant negative market reaction in the four-day window around the WSJ announcement of a reverse split. The market reaction is positively related to the earnings performance before and after the reverse split suggesting that reverse splits provide information on the persistence of past earnings and the probability of future earnings declines. This conclusion is supported by the findings that analysts significantly revise their long-term earnings forecasts downwards and that these forecast revisions are positively related to abnormal announcement returns. I also find smaller ERCs after a reverse split consistent with a weaker earnings–returns relation after a reverse split.

The results of this study suggest that reverse splits provide information about a firm's future earnings. This information causes investors to decrease the weight they place on earnings releases after the reverse split resulting in lower earnings response coefficients or ERCs after the reverse split. Thus, reverse split information appears to change investors' expectations about future earnings.

NOTES

1. Even though increasing the par value is typical, some firms elect to leave par value unchanged.
2. Several managers in proxy statements list reasons for reverse splitting. Marginability refers to the ability of investors to purchase the stock on margin.
3. Han (1995) documents an average abnormal return of −4.69% over two days (day 0 through +1) for his sample of reverse split announcements.
4. Costs associated with reverse splits include taxes, fees, management time, and mailing costs (Sosnick, 1961).
5. This is based on the theoretical arguments in the Holthausen and Verrecchia (1988) model.
6. Managerial actions such as stock split, equity-for-debt swaps, share repurchases, and dividend changes alter investor expectations about future returns to the underlying security (King, Pownall, & Waymire, 1990).

7. The Center for Research on Security Prices at the University of Chicago data reveal 1603 recorded reverse splits over the 1962–2000 period.

8. In a sample of 153 firms with proxy data, most cite increasing marketability and marginability, as well as reducing transaction costs as their primary motivation for the reverse split.

9. Several firms report in their proxy statements a general dissatisfaction with the price at which their shares are trading in the market. For example, Mallon Resources states that the "reverse split is being done to increase acceptance of the stock by the financial community and the investing public, and increase shareholder value."

10. For example, brokers at Merrill Lynch & Co. cannot recommend shares not rated by their analysts, who generally do not rate "penny stocks" … stocks trading below $5. The *Wall Street Journal* – June 21, 2001.

11. While a reverse split Ceteris paribus, increases the per share stock price which may satisfy the minimum share price for the exchange, it also reduces the number of publicly traded shares which may violate the traded shares criterion for listing – AMEX and NASDAQ listing requirements (1998). There was no indication from proxy data that a reverse split resulted in this situation. In fact, most firms indicated that the number of shares listed was much larger than the listings for similar firms, and the reverse split would result in a closer alignment in this regard.

12. The *New York Times* – November 18, 1991, Section 1, page 32, column 3.

13. The *Wall Street Journal*, "Deals & Deal Makers: Battered Companies Do Reverse Split – While Tactics Boosts Share Prices, It seldom Reverses Fortune" – June 21, 2001.

14. Several studies including Lakonishok and Lev (1987), Asquith et al. (1989), Doran (1994) and others provide empirical evidence that stock splits signal that past earnings will persist.

15. Annual earnings announcement dates are obtained from the quarterly Compustat tapes. These dates are then used to center the earnings changes relative to the reverse split announcement, so that earnings in year $+1$ is the first annual earnings after the reverse split announcement. Earnings changes are scaled by the fiscal-year-end closing price in the event year.

16. Compustat data item #58 (income before extraordinary items/Compustat data item #25 (common stock outstanding). The results are qualitatively unchanged when Compuatat data item #18 (EPS before extraordinary items) is used instead.

17. Market-adjusted returns are computed using the equal-weighted market returns and NASDAQ composite returns obtained from CRSP. Size-adjusted returns are computed as the raw returns for firm j over the event period minus the returns for firms in the same size-decile classification as firm j in the year of the reverse split, obtained from CRSP.

18. This finding is consistent with prior research such as Ball and Brown (1968) and Asquith et al. (1989).

ACKNOWLEDGMENT

I am grateful for the helpful comments of Linda Bamber, Jim Boatsman, Paul Copley, Margaret Conchar, Mark Dawkins, Jennifer Gaver, Kenneth

Gaver, Graeme Rankine, Philip Reckers, Ena Rosegreen, and workshop participants at the University of Georgia, Georgia Tech, Emory, Northeastern, Arizona State, and University of Houston.

REFERENCES

Asquith, P., Healy, P., & Palepu, K. (1989). Earnings and stock splits. *The Accounting Review*, *64*, 387–403.

Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, *6*, 159–178.

Bhardwaj, R., & Brooks, L. D. (1992). The January anomaly: Effects of low share price, transaction costs, and bid-ask bias. *Journal of Finance*, *47*, 552–576.

Bhushan, R. (1994). An informational efficiency perspective on the post-earnings announcement drift. *Journal of Accounting and Economics*, *18*, 45–65.

Doran, D. T. (1994). Stock splits: Tests of the earnings signaling and attention directing hypotheses using analyst forecasts and revisions. *Journal of Accounting, Auditing, and Finance*, *9*, 411–422.

Doran, D. T., & Nachtmann, R. (1988). The association of stock distribution announcements and earnings performance. *Journal of Accounting, Auditing, and Finance*, *3*, 113–132.

Dravid, A. R. (1987). A note on the behavior of stock returns around ex-dates of stock distributions. *Journal of Finance*, *42*, 163–168.

Grinblatt, M. S., Masulis, R. W., & Titman, S. (1984). The valuation effects of stock splits and stock dividends. *Journal of Financial Economics*, *13*, 461–490.

Han, K. (1995). The effects of reverse splits on the liquidity of the stock. *Journal of Financial and Quantitative Analysis*, *30*, 159–169.

Hayn, C. (1995). The information content of losses. *Journal of Accounting and Economics*, *20*, 125–153.

Holthausen, R. W., & Verrecchia, R. E. (1988). The effect of sequential information releases on the variance of price changes in an intertemporal multi-asset market. *Journal of Accounting Research*, *26*, 82–106.

Ikenberry, D. L., Rankine, G., & Stice, E. K. (1996). What do stock splits really signal? *Journal of Financial and Quantitative Analysis*, *31*, 357–375.

King, R., Pownall, G., & Waymire, G. (1990). Expectations adjustment via timely management forecasts: Review, synthesis, and suggestions for future research. *Journal of Accounting Literature*, *9*, 113–144.

Klein, L. S., & Peterson, D. (1989). Earnings forecast revisions associated with stock split announcements. *The Journal of Financial Research*, *4*, 319–328.

Lakonishok, J., & Lev, B. (1987). Stock splits and stock dividends: Why, who, and when. *Journal of Finance*, *42*, 913–932.

Lamoureux, C., & Poon, P. (1987). The market reaction to stock splits. *Journal of Finance*, *45*, 1347–1370.

Masse, I., Hanrahan, J. R., & Kushner, J. (1997). The effect of Canadian stock splits, stock dividends, and reverse splits on the value of the firm. *Quarterly Journal of Business and Economics*, *36*, 51–63.

McNichols, M., & Dravid, A. (1990). Stock dividends, stock splits, and signaling. *Journal of Finance*, *45*, 857–880.

Peterson, D. R., & Peterson, P. P. (1992). A further understanding of stock distributions: The case of reverse stock splits. *Journal of Financial Research*, *15*, 189–205.

Pilotte, E. (1997). The wealth and earnings implications of stock splits by nondividend-paying firms. *Quarterly Journal of Business and Economics*, *36*, 81–95.

Pilotte, E., & Manuel, T. (1996). The market's response to recurring events in stock splits. *Journal of Financial Economics*, *41*, 111–127.

Radcliffe, R. C., & Gillespie, W. B. (1979). The price impact of reverse splits. *Financial Analysts Journal*, *35*, 63–67.

Rankine, G., & Stice, E. K. (1997). The market reaction to the choice of accounting method for stock splits and large stock dividends. *Journal of Financial and Quantitative Analysis*, *32*, 161–182.

Sosnick, S. (1961). Stock dividends are lemons, not melons. *California Management Review*, *3*, 61–82.

Spudeck, R. E., & Moyer, R. C. (1985). Reverse splits and shareholder wealth: The impact of commissions. *Financial Management*, *14*, 52–56.

West, R. R., & Brouilette, A. B. (1970). Reverse stock splits … harbinger of bad times or valid management technique. *Financial Executive*, *38*, 12–17.

Woolridge, R., & Chambers, D. (1983). Reverse splits and shareholder wealth. *Financial Management*, *12*, 5–15.

This page intentionally left blank

NEW EVIDENCE ON AUDITOR INDEPENDENCE POLICY

Philip M. J. Reckers and Dahlia Robinson

ABSTRACT

Auditor independence has been extensively examined from the perspective of independence in appearance, driven by the perception that the provision of non-audit services (NAS) leads to the diminution of auditor independence as well as appearance. This study reports on experimental laboratory research using experienced auditors and finds, similar to archival studies, that independence in fact is not compromised by NAS. On the contrary, our results suggest that greater NAS associates with more conservative auditor going-concern judgments among more experienced auditors. One explanation is that greater NAS associates with documented firm-performance deficiencies and experienced auditors recognize that association and respond accordingly.

1. INTRODUCTION

This study utilizes an experimental research design with auditor subjects to reexamine the question of whether auditor independence (measured by the auditor's propensity to issue a going-concern modification) and is compromised by the provision of non-audit services (NAS). The issue of auditor independence gained prominence at the Securities and Exchange

Advances in Accounting, Volume 23, 207–229

ISSN: 0882-6110/doi:10.1016/S0882-6110(07)23008-1

Commission (SEC) during the late 1990s, when both the SEC and users of financial statements became increasingly concerned about the potential negative effects on auditor independence when auditors provide both audit and NAS to their audit clients. These concerns were driven by dramatic changes in the accounting profession and in the types of services that auditors were providing to their audit clients, as well as increases in the absolute and relative size of the fees charged for NAS (SEC, 2000). Then SEC Chairman, Arthur Levitt, verbalized these concerns stating "the Big 5 public accounting firms position themselves globally as 'multidisciplinary professional service organizations' rather than accounting firms, so that auditor "independence if not in fact, then certainly in appearance, becomes a more elusive proposition" (Levitt, 2000). Similarly, Paul Volcker, former Chairman of the Federal Reserve Board, responding to a question about investors' perceptions of conflict of interest when auditors provide NAS, said, "The perception is there because there is a real conflict of interest …" (Volcker, 2000).

In response to the widespread concern about the potential lack of auditor objectivity, the SEC revised its auditor independence rule in November 2000. The new rule limits the types of NAS audit firms can perform for their audit clients, and requires detailed disclosures by SEC registrants of both audit and non-audit fees paid to auditors (in their proxy statements filed on or after February 5, 2001 (SEC, 2000)).[1] In a similar move, Congress, through the Sarbanes-Oxley Act of 2002 (hereafter, SOX) (a direct response to the large corporate failures at Enron, WorldCom, and others) prohibits audit firms from providing any financial information system design and implementation (FISDI) services, internal audit, and "certain other services" to audit clients. Thus, regulators and legislators by their actions, convey the belief that auditor independence in fact is compromised by the provision of certain NAS, resulting in lower quality audits and reviews, and lower quality financial reporting. In addition, where not strictly prohibited, NAS that are permitted to continue (such as tax advisory services) are operationally discouraged by regulations requiring special ex-ante approval by corporate boards of directors. These reform-minded initiatives are not without significant continuing costs and impact on the auditing profession and their corporate clients. Non-audit tax service fees, which represent the largest category of NAS fees, have declined by 8% from 2002 (the year the SOX took effect) to 2003 (Markelevich, Barragato, & Hoitash, 2005). Regulation effectively discouraging the provision of tax services to audit clients is tantamount to requiring a complete reconfiguration of the services and client portfolios of public accounting firms.

Unlike the prior research examining the relation between auditor independence and the provision of NAS, this research uses auditors as the experimental subjects. The main objective with this approach is to gain further insights on the competing environmental pressures that auditors face that potentially impact their independence (DeZoort & Lord, 1997). On the one hand, auditors face pressures to earn "economic rents" through retention of important clients (Levitt, 2000; McKeown, Mutchler, & Hopwood, 1991; Hackenbrack & Nelson, 1996), while on the other hand, there are pressures to retain audit firm reputation and avoid costly litigation (McCracken, 2003; Moreno & Bhattacharjee, 2003). Arguably, the balance of those pressures may have changed as a result of the accounting debacles at the turn of the century and the resultant litigation, SOX and SEC legislation, Public Company Accounting Oversight Board (PCAOB) and stock exchange reforms, as well as the intense press coverage of the auditor independence issue. Our manipulation of NAS fees seeks to shed some light on the current balance of these pressures.

Our experimental subjects are audit seniors from a Big 4 accounting firm undergoing required CPE training. Participants were provided with experimental materials describing a scenario of a typical audit (see Appendix A). The materials included background information on the corporate audit client, a toy manufacturer facing standard but major industry risks. The materials also addressed the relationship (audit fees, non-audit fees, and tenure) of the audit firm and audit client. The client had received unqualified audit opinions for the prior 5 years and had been a client of the audit firm for 2 years. The background information also included the manipulated variables – existence of significant non-audit fees (or not) and internally assessed risk of the audit client (high/low). Additionally, the scenario included summarized income statement and balance sheet information for the current year.

We find that the probability that the auditor will render a going-concern qualification is greater for firms exhibiting higher risks (i.e., with below average performance ratios, above average leverage, and with higher risk scores based on the firm's internal risk-scaling technology software) than for firms exhibiting lower risks. More importantly, we also find that higher NAS fees are associated with higher probabilities of issuing a going-concern opinion ($F = 11.695$, $p = .001$). This experimental effect is contrary to much of the prior research that has associated higher NAS fees with more client deferential audit judgments.

One potential explanation for this result is that auditors have become highly sensitized in recent years to the public's concern over high levels of NAS being provided by auditors.[2] However, this explanation would suggest that our manipulation should not yield any significant main or interaction effects. That is, if auditors were previously significantly less inclined to qualify the opinions registered for important clients (directionally opposite results), we might now expect these same auditors (being more conservative) to discount client financial importance in the performance of the audit, and to treat less and more financially important clients equally. It is not apparent why enhanced sensitivity to this issue of auditor independence and the level of NAS fees would lead to an increased probability of opinion qualification for high NAS fees audit clients. Further, prior research in general, finds that more naïve and less experienced individuals tend to react more to issues that have gained significant attention in the popular press and arguably are typical (McDaniel, Martin, & Maines, 2002; Tversky & Kahneman, 1973; Fiske & Larter, 1983; Bouwman, 1984). Our research findings are contrary to this expectation. Specifically, we document a significant association between the probability of issuing a going-concern qualification and NAS fees only among the more experienced audit seniors.

Another potential explanation for our results relate to associations which audit seniors (but possibly not audit associates) make about firms that engage significant NAS services. Iyer, Ravrindran, and Reckers (2006) recently report that firms that contract for greater NAS are more likely to display greater strategic deficiencies (e.g., inferior margins, turnovers, etc.). Similarly, Brandon, Crabtree, and Maher (2004) report a negative association between the extent of NAS and bond ratings: that is, the investment community associates greater NAS with corporate weaknesses. If audit seniors have gained a similar understanding via experience (i.e., more developed schemas) that clients primarily engage NAS because they need help to remedy non-trivial deficiencies, they might well associate greater engagement of NAS with client firm weakness. Research tells us that one characteristic that varies with the level of experience is how professionals handle information that varies in typicality. Bouwman (1984), for example, shows how experienced financial analysts focus on potential contradictions, whereas inexperienced analysts focus on the typical. Choo and Trotman (1991) report similar results for audit seniors versus audit associates in going-concern decisions. Experienced auditors may assimilate in a more complex fashion (than inexperienced auditors) information regarding client risk and NAS. This understanding may indeed have been further enhanced as a result of SOX Section 404 work which has revealed client deficiencies in a number of areas.[3]

Our result is also consistent with suggestions that NAS potentially improve audit effectiveness through information spillover (the Panel on Audit

Effectiveness, 2000; Simunic, 1984; Kinney, Palmrose, & Scholz, 2004). For example, knowledge of a client's tax accounting and computer system could spill over to the audit and improve audit effectiveness and quality, which in turn results in an increased probability of detecting material weaknesses that would affect a company's ability to continue as a going concern. Finally, the result is also consistent with Dopuch, King, and Schwartz (2003)'s suggestion that for high-profile clients, economic incentives to maintain reputational capital, may in turn motivate audit thoroughness and independence in audit reporting decisions.

Our experimental results also corroborate the findings from several empirical papers that NAS does not appear to impair auditor independence (Antle, Gordon, Narayanamoorthy, & Zhou, 2002; Defond, Raghunandan, & Subramanyam, 2002; Ashbaugh, Lafond, & Mayhew, 2003; Chung & Kallipur, 2003; Francis & Ke, 2003; Reynolds, Deis, & Francis, 2004; Larcker & Richardson, 2004; Kinney et al., 2004; Robinson, 2006). Taken together, the results from these studies have some potentially important policy implications. First, there is limited empirical justification for using the level of NAS as a measure of auditors' independence. Second, disclosures of NAS may, by suggestion, inappropriately reduce investors' perceptions of auditor independence (e.g. Lowe & Pany, 1995, 1996; Dopuch et al., 2003; Church & Zhang, 2003; Davis & Hollie, 2004), which may reduce the accuracy of investors' beliefs in firms' financial information resulting in lower market efficiency (Dopuch et al., 2003; Davis & Hollie, 2004). Finally, the disclosures of NAS and the impact on investors' perceptions of auditor independence may serve as a trigger for shareholder lawsuits. Accordingly, Church and Zhang (2003) find that the frequency of auditor lawsuits is significantly higher when the auditor provides NAS, regardless of the fees generated from such services.

The remainder of the paper is organized as follows. Section 2 discusses the relevant prior research and develops our research hypotheses. Section 3.1 describes the experimental methods and design. Section 4 reports the research results. Section 5 concludes with a summary.

2. PRIOR RESEARCH AND TESTABLE HYPOTHESES

The regulatory reforms have been fueled, in part, by regulators' and investors' beliefs that auditors' provision of significant amounts of certain types of NAS to their audit clients implies the absence of independence in appearance (Pitt & Birenbaum, 1997; Dopuch et al., 2003). Since independence in fact is unobservable, many supporters of this regulation also infer an absence of independence in fact when they conclude that independence in appearance is lacking. Thus, the mandatory disclosure of audit and non-audit fees paid to auditors may serve as a proxy for the unobservable construct, in fact independence (Dopuch et al., 2003).

The detailed auditor fee disclosure, mandated by the SEC in 2000, has generated a number of empirical archival research studies investigating whether auditor independence in fact is adversely affected by the provision of NAS (e.g. Chung & Kallapur, 2003; Francis & Ke, 2003; Frankel, Johnson, & Nelson, 2002; Defond et al., 2002; Reynolds et al., 2004; Larcker & Richardson, 2004; Kinney et al., 2004). The argument then is that if economic bonding between the auditor and the client increases because of increased NAS, then the auditor will be favorably disposed towards managers' financial reporting choices.

These studies operationalize auditor independence in fact in a variety of ways including, firms' propensity to manage accruals, meet earnings targets, and issue restatements, or the auditor's propensity to issue going-concern opinions. Overall, these studies do not find consistent support for a diminution in auditor independence with the provision of NAS. Frankel et al. (2002), document a positive relation between NAS fees and the level of discretionary accruals consistent with increased economic bonding between the auditors and clients from the provision of NAS. However, this finding by Frankel et al. (2002) has not been corroborated by other studies. For example, Antle et al. (2002), Chung and Kallipur (2003), Francis and Ke (2003), Reynolds et al. (2004), and Larcker and Richardson (2004) find no consistent evidence that the ratio of NAS fees to total fees is positively correlated to accruals and attribute the Frankel et al. (2002)'s findings to measurement error in the dependent variable (fees) that is correlated with the independent variable. Larcker and Richardson (2004) use latent class mixture models approach to investigate the question and consistently document a negative relation between the level of fees paid (both audit and non-audit) to auditors and the level of discretionary accruals.

Similarly, Kinney et al. (2004) examines firms' propensity to restate previously filed financial statements as a proxy for lower financial reporting quality. They document a consistent negative association between tax service fees and restatements suggesting that there are "net benefits from acquiring tax services from a registrant's audit firm" (Kinney, 2004, p. 560). One disadvantage of focusing on accounting accruals, meeting earnings forecast, or earnings restatements is that auditors have only an indirect effect on these client outcomes.

A number of other studies conduct more direct tests of the effects of NAS on auditor independence by focusing on auditors' going-concern opinions. Going-concern opinions are well suited for this situation for a number of reasons. First, SAS No. 59 (AICPA, 1988) requires auditors in each engagement to assess and report on the continued viability of their clients and to issue a going-concern qualification in the audit opinion if substantial doubt exists as to the firm's ability to continue as a going concern. Both the initial going-concern assessment and the evaluation of management's plans to remedy the situation involve highly subjective judgments (Mutchler, Hopwood, & McKeown, 1997; Behn, Kaplan, & Krumweide, 2001; Geiger & Rama, 2003). Second, going-concern opinions are also costly for both the auditor and the client (Kida, 1980; Blay & Geiger, 2001; Weil, 2001). Finally, prior research documents that large audit clients are less likely to receive going-concern opinions (McKeown et al., 1991; Mutchler et al., 1997) and are likely to have larger audit fees (Simunic, 1980, 1984; Francis, 1984) and non-audit fees (Abbott, Parker, Peters, & K. Raghunandan, 2003). Thus, it is reasonable to assume that auditors' going-concern opinions may be correlated with the magnitude of both audit and non-audit fees they receive from their clients.

Defond et al. (2002), Geiger and Rama (2003) examine the correlation between NAS and auditors' propensity to issue going-concern opinions for distressed firms. Defond et al. (2002) find no significant correlation between NAS fees and the probability of receiving a going-concern opinion for distressed firms while Geiger and Rama (2003) find a significant positive correlation only for audit fees. Robinson (2006) decomposes NAS fees into a tax and a non-tax component. She documents a positive and significant correlation between the tax component of NAS fees and receiving a going-concern opinion in bankrupt firms. However, none of these studies utilize an experimental approach with auditor subjects. Beeler and Hunton (2002) did use an experimental approach to examine the issue. Notably, they report a negative association between NAS and independence. However, their study required their auditor subjects to follow a directed information deliberation process not consistent with current audit practices; and they recognized that their experimental materials lacked "a host of (relevant) endogenous and exogenous factors" normally found in practice.[4] Beeler and Hunton, for example, did not examine the influence of high versus low risk of client bankruptcy and audit failure; whereas we do. Their work also pre-dates critical SOX legislation. We extend these studies by examining the auditor reporting decision and its association with NAS using an experimental research design using auditor subjects and incorporating both risk and reward task-tension dimensions.

We develop and test two research hypotheses. First, given the intense press coverage of auditor independence issues and the regulatory initiatives restricting the kinds of NAS auditors may provide to their audit clients, we argue that auditors have become more conservative in their decisions (consistent with research by Lindberg and Beck, 2004). We expect that more conservative auditors would be more likely to discount client financial importance in the performance of the audit, treating all audit clients more or less the same. Second, we suggest that the potential relation between the going-concern opinion and the level of NAS fees may be affected by auditor experience. Prior research suggests that experienced professionals are more likely to focus on potential contradictions in information, whereas inexperienced professionals tend to focus on the typical aspects of the information (Bouwman, 1984; Choo & Trotman, 1991). Inexperienced auditors may therefore be more influenced by the negative publicity surrounding auditor independence and the provision of NAS. We therefore hypothesize that:

H1. The auditor's probability of issuing a going-concern decision is uncorrelated with the level of non-audit service fees paid by the audit client.

H2. Experienced and inexperienced auditors differ in their response to the provision of significant NAS to the audit client.

3. EXPERIMENTAL METHODS AND DESIGN

3.1. Participants

The experiment was conducted with 92 audit seniors attending a CPE training program conducted by a Big 4 accounting firm. Table 1 provides a general demographic data profile. Participants under the various experimental conditions did not differ on any of the profiling items. Participants who failed manipulation checks were removed from analyses. The primary findings are based on the responses of 92 audit seniors.

Audit seniors have been frequently used in similar audit research as experienced professionals. Hackenbrack (1992)'s research on auditor's fraud risk assessment and non-diagnostic cues employed auditor seniors. Zimbelman (1996, 1997) employed 108 audit seniors in his examination of how *SAS No. 82* affects auditors' attention to fraud. On the subject of client

Table 1. Profile of Subjects ($n = 92$).

	Mean	Standard Deviation
Age	27	3.4
Gender (% female)	49%	
In-charge engagements	7.5	7.0
Years as auditor	3.2	1.5
Owner of stock	66%	

integrity and competence of auditors' assessment of material misstatements, Bernardi (1994) employed 342 seniors from Big 6 accounting firms. Chang and Hwang (2003) employed 43 audit seniors to examine the individual and interactive effects of client retention and clients' business risk on acquiescence to aggressive reporting practices. More recently, Low (2004) had 98 audit seniors as subjects to investigate the effect of industry specialization on auditors' risk assessment and audit-planning decisions. Choo and Trotman (1991) used 22 experienced auditors (audit seniors) and 22 non-experienced participants (audit associates) in a going-concern task; they found audit seniors responded differently than audit associates. Finally Shelton, Whittington, and Landsittel (2001) state that ''representatives from several firms indicated that in practice (fraud risk assessments) may be performed by the senior with review and concurrence by the manager and partner''. Thus, audit seniors have been widely used in prior accounting research and we believe them to be appropriate for this research. Nevertheless, we acknowledge the limitations posed by our choice of participants on our findings in the conclusion section.

3.2. Research Task and Dependent Variable

Participants were provided with experimental materials describing a scenario of a typical audit (see Appendix A). The materials included background information on the corporate audit client, a toy manufacturer facing standard but major industry risks including increasing foreign competition, trying to predict consumer trends, compliance with changing government safety regulations and increasing product liability litigation and firm-specific risks related to labor negotiations and stability of major customer relations. Materials also disclosed the relationship (audit fees, non-audit fees, and tenure) of the audit firm and audit client. The client had received unqualified audit opinions for the prior 5 years and had been a client of the audit firm for 2 years. The background information also included the manipulated variables – existence of significant non-audit fees (or not) and internally assessed risk of the audit client (high/low). Additionally, the scenario included summarized income statement and balance sheet information for the current year.

Case materials were modified versions of those used by Lowe (1994) and depicted a corporation facing various environmental risks. In every audit, risks associated with business continuity must be evaluated. Accordingly, participants in this instance were instructed to assess whether a going-concern audit opinion qualification was appropriate for Playworks Corporation, the fictitious client. Specifically, participants were asked:

Given the admittedly incomplete information provided above, do you believe that Playworks warrants a going concern audit opinion?

Absolutely No	Probably No	Probably Yes	Absolutely Yes
0……..1……..2……..3……..4……..5……..6……..7……..8……..9……..10			

Participants knew there were additional research tasks they would address during the dedicated research session. They did not know that this task was the only task related to the Playworks' audit.

3.3. Manipulated Variables: Client Risk

Client risk was varied at two levels: low and high. In the low-risk conditions, the client was described as having performance ratios above industry averages and leverage ratios below average, a risk score of 2 on a 6 point ascending scale using an in-house risk scaling technology software, and an explicit interpretation that the client exhibited "relatively low risk". Specific language was as follows:

Currently, performance ratios are above industry averages and leverage is below average. New management sought out new auditors when first assuming control 2 years ago. While KRisk client scaling technology was not available at KPMG when Playworks was accepted as a client 2 years ago, Playworks' KRisk score would have been 1 and Playworks' current KRisk score remains at 2, indicating relatively low risk.[5]

In the high-risk condition, the client was described as having performance ratios below industry averages and leverage ratios above average, and a risk score that had declined sharply from 2 to 4 on a 6 point ascending scale over the past 2 years using an in-house risk scaling technology software. Specific language was as follows:

Currently, performance ratios are below industry averages and leverage is above average, although the firm continues to report a profit. While KRisk client scaling technology was not available at KPMG when Playworks was accepted as a client 2 years ago, Playworks would have received a KRisk score of 2. Playwork's current KRisk score has declined to 4- (out of 6), indicating a shift (increase in) risk.

3.4. Non-Audit Services

NAS fees were varied at two levels: none and significant. A significant body of pre-SOX research has examined the association of significant NAS, and implied (or explicitly noted) client economic importance (to the audit firm, office, and/or individual partners) with auditor judgments and decisions. This research has been conducted in an effort to gain a better understanding of environmental pressures related to auditor independence. Competing pressures have been consistently noted in the literature (DeZoort & Lord, 1997). Auditors face pressures to earn ''economic rents'' through retention of important clients (Levitt, 2000; McKeown et al., 1991; Hackenbrack & Nelson, 1996), but also face pressures to retain audit firm reputation and avoid costly litigation (McCracken, 2003; Moreno & Bhattacharjee, 2003). The balance of those pressures may have arguably shifted under the intense scrutiny the audit profession has been subjected to because of the accounting scandals and increased litigation in the early 2000s and the increased legislative reform by the SEC and Congress. We manipulate NAS fees in an attempt to learn more about the current balance of those pressures. Specific language in the no NAS fess/low client importance condition was as follows:

The Playworks' engagement is only for audit services. The engagement is not a material source of office revenues.

Specific language in the significant NAS fess/high client condition was as follows:

The engagement is one of the largest in the office when board-of-directors-approved fees for tax services are included. Tax services are about twice audit fees.

3.5. Statistical Model

The basic experimental design is a 2×2 fully crossed ANOVA. To this basic design a third variable is added: ownership of stock by the audit senior (see Fig. 1). Participants were asked whether they currently owned stock with the expectation that significant attitudinal differences might manifest between groups reflecting different perceptions of the role of the auditor. Other researchers, practitioners, and policy makers frequently refer to "the expectations' gap": a difference in perspective between what stockholders expect of auditors and what auditors believe they are professionally obligated to do. While inclusion of this variable was somewhat speculative in nature, a near-significant interaction between stock ownership and NAS fees is reported in the findings section below, attesting to the importance of this factor and helping to explain results regarding NAS fees that are contrary to much prior pre-SOX experimental behavioral research.

Also included in our model were two covariates: extent of in-charge experience and gender. These variables, as well as participant age and years or audit experience, were examined in an effort to avoid "the critical missing variable problem" and because of the ease of gathering this data. Time restrictions imposed by the participating Big 4 firm did not allow us to collect responses to extensive batteries of attitude questionnaires or psychometric scales (e.g., MacDonald's Tolerance of Ambiguity, Crowne & Marlowe's Social Desirability Scale, etc.). The extent of in-charge experience and gender proved statistically significant; age and years as an auditor did not. The gathering of this information and its inclusion was not purely speculative. A significant body of research has addressed the importance of considering experience and expertise in the conduct of behavioral audit experiments (e.g., Abdolmohammadi & Wright, 1987; Waller & Felix, 1984; Haynes, Jenkins, & Nutt, 1998; Vera-Munoz, Kinney, & Bonner, 2001; Knapp & Knapp, 2001). Gender has also been found significant frequently in a variety of decision contexts.

	Low Client Firm Risk		**High Client Firm Risk**	
	Do Not Own Stock	**Own Stock**	**Do Not Own Stock**	**Own Stock**
No NAS/Low Client Importance				
Significant NAS/ High Client Importance				

Fig. 1. Experimental Design.

4. RESULTS

Table 2, Panels A, B, and C present our primary findings. Before examining the specific outcomes of Table 2, we should note that preceding these analyses, we examined a larger participant pool (160 auditors) that included 58 audit associates as well as 112 audit seniors. Of this group, 50 audit associates and 92 audit seniors passed our manipulation checks. In that parallel ANCOVA the following basic findings emerged:

- Firm Risk was significant ($F = 5.869$, $p = .017$)
- NAS Fees/Client Importance was significant ($F = 4.422$, $p = .038$)
- The Rank by NAS Fees/Client Importance interaction was significant ($F = 4.184$, $p = .043$)
- The Rank by NAS Fees by Firm Risk interaction was significant ($F = 5.064$, $p = .026$)

On examination of the interactions, we found no significant main effects or interactions among audit associates. The results reported in Table 2 were based on the audit seniors, implying an auditor rank effect. One interpretation of this finding is that the influence of our manipulations might be viewed as somewhat subtle and not the result of obvious demand effects for a politically correct response.

The results reported in Table 2 show that the Firm Risk manipulation is statistically significant ($F = 3.899$, $p = .052$) among audit seniors. Table 2, Panel B displays treatment means. For the low-risk condition, the mean response is 4.64; for the high-risk condition, the mean response is 5.67. The outcomes are thus in the predicted direction. The perceived propriety of a going-concern qualification is greater for firms with below average performance ratios, above-average leverage, and scoring high using internal firm risk-scaling technology software. Among our participants, our NAS fees manipulation was also statistically significant ($F = 11.695$, $p = .001$). However, the direction of the experimental effects is contrary to much prior research that has associated higher NAS fees with more client deferential audit judgments. In our study, we find that higher NAS fees associate with higher probabilities of rendering a going-concern qualification. Under the No NAS fees condition, the treatment mean is 4.25. Under the Significant NAS fees condition, the treatment mean is 6.01.

We also found a marginally significant NAS Fees by Stock Ownership interaction ($F = 3.420$, $p = .068$). Given that our findings relating to NAS fees are contrary to much prior behavioral research, we look to this interaction for any additional light it may shed. Table 2, Panel C provides

Table 2. Experimental Findings.

Panel A: Principal findings tests of between-subjects effects dependent variable: Going concern

Source	Type III Sum of Squares	*df*	Mean Square	*F*	Significance
Corrected model	159.861(a)	9	17.762	3.737	.001
Intercept	128.714	1	128.714	27.078	.000
In-charge experience	28.878	1	28.878	6.075	.016
Gender	25.703	1	25.703	5.407	.023
Firm risk	18.535	1	18.535	3.899	.052
NAS fees/client importance	55.595	1	55.595	11.695	.001
Stock ownership	6.282	1	6.282	1.322	.254
Firm risk × NAS fees	.211	1	.211	.044	.833
Firm × stock ownership	.274	1	.274	.058	.811
NAS fees × stock ownership	16.256	1	16.256	3.420	.068
Firm risk × NAS fees × stock	2.530	1	2.530	.532	.468
Error	389.791	82	4.754		
Total	3,054.000	92			
Corrected total	549.652	91			
$R^2 = .291$					
Adj. $R^2 = .213$					

Panel B: Main effects

1. Firm Risk Dependent Variable: Going Concern

Firm risk	Mean	Standard error	95% confidence interval	
			Lower bound	Upper bound
High firm risk	5.627(a)	.348	4.934	6.320
Low firm risk	4.635(a)	.350	3.939	5.331

2. NAS Fees/Client Importance Dependent Variable: Going Concern

NAS fees/client importance	Mean	Standard error	95% confidence interval	
			Lower bound	Upper bound
No NAS fees	4.254(a)	.348	3.563	4.946
Significant audit fees	6.008(a)	.358	5.296	6.719

Panel C: Interaction effects

NAS Fees (Client Importance) × Stock Ownership Dependent Variable: Going Concern

Client importance	Stock ownership	Mean	Standard error	95% confidence interval	
				Lower bound	Upper bound
No NAS fees	No ownership	3.519(a)	.535	2.454	4.584
	Owns stock	4.990(a)	.433	4.129	5.851
Significant NAS fees	No ownership	6.166(a)	.591	4.989	7.342
	Owns stock	5.850(a)	.400	5.053	6.646

related cell means. For both participant groups (those who own stock and those who do not), we find significant and directionally similar results. However, the affect of the NAS fees manipulation is greater among audit seniors who do not own stock and is driven by the relatively low mean response registered under conditions of No NAS fees (3.519 vs. 4.990 for auditor seniors who own stock). Mean responses across both groups are similar under conditions of Significant NAS fees (5.85 vs. 6.17).

There are potentially two interpretations for these results. One possible explanation is that auditors have become more conservative about NAS because of the intense publicity over recent years. As we explain earlier, we would then expect that our manipulation would render no significant main or interaction effects. That is, whereas previously one might expect auditors to be more inclined to not qualify the opinions registered for important clients (directionally opposite results), we would now expect auditors to discount client financial importance in the performance of the audit and treat less and more financially important clients equally. Thus, it is unclear how a more conservative view on this issue would lead to an increased likelihood of opinion qualification for high NAS fees audit clients. Further, prior research in general finds that more naïve and less experienced individuals tend to react more to issues that have gained significant attention in the popular press. Our findings are to the contrary: our NAS fees manipulation does not elicit a significant response among less experienced audit associates, but does so among more experienced audit seniors.

The other possible explanation in our minds relates to associations which audit seniors (but possibly not audit associates) make to firms that engage significant NAS services. Ravindran et al. (2006) recently report that firms that contract for greater NAS are those firms that display greater strategic deficiencies (e.g., inferior margins, turnovers, etc.). Brandon et al. (2004) report a negative association between extent of NAS and bond ratings. If seniors have gained a similar understanding via experience that clients primarily engage NAS because they need help to remedy deficiencies, they might well associate greater engagement of NAS with client firm weakness. More developed schemas indeed may be been further advanced as a result of SOX Section 404 work which has revealed client deficiencies in a number of areas.

Our result is also consistent with suggestions that NAS potentially improve audit effectiveness through information spillover (the Panel on Audit Effectiveness, 2000; Simunic, 1984; Kinney et al., 2004). For example, knowledge of a client's tax accounting and computer system could spill over to the audit and improve audit effectiveness and quality, which in turn results in an increased probability of detecting material weaknesses that would affect a company's ability to continue as a going concern.

There are two possible explanations for why a client firm might engage NAS services. The first is an attempt to buy influence over the auditor. The second is because of internal deficiencies or weaknesses to address important/critical needs. Audit seniors can be expected to have been exposed to both explanations. The greater salience of the second explanation alone is consistent with our results.[6]

5. SUMMARY AND CONCLUSIONS

The legislative and regulatory reform restricting the types of NAS auditors can perform for their audit clients is motivated by the belief the auditor objectivity is impaired by the provision of NAS. Recent archival research on the issue provides little empirical support for this belief. We reexamine the issue of auditor independence and NAS using an experimental laboratory research setting with auditors subjects. By using real auditors we hope to gain additional insights on the competing environmental pressures that auditors face that potentially impact their independence, since the balance in those pressures may have shifted in recent years.

We find no evidence that auditors' reporting decisions are adversely affected by the provision of NAS. Specifically, we find that the probability of the client receiving a going-concern opinion increases with the level of NAS. This finding is driven by the more experienced auditor subjects, suggesting that experience may allow auditors to associate greater NAS with documented firm-performance deficiencies. The result is also consistent with suggestions that NAS potentially improve audit effectiveness through information spillover.

NOTES

1. This was a compromise position for the SEC as it had originally sought to ban all NAS. The expanded fee disclosure was mandated by the SEC in the belief that these disclosures would be subject to increased scrutiny by investors.

2. Lindberg and Beck (2004), using survey instruments, find results that suggest that Certified Public Accountants (CPA) s' perceptions of the effects of non-audit services on auditor independence are more negative after the Enron bankruptcy.

3. Seniors may also inherit their understanding of the negative association of NAS and risk from supervisors and managers. Wilks (2002) reports seniors' judgments in part reflect observed supervisors' preferences and beliefs.

4. Beeler and Hunton examined whether conditions in which the auditor low balled audit fees to secure lucrative non-audit service fees would influence predecisional commitments and audit evidence interpretations consistent with securing economic rents. Unlike our study, Beeler and Hunton did not consider alternative environments exhibiting higher and lower risk of client bankruptcy and audit failure. It is our contention that findings under conditions of low risk do not generalize to conditions of high risk.

5. For client evaluative purposes related to acceptance and audit continuation, KPMG introduced in 2005 the KRisk scale. This 1–6 scale (one least risky, 6 most risky) gives consideration to a number of client dimensions including, but not limited to, financial performance, industry, management integrity, internal control environment and governance structure, and aggressive/conservative past accounting reporting practices.

6. Both covariates in our model proved to be statistically significant. The mean response of female participants was 5.8; of men, 4.6. In-Charge experience exhibited a negative association with our dependent measure: greater in-charge engagement assignments associated with lower going concern probabilities.

REFERENCES

Abbott, L., Parker, S., Peters, G., & Raghunandan, K. (2003). The association between audit committee characteristics and audit fees. *Auditing: A Journal of Practice and Theory*, *22*, 17–32.

Abdolmohammadi, M., & Wright, A. (1987). An examination of the effects of experience and task complexity on audit judgments. *The Accounting Review*, *62*, 1–13.

American Institute of Certified Public Accountants (AICPA). (1988). "*The Auditor's Consideration of an Entity's Ability to Continue as a Going Concern*." Statement on Auditing Standards No. 59. New York, NY: AICPA.

Antle, R., Gordon, E., Narayanamoorthy, G., & Zhou, L. (2002). *The joint determination of audit and non-audit fees, and abnormal accrual*. Working paper. Yale School of Management.

Ashbaugh, H., Lafond, R., & Mayhew, B. (2003). Do non-audit services compromise auditor independence? Further evidence. *The Accounting Review*, *78*, 611–639.

Beeler, J., & Hunton, J. (2002). Contingent economic rents: Insidious threats to audit independence. *Advances in Accounting Behavioral Research*, *5*, 21–50.

Behn, B., Kaplan, S., & Krumweide, K. (2001). Further evidence on the auditors' going-concern report: The influence of management plans. *Auditing: A Journal of Practice and Theory*, *20*, 13–28.

Bernardi, R. (1994). Fraud detection: The effect of client integrity and competence and auditor cognitive style. *Auditing: A Journal of Practice and Theory*, *13*, 68–84.

Blay, A., & Geiger, M. (2001). Market expectations for first-time going-concern recipients. *Journal of Accounting, Auditing and Finance*, *16*, 209–226.

Bouwman, M. (1984). Expert versus novice decision making in accounting: A summary. *Accounting, Organizations and Society*, *9*, 325–328.

Brandon, D., Crabtree, A., & Maher, J. (2004). Nonaudit fees, auditor independence, and bond ratings. *Auditing: A Journal of Practice and Theory*, *23*, 89–103.

Chang, C., & Hwang, N. (2003). The impact of retention incentives and client business risks on auditors' decisions involving aggressive reporting practices. *Auditing: A Journal of Practice and Theory*, *22*, 207–218.

Choo, F., & Trotman, K. (1991). The relationship between knowledge structure and judgments for experienced and inexperienced auditors. *The Accounting Review*, *66*, 464–485.

Chung, H., & Kallapur, S. (2003). Client importance, non-audit fees, and abnormal accruals. *The Accounting Review*, *78*, 931–955.

Church, B., & Zhang, P. (2003). *Independence in appearance, non-auditing services, and auditor fee disclosures*. Working paper. Georgia Institute of Technology and University of Ontario.

Davis, S., & Hollie, D. (2004). *An experimental investigation of non-audit service fees and investors' perceptions of auditor independence: Post-Enron era*. Working paper. Georgia State University and University of Houston.

Defond, M., Raghunandan, K., & Subramanyam, K. (2002). Do non-audit fees impair auditor independence? Evidence from going concern auditor opinions. *Journal of Accounting Research*, *40*, 1247–1274.

DeZoort, F., & Lord, A. (1997). A review and synthesis of pressure effects research in accounting. *Journal of Accounting Literature*, *16*, 28–85.

Dopuch, N., King, R., & Schwartz, R. (2003). Independence in appearance and in fact: An experimental investigation. *Contemporary Accounting Research*, *20*, 79–115.

Fiske, S., & Larter, W. (1983). The novice and the expert: Knowledge-based strategies in political cognition. *Journal of Experimental Social Psychology*, *19*, 381–400.

Francis, J. (1984). The effect of audit firm size on audit prices: A study of the Australian market. *Journal of Accounting and Economics*, *6*, 133–151.

Francis, J., & Ke, B. (2003). *Do fees paid to auditor increase a company's likelihood of meeting analysts' earnings forecasts*? Working paper. University of Missouri.

Frankel, R., Johnson, M., & Nelson, K. (2002). The relation between auditors' fees for non-audit services and earnings management. *The Accounting Review*, *77*, 71–105.

Geiger, M., & Rama, D. (2003). Audit fees, non-audit fees, and auditor reporting on stressed companies. *Auditing: A Journal of Practice and Theory*, *22*, 53–69.

Hackenbrack, K. (1992). Implications of seemingly irrelevant evidence in audit judgment. *Journal of Accounting Research*, *30*, 54–76.

Hackenbrack, K., & Nelson, M. (1996). Auditors' incentives and their application of financial accounting standards. *The Accounting Review*, *71*, 43–59.

Haynes, C., Jenkins, J., & Nutt, S. (1998). The relationship between client advocacy and audit experience: An exploratory analysis. *Auditing: A Journal of Practice and Theory*, *17*, 88–104.

Iyer, G., Ravrindran, S., & Reckers, P. (2006). Procurement of IT consulting services and firm specific characteristics. *Journal of the Association of Information Systems*, *7*, 207–237.

Kida, T. (1980). An investigation into auditors' continuity and related qualification judgments. *Journal of Accounting Research*, *18*, 506–523.

Kinney, W., Palmrose, Z-V., & Scholz, S. (2004). Auditor independence, non-audit services, and restatements: Was the U.S. Government right? *Journal of Accounting Research*, *42*, 561–588.

Knapp, C., & Knapp, M. (2001). The effects of experience and explicit fraud risk assessment in detecting fraud with analytical procedures. *Accounting, Organizations and Society*, *26*, 25–37.

Larcker, D., & Richardson, S. (2004). Fees paid to audit firms, accrual choices and corporate governance. *Journal of Accounting Research*, *42*, 625–658.

Levitt, A. (2000). Renewing the covenant with investors. Speech at the New York University, Center for Law and Business (May 10).

Lindberg, D., & Beck, F. (2004). Before and after Enron: CPA's views of auditor independence. *The CPA Journal*, *74*, 36–39.

Low, K. (2004). The effects of industry specialization on audit risk assessments and audit-planning decisions. *The Accounting Review*, *79*, 201–219.

Lowe, D., & Pany, K. (1995). CPA performance of consulting engagements with audit clients: Effect on financial statement users' perceptions and decisions. *Auditing: A Journal of Practice and Theory*, *14*, 35–53.

Lowe, D., & Pany, K. (1996). An examination of the effects of type of engagements, materiality, and structure of CPA consulting engagements with audit clients. *Accounting Horizons*, *10*, 32–51.

Markelevich, A., Barragato, C., & Hoitash, R. (2005). The nature and disclosure of fees paid to auditors: An analysis before and after the Sarbanes-Oxley Act. *The CPA Journal*, November (Special Auditing Issue), 6–10.

McCracken, S. (2003). Auditors' strategies to protect their litigation reputation: A research note. *Auditing: A Journal of Practice and Theory*, *22*, 165–179.

McDaniel, L., Martin, R., & Maines, L. (2002). Evaluating financial reporting quality: The effects of financial expertise vs. financial literacy. *The Accounting Review*, *77*, 139–167.

McKeown, J., Mutchler, J., & Hopwood, W. (1991). Towards an explanation of auditor failure to modify the audit opinion of bankrupt companies. *Auditing: A Journal of Practice and Theory*, *10*, 1–23.

Moreno, K., & Bhattacharjee, S. (2003). The impact of pressure from potential client business opportunities on the judgments of auditors across professional ranks. *Auditing: A Journal of Practice and Theory*, *22*, 13–28.

Mutchler, J., Hopwood, W., & McKeown, J. (1997). The influence of contrary information and mitigating factors on audit opinion decisions on bankrupt companies. *Journal of Accounting Research*, *35*, 295–310.

Panel on Audit effectiveness. (2000). *The panel on audit effectiveness report and recommendations*. Stamford, CT: Public Oversight Board.

Pitt, H., & Birenbaum, D. (1997). *Serving the public interest: A new conceptual framework for auditor independence*. Report prepared for the AICPA in connection with a presentation to the Independence Standards Board.

Reynolds, K., Deis, D., & Francis, J. (2004). Professional service fees and auditor objectivity. *Auditing: A Journal of Practice and Theory*, *23*, 29–52.

Robinson, D. (2006). *Tax service fees and auditor independence: Evidence from going-concern audit opinions prior to bankruptcy filings*. Working paper. Arizona State University.

Securities and Exchange Commission (SEC). (2000). *Final Rule*: *Revision of the Commission's Auditor Independence Requirements*. FRR No. 56. Washington, D.C.

Shelton, S., Whittington, O., & Landsittel, D. (2001). Auditing firms' fraud risk assessment practices. *Accounting Horizons*, *15*, 19–33.

Simunic, D. (1980). The pricing of audit services: Theory and evidence. *Journal of Accounting Research*, *18*, 161–190.

Simunic, D. (1984). Auditing, consulting and auditor independence. *Journal of Accounting Research*, *22*, 679–702.
Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *4*, 207–232.
Vera-Munoz, S., Kinney, W., & Bonner, S. (2001). The effects of domain experience and task presentation format on accountants' information relevance assurance. *The Accounting Review*, *76*, 405–429.
Volcker, P. (2000). Securities and Exchange Commission (SEC)'s Public Hearings on *Proposed Auditor Independence Rules*. New York (September 13) 2000.
Waller, W., & Felix, W., Jr. (1984). The auditor and learning from experience: Some conjectures. *Accounting, Organizations and Society*, *9*, 383–406.
Weil, J. (2001). Going concerns: Did accountants fail to flag problems at dot-com casualties? *Wall Street Journal* (February 9), C1.
Wilks, J. (2002). Predecisional distortion of evidence as a consequence of real-time audit review. *The Accounting Review*, *77*, 51–67.
Zimbelman, M. (1996). *Assessing the risk of fraud in audit planning*. Unpublished Ph.D. dissertation, University of Arizona.
Zimbelman, M. (1997). The effects of SAS No. 82 on auditors' attention to fraud risk factors and audit planning decisions. *Journal of Accounting Research* (Supplement), *35*, 75–97.

APPENDIX A. ILLUSTRATION OF EXPERIMENTAL INSTRUMENTS

PLAYWORKS CORPORATION CASE

Playworks Corporation was founded in 1981 and is publicly traded on NASDAQ. Playworks produces toys that it sells to national chain stores and independent retail outlets. About half of Playworks' sales come from infant/toddler toys, with electronic toys making up most of the rest.

The toy industry changes constantly. The major risks in the industry rest with foreign competition, trying to predict consumer trends and in compliance with changing government safety regulations. Litigation costs have risen in recent years industry-wide. Additionally, during the last few years, new domestic and foreign competition has emerged, causing Playworks' market share to decline. New management was brought in 24 months ago with the purpose of improving performance.

As part of the audit, auditors have a duty to determine if a company will likely continue to exist for at least another year. The auditor's evaluation is based on their knowledge of conditions at the time of the audit, as well as management's plans to overcome any problems. If the auditors conclude there is substantial doubt whether the company will exist for at least another year, the auditors must indicate this in the audit report. If the auditors conclude the company has no problems that cannot be resolved by management's plans, a standard, favorable audit report is appropriate.

Manipulated Risk Paragraphs

Version #1:

Playworks had been audited by another large international CPA firm until 2 years ago. Over the last 5 years, Playworks has received clean opinions, although last year, significant internal control deficiencies were noted. Currently, performance ratios are below industry averages and leverage is above average, although the firm continues to report a profit. While KRisk client scaling technology was not available at KPMG when Playworks was accepted as a client 2 years ago, Playworks would have received a KRisk score of 2. Playwork's current KRisk score has declined to 4- (out of 6), indicating a shift (increase in) risk.

Version #2:

Playworks had been audited by another large international CPA firm until 2 years ago. Over the last 5 years, Playworks has received clean opinions, although last year, significant internal control deficiencies were noted. Currently, performance ratios are above industry averages and leverage is below average. New management sought out new auditors when first assuming control 2 years ago. While KRisk client scaling technology was not available at KPMG when Playworks was accepted as a client 2 years ago, Playworks' KRisk score would have been 1 and Playworks' current KRisk score remains at 2, indicating relatively low risk.

Common Footnote:

For client evaluative purposes related to acceptance and audit continuation, KPMG introduced in 2005 the KRisk scale. This 1–6 scale (one least risky, 6 most risky) gives consideration to a number of client dimensions including, but not limited to, financial performance, industry, management integrity, internal control environment and governance structure, and aggressive/conservative past accounting reporting practices.

Manipulated Risk Paragraphs

Version #1:

The engagement is one of the largest in the office when board-of-directors-approved fees for tax services are included. Tax services are about twice audit fees.

Version #2:

The Playworks' engagement is only for audit services. The engagement is not a material source of office revenues.

Summarized, pre-audit financial information:

Sales $ 487,940,000 (down 9%), Plant & Equipment $201,005,107 (constant), Net Income 7,717,020 (down 2%), Total Assets 462,814,288 (up 1%), Cash Flow 2,444,260 (down 17%), Cash 799,423 (down 8%)

During the current year audit, standard audit procedures addressed conditions relevant to whether Playworks could continue to exist for another year. These conditions are summarized below:

SAFETY STANDARDS: The federal government is reviewing current safety standards for toys. Commentators and industry association spokespersons suggest that new laws with tighter standards for infant and toddler toys are probable and new standards for electronic toys are likely. Redesign of several of Playworks leading toys may be necessitated. Litigation effects are indeterminate.

LABOR NEGOTIATIONS: The current labor contract will expire in April. Management and labor have reached oral agreement regarding benefits programs only. Job security and wage increases remain contentious issues. Labor has insisted that unless progress is made, union members will strike. Management argues additional concessions will imperil competitiveness and profitability.

STABILITY OF MAJOR CLIENTS: Playworks' biggest customer, Toyland (about 30% of annual sales) is having significant cash flow problems, as shown by Toyland's payments to Playworks being 60–90 days late during the last 6 months. When pressed, Toyland paid overdue accounts shortly before fiscal year end but after the Christmas season. Significant current receivables continue to exist. Toyland significantly expanded operations immediately before the recent recession and decline in industry sales and simultaneous with the entrance of new competitors in the marketplace. Toyland will likely need to downsize and even then can continue to exist only if it can accomplish a major re-financing in a timely fashion. If Toyland goes bankrupt, ceases or reduces operations, Playworks will lose sales, cash flows, and profits.

TOY INDUSTRY: According to recent government and industry association statistics, retail toy sales were down 7.9% in the last 12 months and expected to remain stagnant over the next year, if not decline further.

SUPPLIERS: In its manufacturing operations, Playworks depends upon various suppliers for its material needs. No disruption of deliveries of supplies is foreseen, unless new cash flow problems emerge.

Given the admittedly incomplete information provided above, do you believe that Playworks warrants a going-concern audit opinion?

Absolutely No	Probably No	Probably Yes	Absolutely Yes
0……..1……..2……..3……..4……..5……..6……..7……..8……..9……..10			

This page intentionally left blank

FINANCIAL REPORTING PRACTICES OF FAMILY FIRMS

Yen H. Tong

ABSTRACT

This paper investigates whether the financial reporting practices of family firms differ from non-family firms. Results indicate family firms have lower absolute discretionary accruals, report fewer small positive earnings surprises compared to non-family firms, have more informative earnings and have less earnings restatements relative to non-family firms. Overall, the findings indicate that the financial reporting practices of family firms are of better quality than those of non-family firms. Better quality financial reporting practices in family firms is consistent with a long-run investment horizon, reputation concerns and better monitoring of managers, and is indicative of less opportunistic rent extraction.

1. INTRODUCTION

In this paper, I examine whether the financial reporting practices of family firms differ from non-family firms. Family firms refer to firms in which founding families have a controlling interest. The founding families typically derive their controlling interest from ownership of large blocks of shares or ownership of shares with special provisions that give them disproportionate voting and control rights. In June 2005, 80-year old John Rigas, the founder

Advances in Accounting, Volume 23, 231–261

ISSN: 0882-6110/doi:10.1016/S0882-6110(07)23009-3

of Adelphia Communications who had run the company for five decades before driving it to bankruptcy in 2002, was sentenced to 15 years in prison after being convicted of his role in a multibillion dollar fraud that led to the collapse of the nation's fifth-largest cable company. In March 2005, a federal jury in New York found former WorldCom CEO and co-founder Bernard Ebbers guilty on charges related to an $11 billion accounting scandal at the telecom giant. Such high-profile accounting frauds raise the question of whether founding families systematically exploit their powerful positions in the companies to extract private benefits and attempt to cover up such rent extraction via improper financial reporting (e.g., Colvin, 2005; Anders, Hymowitz, Lublin, & Clark, 2005).

This question becomes especially important in light of several recent studies which show that family firms are at least as common among public corporations around the world as are widely held and other non-family firms (Shleifer & Vishny, 1986; La Porta, López-de-Silanes, & Shleifer, 1999; Claessens, Djankov, & Lang, 2000; Faccio & Lang, 2002; Anderson & Reeb, 2003). Some of the largest publicly traded firms in the US, such as Wal-Mart Stores and Ford Motors, are controlled by families.

Based on the assumption that founding families have significant controlling interest, two agency conflicts can potentially arise in family firms: (1) the conflict between influential shareholders with concentrated ownership or significant control (the family) and atomistic shareholders and (2) the classic conflict between owners (the family) and the professional manager.[1] The two distinct agency conflicts imply divergent theoretical predictions on the financial reporting practices of family firms. On the one hand, Fama and Jensen (1983) and Demsetz (1983) argue that concentrated ownership facilitates large owners, such as families, to expropriate private benefits from smaller shareholders. Such rent extraction is likely to result in manipulation of financial reporting to cover up evidence of wealth expropriation activities such as related-party transactions and improper consumption of the firm's resources.[2] On the other hand, Stein (1988, 1989) suggests that large shareholders, such as families, have longer investment horizons. This view suggests that the financial reporting practices of family firms are likely to be of higher quality compared to non-family firms because of concerns over reputation and long-term viability of the firm. In addition, founding families with their controlling stake in the firm can better monitor professional managers, thereby mitigating managers' opportunistic behavior, including manipulation of earnings to achieve better compensation outcomes.

In this paper, I investigate which of the two agency conflicts discussed above is likely to dominate and has greater influence on the financial reporting practices of family firms compared to non-family firms. Specifically, I first investigate whether family firms exercise greater reporting discretion relative to their non-family counterpart in the Standard & Poor's 500 (S&P 500) index. I find family firms report lower absolute value of discretionary accruals and fewer small positive earnings surprises. Second, I examine the informativeness of earnings of family firms and find significantly higher coefficients on earnings in a regression of returns on the levels and changes in earnings, indicating that family firms exhibit higher earnings informativeness relative to that of non-family firms. Finally, I examine the propensity of family firms to restate earnings relative to non-family firms because restatements are ex post manifestations of impaired reporting practices. The results show that family firms are less likely to restate earnings relative to non-family firms. Taken together, I interpret the evidence as consistent with family firms having better quality financial reporting compared to non-family firms. The better quality financial reporting is consistent with a long-run investment horizon, concerns over reputation and better monitoring of professional managers and is indicative of less opportunistic rent-extraction activities by founding families.

This paper contributes to the literature on understanding the financial reporting practices of organizational forms with strong insider ownership. The results in this paper are consistent with the findings in Warfield, Wild, and Wild (1995), who provide evidence that managerial ownership is inversely related to the magnitude of accounting accrual adjustments and positively related to the information content of earnings.

This paper also contributes to prior research that examines governance factors affecting the informativeness of earnings, as captured by the coefficient relating returns to earnings (e.g., auditor choice in Teoh & Wong, 1993; institutional holdings in Jiambalvo, Rajgopal, & Venkatachalam, 2002; dual class ownership structure in Francis, Schipper, & Vincent, 2005). In particular, I show that even in the US environment, characterized by relatively strong shareholder protection and financial reporting arrangements, the nature of insider ownership (in particular, family ownership) affects the informativeness of earnings.

This paper is closely related to concurrent papers on family firms by Ali, Chen, and Radhakrishnan (2005) and Chen and Radhakrishnan (2005). Ali et al. (2005) examine the earnings quality and voluntary disclosure practices of family firms, while Chen and Radhakrishnan (2005) examine the relations among family firms, fees paid to auditor, and earnings management. However, this paper differs from Ali et al. (2005) and Chen and Radhakrishnan (2005) in two important respects.[3] First, Ali et al. (2005)

broadly examine the information environment surrounding family firms, including the issuance of management forecasts and the properties of analyst forecasts. Chen and Radhakrishnan (2005) examine the relation between family firms as an internal governance mechanism and auditors as an external governance mechanism. In contrast, I focus on the nature of family firms' reporting practices. Second, although Ali et al. (2005) also examine earnings response coefficients, I consider four other proxies for financial reporting quality not examined in their paper – discretionary accruals, earnings smoothness relative to cash flows, frequency of small positive earnings surprises, and earnings restatements.[4] I believe that this paper, along with Ali et al. (2005) and Chen and Radhakrishnan (2005), are complimentary and together facilitate a better understanding of the financial reporting practices of family firms. Finally, this paper complements finance and economics research on the implications of family firms for firm performance. For example, Anderson and Reeb (2003) and Villalonga and Amit (2006) find that family firms outperform non-family firms in terms of operating and stock return performance.

The rest of the paper is organized as follows: Section 2 reviews prior literature; Section 3 delineates the sampling, research design, and variable measurements; Section 4 presents the empirical findings; and Section 5 concludes.

2. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

Studies in accounting literature on the effects of ownership structures typically rely on the premise that one or more aspects of these ownership structures influence the quality of accounting reports. For example, Dhaliwal, Salamon, and Smith (1982) assume that managerial opportunism decreases with managerial ownership and find that owner-managed firms are less likely to choose income-increasing depreciation methods. Warfield et al. (1995) predict and find an inverse relation between the extent of reporting discretion, proxied by the absolute value of discretionary accruals, and the extent of managerial ownership. Warfield et al. (1995) also find that the earnings response coefficient obtained from a regression of returns on earnings increases, on average, with the level of managerial ownership. Alternatively, Fan and Wong (2002) examine East Asian firms characterized by concentrated family holdings and hypothesize that concentrated ownership is associated with lower credibility of earnings because the controlling shareholder has the incentive to misreport accounting information for self-interested purposes and to publicly disclose as little proprietary information as possible.

Consistent with prior studies that assume ownership structures affect reporting quality, this paper examines the effects of family firms ownership structure on financial reporting practices. Family firms are both prevalent and substantial among US firms. Shleifer and Vishny (1986) report that founding families have significant stakes in one-third of the largest US companies and control nearly 20% of all board seats in these firms. Villalonga and Amit (2006) document that family firms comprise 37% of Fortune 500 firms from 1994 to 2000 and own an average of 16% of the equity of these firms.[5] Anderson and Reeb (2003) show that family ownership is found in one-third of firms on the S&P 500 index and such ownership accounts for 18% of outstanding equity in the index from 1992 through 1999.

In the family firm setting, agency conflicts exist not only between family owners and managers, but also between family owners as shareholders with controlling interest and all other minority shareholders. The unique position of founding families is a double-edged sword. On the one hand, it affords family owners significant influence and control over management to effectively monitor them and mitigate potential managerial expropriation. On the other hand, their controlling interest places them at a significant advantage over atomistic shareholders and could therefore provide greater opportunities for private rent extraction. As follows, I discuss the implications of these two forms of agency conflicts and their potential effects on the financial reporting practices of family firms.

The long-run investment horizon of family owners is likely to discourage family firms from engaging in myopic and value-destructing rent seeking behavior (Stein, 1988, 1989). Indeed, one of the biggest strategic advantages of a family firm is that family owners can have passion for the enterprise that goes far beyond that of any hired executives. To counter a perception of low accountability and to entice investors to buy non-controlling interests, family firms have incentives to provide more precise and transparent earnings. Therefore, a family firm is less likely to engage in low-quality financial reporting practices in order to maintain its reputation and to facilitate the long-term viability of the firm. Consistent with this, Anderson and Reeb (2003) document that family firms significantly outperform their non-family counterpart S&P 500 firms in profitability and stock return performance. They conclude that the classic owner–manager conflict in non-family firms is more costly than the conflict between family and non-family shareholders in founder-CEO firms.

In addition to a long-run investment horizon, family owners are better able to mitigate management myopia because they can more effectively monitor professional managers (Demsetz & Lehn, 1985). For example, Weber, Lavelle, Lowry, Zellner, and Barrent (2003) argue that family owners are often as knowledgeable as management about the firm and therefore, provide effective checks on professional managers. In addition, Weber et al. (2003) report that the large personal and financial stakes in the firm give family owners a powerful incentive to hold management accountable. As a result, family owners who sit on the board of directors, unlike some independent directors, often play a proactive role to monitor management. The better monitoring of management is likely to mitigate managerial opportunistic behavior designed to maximize the managers' own wealth, including activities such as manipulation of financial reporting. The mitigation of managerial manipulation of accounting numbers is likely to result in better quality financial reporting practices by family firms.

However, Fan and Wong's (2002) findings on concentrated ownership, including family ownership, suggest that family firms would be associated with higher misuse of reporting discretion and lower earnings informativeness. The controlling interest held by founding families and the insularity of family outfits sometimes breeds an apparent disregard for outside shareholders. Founding families are likely to engage in self-dealing behavior by directly expropriating wealth from managers, employees, or other investors, seeking non-profit maximizing objectives, or generally putting their interests over those of the firm's other stakeholders (Faccio, Lang, & Young, 2001; Anderson, Mansi, & Reeb, 2003).

Family firms also represent cases in which many well-recognized corporate governance devices are less prevalent compared to non-family firms. For example, Barclay and Holderness (1989) observe that the presence of large shareholders deters bidding by outside agents, suggesting that the market for corporate control is less effective in constraining family actions. Shivdasani (1993) suggests that unaffiliated blockholder ownership is substantially lower in family firms than in non-family firms, indicating absence of external discipline to control family actions.

The family's controlling interest combined with potentially weaker corporate governance structures places founding families in an extraordinarily powerful position to seek private gains at the expense of other atomistic shareholders (Fama & Jensen, 1983; Demsetz, 1983). These rent-extraction activities likely lead to accounting manipulation to conceal the true financial situation of the firm and the magnitude of the wealth expropriation by founding families. For example, actions such as manipulating reported earnings to show a string of consecutive positive earnings surprises or managing reported earnings upwards that are not supported by underlying fundamentals, are likely to present the firm's financial performance in a more favorable light than is warranted. Such actions are designed to mislead, at least over a period of time, investors and other stakeholders while the founding families continue to engage in inappropriate wealth expropriation that adversely impact the underlying financial performance.[6]

In summary, two possible scenarios emerge. In the first scenario, interests in the long-term viability of the firm, concerns over reputation and better monitoring of managers by founding families could result in higher quality financial reporting by family firms. In the second scenario, attempts to mislead other stakeholders about the financial performance of the firm and to cover up the extent of wealth expropriation by founding families might lead to lower quality financial reporting. Hence, given the competing predictions on the expected quality of financial reporting by family firms, I test the following hypothesis stated in the null:

Hypothesis. *There is no difference in the quality of financial reporting practices of family firms compared to non-family firms.*

3. SAMPLE AND RESEARCH DESIGN

3.1. Sample

I obtain the sample of 177 family firms from *Business Week* (BW) November 10, 2003. Specifically, BW defines a family company as one where founders or descendants continue to hold positions in top management, are on the board, or are among the company's largest stockholders. BW uses regulatory filings, company websites, and corporate histories to further confirm their identification of the 177 family companies in the S&P 500 stock index as of July, 2003.[7] The control sample consists of all other S&P 500 firms that otherwise satisfy data requirements.[8] The sampling period is from 1992 to 2003. I require firms to have data on Compustat and CRSP to construct the test variables. I obtain 3,054 firm-year observations for the reporting discretion tests from 1992 to 2003, of which 1,065 observations are family firm observations and 1,989 are non-family firm observations.

For earnings informativeness tests, the sample size is reduced to 3,040 firm-year observations because of additional data requirements. Of the 3,040 observations, 1,057 are family while 1,983 are non-family firms.

Data on restatements come from the General Accounting Office's (GAO) study on financial statement restatements. The GAO database identifies 919 restatement firm-years, representing over 800 distinct firms. As the GAO database contains firm-year observations from January 1, 1997 to June 30, 2002 (compared to my sampling period from 1992 to 2003), I obtain 1,664 firm-year observations after I merge the GAO restatement sample with the reporting discretion test sample. Of the 1,664 firm-year observations, 66 are restatement observations and 14 of these are attributable to family firms.

3.2. Research Design

I use five proxies for the quality of financial reporting in order to test the hypothesis on differences in financial reporting practices of family compared to non-family firms. The first three proxies measure the magnitude of reporting discretion. They are the absolute magnitude of discretionary accruals, the smoothness of earnings relative to cash flows and the propensity of firms to report small positive earnings surprises. I examine discretionary accruals because greater absolute levels of discretionary accruals are assumed to be associated with lower financial reporting quality as actual reported accruals deviate from expected accruals arising from the underlying fundamentals of the firm (Warfield et al., 1995; Dechow, Sloan, & Sweeny, 1996; Bowen, Rajgopal, & Venkatachalam, 2005). I examine smoothness of earnings because earnings patterns that exhibit significantly lower variance compared to cash flows patterns has been interpreted as evidence of lower reporting quality because earnings are artificially smooth (Hunt, Moyer, & Shevlin, 1997; Pincus & Rajgopal, 2002). I also investigate frequency of positive earnings surprises because evidence presented by Burgstahler and Dichev (1997), DeGeorge, Patel, and Zeckhauser (1999), and Matsumoto (2002) suggests that managers use reporting discretion to avoid reporting negative earnings surprises.

I provide evidence based on the above three proxies to be consistent with a long tradition in accounting literature for correlating aspects of the organizational status with proxies for reporting discretion derived from reported earnings numbers.[9] However, prior research has also questioned the statistical power of the three proxies for inferring the quality of financial reporting practices (e.g., see Dechow et al., 1995; Healy & Wahlen, 1999;

Dechow & Skinner, 2000; Durtschi & Easton, 2005). As such, I use two other alternative proxies for the quality of financial reporting practices – the informativeness of earnings and the propensity of family firms to restate GAAP earnings.

The test on earnings informativeness follows research by Holthausen and Verrecchia (1990), Teoh and Wong (1993), and Francis et al. (2005). In particular, Holthausen and Verrecchia (1990) model the determinants of the magnitude of the price response to an earnings release (earnings informativeness) as an increasing function of the amount of prior uncertainty about firm value and a decreasing function of the noise (or the lack of credibility) of the earnings signal. Examining earnings informativeness therefore provides evidence on the relation between the credibility of earnings signal and the incentives of family firms to report transparent earnings.

Examining earnings restatements is useful because restatements represent violations of GAAP in previous periods and are known to result in ex post manifestations of loss in shareholder wealth. For example, the GAO report on earnings restatements finds an average negative 9% loss in market value of restating firms over a 3-day period surrounding the restatement announcement. Hence, I evaluate whether family firms are likely to report an earnings restatement as often as non-family firms. In the next three sections, I discuss the measurement of the five proxies for financial reporting practices and the associated control variables.

3.2.1. Reporting Discretion

Consistent with Bowen et al.'s (2005) (hereafter referred as BRV) proxies for reporting discretion, I examine whether family firms exhibit differential reporting discretion relative to non-family firms in terms of (1) absolute value of discretionary accruals obtained using the modified Jones model (|ABACC|); (2) smoothing of earnings via accruals (SMOOTH); and (3) avoiding earnings decreases by reporting small quarterly positive earnings surprises (FREQ).[10]

Specifically, I estimate the following regression:

$$\mathrm{RD}^{\mathrm{mean}}_{jt+1\sim 3} = \alpha + \beta f_{jt} + \sum_{e=1}^{N} \lambda^{e}_{jt} X^{e}_{jt} \tag{1}$$

Where $\mathrm{RD}^{\mathrm{mean}}_{jt+1\sim 3}$ are reporting discretion proxies, f_{jt} an indicator variable coded as 1 for family firms and 0 otherwise, X^{e}_{jt} a vector of control variables, e is the indicator superscript that runs from $e = 1$ to $e = N$ as per the summation function and $t = 1992$–2000. Following BRV (2005) I measure the right-hand side variables each year for 9 years from 1992 to 2000, and measure the reporting discretion variables over the subsequent 3 years using nine 3-year rolling window averages (1993–1995, 1994–1996, 1995–1997, 1996–1998, 1997–1999, 1998–2000, 1999–2001, 2000–2002, 2001–2003). If family firms are less (more) likely to engage in reporting discretion, then β should be negative (positive). As in BRV (2005), I capture reporting discretion in three ways. To measure |ABACC|, I first estimate normal accruals each year for 11 years (1993–2003) for each two-digit SIC code using the cross-sectional modified Jones model after controlling for operating cash flows. |ABACC| is the 3-year rolling average of the absolute value of discretionary accruals obtained by subtracting the estimated normal accruals from total accruals. I measure earnings smoothing (SMOOTH) as the ratio of the standard deviation of quarterly cash flows scaled by the standard deviation of quarterly net income (Leuz, Nanda, & Wysocki, 2003; BRV 2005). A ratio greater than 1 indicates greater variability of cash flows over earnings, and a higher ratio is consistent with greater smoothing of earnings using accruals. I measure the standard deviation using the same nine 3-year rolling windows as I use in capturing |ABACC|. I capture the frequency of small positive earnings surprises (FREQ) as the number of times the seasonally adjusted quarterly earnings scaled by beginning total assets falls within the (0.00–0.0025) window, measured over the same nine 3-year rolling windows as in |ABACC|.[11]

Following prior literature that examines family firms, I include several control variables in the regression of reporting discretion on family firms (Ali et al., 2005; Anderson & Reeb, 2003; Villalonga & Amit, 2006). Since Villalonga and Amit (2006) find that family firms have significantly lower debt than non-family firms and that prior research documents that managers have incentives to manipulate earnings either to avoid violating debt-covenants or to prevent adverse effects on their debt ratings, I include a leverage variable, measured as the ratio of long-term debt to total assets (DeFond & Jiambalvo, 1994; Minton & Schrand, 1999). Following prior research I expect a positive association between this variable and reporting discretion.

Ali et al. (2005) show that family firms have higher growth than non-family firms based on the market-to-book ratio. Growth firms have higher incentives to use reporting discretion to either meet earnings benchmarks or to smooth earnings. Skinner and Sloan (1999) find that the market severely punishes growth firms for negative earnings surprises, and earlier research also documents that earnings volatility increases perceived firm risk (Beaver, Kettler, & Scholes, 1970). I use the book-to-market ratio as a proxy for growth and since a higher ratio indicates lower growth, I expect a negative relation between book-to-market ratio and reporting discretion.[12]

Prior research suggests that frequent access to capital markets provides managers with incentives to influence reported earnings numbers (Frankel, McNichols, & Wilson, 1995; Teoh, Welch, & Wong, 1998a, 1998b). Ali et al. (2005) show that family firms are less likely to issue debt or equity compared to non-family firms. Following Dechow et al. (1996), I measure a firm's ex ante demand for financing and access to capital markets as a firm's free cash flow (FCF) scaled by current assets. I define FCF as the difference between cash flow from operations for year t–1 and the past 3-year average ($t-1$, $t-2$, $t-3$) of the firm's capital expenditures scaled by current assets at $t-1$. I set a dummy variable (D_{CAP}) to one if the FCF is less than -0.50 and zero otherwise. I expect a positive association between D_{CAP} and reporting discretion.

I also include a control for firm size because Ali et al. (2005) and Anderson and Reeb (2003) find family firms are significantly smaller than non-family firms. Smaller firms face less political costs and hence might have less incentive to exercise reporting discretion to reduce unwanted political visibility (Watts & Zimmerman, 1990). I use the natural logarithm of sales (LnSALES) to proxy for size and expect a positive association between reporting discretion and LnSALES.

Following Ali et al. (2005) and Anderson and Reeb (2003), I include proxies for risk because family firms and non-family firms are likely to have different risk profiles. Minton and Schrand (1999) find that firms with greater earnings volatility have higher costs of equity and debt capital. Hence, riskier firms might utilize more reporting discretion to reduce the perception of risk or to smooth earnings and lower their cost of equity capital (Warfield et al., 1995). I measure risk using both the standard deviation of quarterly operating cash flows computed over the 3-year window prior to the window over which the reporting discretion variable is computed (STDCFO) and the annualized standard deviation of monthly returns calculated over the 60 months prior to the fiscal year end month (STDRET).

Kothari, Leone, and Wasley (2005) argue that tests related to reporting discretion that do not control for performance are often misspecified. Anderson and Reeb (2003) find that family firms have better performance in terms returns on assets (ROA) compared to non-family firms. To control for the effect of performance on reporting discretion, I include return on total assets in the model. Following Anderson and Reeb (2003), I include the natural log of the number of years since the firms' inception to control for firm age (LnAGE) because they find that family firms are on average younger than non-family firms. I do not have directional predictions for the signs of the coefficients on ROA and LnAGE.

I also include two control variables to proxy for external influence on a firm's financial reporting practices. First, following Anderson and Reeb (2003), I include a proxy for institutional ownership as institutional owners are often characterized as sophisticated investors who can potentially monitor abuse of reporting discretion by managers (Lev, 1988; Jiambalvo et al., 2002). However, another body of literature has argued that some institutional investors are "transient owners" who are overly focused on short-term earnings and pressure managers to deliver earnings momentum via abuse of reporting discretion (Bushee, 1998; Graham, Harvey, & Rajgopal, 2005). To control for institutional ownership, I include the variable INST that is the proportion of firm's shares held by institutional investors from the SPECTRUM database. Given the above discussion, I do not predict a sign for INST.

Second, auditors have a direct responsibility over the financial reports of a firm, and prior research finds that audit quality increases with auditor's market share (Craswell, Francis, & Taylor, 1995). Therefore, I include a proxy for auditor influence on financial reporting based on auditor specialization. I sort all the firms on COMPUSTAT by their two-digit SIC code. Based on the sort, I define a dummy variable AUDEXP, which is set to one (zero) if the audit firm for a particular company audits more than 15% (less than 15%) of firms in that company's two-digit SIC code (Dunn & Mayhew, 2004). I expect a negative coefficient on AUDEXP.

Finally, I introduce two-digit industry dummies (IND) and time dummies (YEAR) to account for any unobserved variation in the contracting environment of the firm (Himmelberg, Hubbard, & Palia, 1999).

3.2.2. Earnings Informativeness

To examine earnings informativeness, I regress 12- and 15-month abnormal buy-and-hold returns on earnings and change in earnings, after controlling for firm size, book-to-market ratio, leverage, and institutional ownership as follows:

$$\begin{aligned}\mathrm{ABRET}_{jt} = {} & \alpha_0 + \alpha_1 \mathrm{EARN}_{jt} + \alpha_2 f_{jt} \times \mathrm{EARN}_{jt} \\ & + \sum_{k=1}^{5} \alpha_k X_{jt}^k \times \mathrm{EARN}_{jt} + \varepsilon_{jt} \end{aligned} \tag{2a}$$

$$
\begin{aligned}
\mathrm{ABRET}_{jt} = {} & \alpha_0 + \alpha_1 \mathrm{EARN}_{jt} + \alpha_2 f_{jt} \times \mathrm{EARN}_{jt} + \sum_{k=1}^{5} \alpha_k X_{jt}^k \times \mathrm{EARN}_{jt} \\
& + \beta_1 \Delta\mathrm{EARN}_{jt} + \beta_2 f_{jt} \times \Delta\mathrm{EARN}_{jt} + \sum_{k=1}^{5} \beta_k X_{jt}^k \\
& \times \Delta\mathrm{EARN}_{jt} + \upsilon_{jt} \qquad (2b)
\end{aligned}
$$

Where ABRET_{jt} is measured as (1) firm j's 12-month cumulative abnormal buy-and-hold returns beginning 3 months following the end of fiscal year $t-1$ (ABRET_12) and, (2) 15-month cumulative abnormal buy-and-hold returns beginning at the end of fiscal year $t-1$ (ABRET_15), EARN_{jt} and $\Delta\mathrm{EARN}_{jt}$ are firm j's income and change in income for year t, f_{jt} an indicator variable coded as 1 for family firms and 0 otherwise, and X_{jt}^k a vector of five control variables defined as follows:

$X_{jt}^1 = \mathrm{LOSS} = 1$ if $\mathrm{EARN}_{jt} < 0$, and 0 otherwise

$X_{jt}^2 = \mathrm{LnSALES}_{jt} =$ log of firm j's sales *in* year $t-1$

$X_{jt}^3 = \mathrm{BM}_{jt} =$ firm j's book-to-market ratio in year $t-1$

$X_{jt}^4 = \mathrm{LEV}_{jt} =$ firm j's leverage in year $t-1$

$X_{jt}^5 = \mathrm{INST}_{jt} =$ percent of firm j's shares held by institutions in year $t-1$

In Eq. (2a), the test of the relative earnings informativeness of family firms versus that of non-family firms rests on the sign of α_2: if earnings of family firms are more informative than that of non-family firms, then $\alpha_2 > 0$; if they are less informative, then $\alpha_2 < 0$. In Eq. (2b), the test of the relative earnings informativeness rests on $\alpha_2 + \beta_2$: more earnings informativeness implies $\alpha_2 + \beta_2 > 0$, while less earnings informativeness implies $\alpha_2 + \beta_2 < 0$. Consistent with Francis et al. (2005), I expect negative coefficients on the interaction variables between LOSS, LnSALES, BM, LEV and the earnings variables, and have no directional prediction for the interaction variable between INST and EARN/ΔEARN.

3.2.3. Earnings Restatement

I use the following regression model to examine the propensity of family firms to restate earnings compared to non-family firms.

$$
Prob(\mathrm{RESTATE}_{jt+1} = 1) = F\left\{\alpha + \beta f_{jt} + \sum_{e=1}^{N} \lambda_{jt}^e X_{jt}^e + \upsilon_{jt}\right\} \qquad (3)
$$

RESTATE is an indicator variable coded as 1 if the firm announces an earnings restatement in year $t+1$ per the GAO database, and 0 otherwise. All the right-hand side variables are defined as in equation (1). I should observe more restatements in family firms if they are relatively poorly managed and engage in more rent-extraction activities and manipulation of financial reporting compared to non-family firms. This is because subsequent revelations of misleading financial reports are likely to lead to restatements (e.g., the case of WorldCom in 2002). Conversely, if family firms are better managed because of concerns over reputation and with better alignment of owner's and manager's interests, I expect to find family firms to have lesser likelihood of earnings restatements. A positive coefficient on f_{jt} is consistent with family firms being more likely to restate earnings while a negative coefficient is consistent with family firms being less likely to restate earnings.

4. EMPIRICAL FINDINGS

4.1. Descriptive Statistics

Table 1 lists and provides detail definition of the variables discussed in Section 3.

Table 2 presents descriptive statistics on the mean, median, and standard deviations of the variables listed in Table 1. Table 2, Panel A reports the statistics on the three individual measures of reporting discretion over the nine rolling windows from 1993 to 2003. The 3-year average of the absolute value of discretionary accruals represents about 18% of lagged assets for both the average family firm and non-family firm and the difference in means is not significant. The median absolute value of discretionary accruals for family firms is significantly higher than that of the non-family firms. The relatively smaller median values of |ABACC|and the large standard deviations suggest the presence of extreme values in both samples. I test the sensitivity of removing the extreme values in the OLS regressions and logistic regression in section 4.6.

The mean (median) SMOOTH ratio is 3.02 (2.25) for family firms and 2.96 (2.02) for non-family firms, suggesting that on average cash flows are two to three times as variable as earnings for both family and non-family firms. The difference in medians of SMOOTH for the two sets of firms is significant at the 1% level. Finally, the average family (non-family) firm reports a small positive quarterly earnings surprise ~11.9% (17.4 %) of the

Table 1. Variable Definition.

Variables	Definition
Reporting discretion variables	
$\|ABACC\|$	Absolute value of abnormal accruals (scaled by lagged total assets) computed as per the modified Jones (1991) model after controlling for change in operating cash flows. The normal accruals model is estimated for every calendar year. $\|$ABACC$\|$ is estimated annually for each two-digit SIC code over 11 years (1993–2003) and then averaged across nine 3-year windows: 1993–1995, 1994–1996, 1995–1997, 1996–1998, 1997–1999, 1998–2000, 1999–2001, 2000–2002, 2001–2003. Accruals are defined as earnings (#18) – cash flows adjusted for extraordinary items (#308–124).
SMOOTH	Standard deviation of quarterly cash flows adjusted for extraordinary items (quarterly #108–78) scaled by the standard deviation of quarterly net income (quarterly #76) computed over nine 3-year windows: 1993–1995, 1994–1996, 1995–1997, 1996–1998, 1997–1999, 1998–2000, 1999–2001, 2000–2002, 2001–2003.
FREQ	Frequency of times the firm reports a small quarterly earnings surprise over the 3-year windows, 1993–1995, 1994–1996, 1995–1997, 1996–1998, 1997–1999, 1998–2000, 1999–2001, 2000–2002, 2001–2003. A small surprise occurs when the change in seasonally lagged quarterly earnings after tax scaled by total assets at the end of quarter q-5 falls within the range of 0.00–0.0025.
DISCIND	Average of the fractional ranks of $\|$ABACC$\|$, SMOOTH, and FREQ. Fractional ranks for each individual measure is obtained for every 3-year window by first ranking each measure from smallest to largest. The ranks are then rescaled by the total number of observations in each 3-year window to obtain fractional ranks. The fractional ranks lie between 0 and 1.
General firm characteristics	
MVE	Market value of equity (#199 $\times$ #25)
SALES	Total sales (#12)
ASSETS	Total assets (#6)
F	Indicator variable set equal to one if family firm, zero otherwise
OWN	Family ownership obtained from proxy statements deflated by total shares outstanding (#25).
Control variables	
LEV	Proportion of long-term debt (#9) to total assets (#6)
BM	Ratio of book value (#60) to market value of equity (#199 $\times$ #25).
D_{CAP}	Indicator variable set to one if the free cash flow (FCF) measure is less than -0.50 and zero otherwise. The FCF measure is the difference between cash flow from operations (#308) for current year and the current and past 2-year average of the firm's capital expenditure (#128) scaled by current year's current assets (#4).

Table 1. (*Continued*)

Variables	Definition
LnSALES	Natural logarithm of total sales (#12).
STDCFO	Standard deviation of quarterly cash flows from operations (#108) over current and 11 prior quarters.
STDRET	Annualized standard deviation of monthly returns calculated over current and prior 59 months as of fiscal year end month.
ROA	Income before extraordinary items (#18) scaled by lagged total assets (#6).
LnAGE	Natural logarithim of firm's age. Firm's age (AGE) is determined as the number of years the firm appears on CRSP or Compustat, whichever is longer.
INST	Ratio of institutional share ownership to total shares outstanding.
AUDEXP	Indicator variable that is set to one (zero) if the audit firm for a particular company audits more than 15% (less than 15%) of firms in that company's two-digit SIC code.
Variables in earnings informativeness tests and earnings restatement tests	
ABRET_12	Twelve-month abnormal buy-and-hold returns calculated using cumulated monthly returns beginning 3 months following the end of fiscal year $t-1$ less the corresponding cumulated monthly value-weighted market portfolio on CRSP.
ABRET_15	Fifteen-month abnormal buy-and-hold returns calculated using cumulated monthly returns beginning from the end of fiscal year $t-1$ less the corresponding cumulated monthly value-weighted market portfolio on CRSP.
EARN	Income before extraordinary items (#18) scaled by beginning market value of equity (#199 × #25).
ΔEARN	Current minus prior period income before extraordinary items (#18) scaled by beginning market value of equity (#199 × #25).
LOSS	Indicator variable that is set equal to one if income before extraordinary items (#18) is negative, zero otherwise.
RESTATE	Indicator variable set equal to one if the firm appears in the GAO restatement database, zero otherwise.

Note: Numbers in parentheses refer to Compustat annual data items, unless specified as quarterly data items.

time over a 3-year window. The mean difference test shows that non-family firms report significantly higher frequencies of small positive quarterly earnings surprises. Overall, the differences in mean and median results are mixed on whether family firms exercise greater reporting discretion compared to non-family firms.

Table 2, Panel B reports statistics on general firm characteristics. Consistent with findings in Anderson and Reeb (2003), family firms are

Table 2. Descriptive Statistics.

Variables	Non-Family Firms ($N = 1{,}989$)			Family Firms ($N = 1{,}065$)			Differences	
	Mean	Standard deviation	Median	Mean	Standard deviation	Median	Mean	Median
Panel A: Reporting discretion measures								
|ABACC|	0.189	0.417	0.059	0.181	0.310	0.065	0.008	−0.006*
SMOOTH	2.967	3.759	2.023	3.027	3.428	2.251	−0.060	−0.228*
FREQ	0.174	0.179	0.083	0.119	0.147	0.083	0.055*	0.000
Panel B: General firm characteristics								
MVE *($M)*	13,246	28,831	4,392	12,273	32,111	4,024	973	368**
SALES *($M)*	8,760	13,699	4,421	5,904	12,705	2,466	2,856*	1,955*
ASSETS *($M)*	8,998	14,709	4,216	5,007	8,353	2,431	3,991*	1,785*
OWN[a] (%)	0.000	0.000	0.000	0.100	0.135	0.041	−0.100*	−0.041*
Panel C: Control variables								
Economic determinants variables								
LEV	0.189	0.121	0.178	0.152	0.141	0.132	0.037*	0.046*
BM	0.409	0.262	0.356	0.307	0.218	0.259	0.102*	0.097*
D_{CAP}	0.014	0.116	0.000	0.010	0.101	0.000	0.004	0.000
LnSALES	8.383	1.188	8.394	7.747	1.440	7.810	0.636*	0.584*
STDCFO	0.026	0.024	0.020	0.034	0.033	0.024	−0.008*	−0.004*
STDRET	0.296	0.114	0.269	0.383	0.174	0.343	−0.087*	−0.074*
ROA	0.059	0.074	0.057	0.079	0.081	0.080	−0.020*	−0.023*
AGE	33.68	12.48	35.00	21.62	14.05	21.00	12.07*	14.00*
INST	0.513	0.260	0.590	0.513	0.245	0.554	0.000	0.036***
AUDEXP	0.551	0.497	1.000	0.603	0.489	1.000	−0.052*	0.000*

Table 2. (*Continued*)

Panel D: Earnings informativeness variables

Variables	Non-Family Firms ($N = 1{,}983$)			Family Firms ($N = 1{,}057$)			Differences	
	Mean	Standard deviation	Median	Mean	Standard deviation	Median	Mean	Median
ABRET_12	−0.013	0.391	−0.045	0.233	1.152	0.037	−0.246*	−0.082*
ABRET_15	−0.004	0.472	−0.038	0.321	1.349	0.075	−0.325*	−0.113*
EARN	0.044	0.074	0.054	0.046	0.064	0.049	−0.002	0.005*
ΔEARN	0.008	0.103	0.006	0.011	0.064	0.008	−0.003	−0.002*
LOSS	0.108	0.310	0.000	0.078	0.268	0.000	0.030*	0.000*
LnSALES	8.383	1.190	8.386	7.749	1.445	7.821	0.634*	0.565*
BM	0.409	0.262	0.356	0.303	0.213	0.258	0.106*	0.098*
LEV	0.189	0.121	0.178	0.152	0.142	0.133	0.037*	0.045*
INST	0.513	0.261	0.590	0.515	0.245	0.555	−0.002	0.040***

Note: See Table 1 for variable definitions.
*Significance at 1% (Two-tailed test).
**Significance at 5% (Two-tailed test).
***Significance at 10% (Two-tailed test).
[a]There are only 716 observations for OWN for family firms.

smaller than non-family firms with a mean market capitalization (mean assets) of \$12.3 billion (\$5.0 billion) compared with \$13.2 billion (\$9.0 billion) for non-family firms. Family firms also seem to generate significantly less revenues (mean sales of \$5.9 billion) compared to non-family firms (mean sales of \$8.7 billion). I also gather data on the percentage family ownership in family firms using firms' proxy statements on Edgar and report the statistics on the 716 observations with available data (Edgar starts in 1994). The mean (median) family ownership is 10.0% (4.1%) in the sample.[13]

Table 2, Panel C reports descriptive statistics on the control variables used in the reporting discretion and earnings restatements regressions. Family firms have significantly lower leverage, lower book-to-market ratios and also greater standard deviation of cash flows and returns. Family firms also perform significantly better in terms of ROA than non-family firms (mean of 7.9 vs. 5.9%), consistent with the evidence reported in recent literature (e.g., Anderson & Reeb, 2003; Villalonga & Amit, 2006). Consistent with Anderson and Reeb (2003), I find family firms on average are significantly younger than non-family firms, with a mean family firm age of 21.62 years versus a mean non-family firm age of 33.68 years. The median, but not the mean, percentage of shares held by institutions for non-family firms is significantly higher than that of the family firms. The auditors of family firms exhibit significantly more audit expertise than those of non-family firms.

Table 2, Panel D presents descriptive statistics on the variables used in earnings informativeness tests. On average, family firms earn significantly higher 12- and 15-month buy-and-hold abnormal returns than non-family firms (23.3 vs. −1.3% for 12-month returns, and 32.1 vs. −0.4% for 15-month returns), and experience less losses. These results are consistent with the better performance of family firms on accounting performance variables in Panels B and C. Hence, family firms appear to outperform non-family S&P 500 firms both in terms of accounting and stock performance at a univariate level.

4.2. Correlation Statistics

Table 3 presents the correlation statistics. Results indicate that |ABACC| is significantly negatively correlated with the two other individual reporting discretion measures SMOOTH ($\rho = -0.21$) and FREQ ($\rho = -0.26$), whereas SMOOTH and FREQ are positively correlated ($\rho = 0.29$). The family

Table 3. Correlation Statistics Spearman Correlation among Reporting Discretion, Family Firms and Control Variables ($N = 3,054$).

	|ABACC|	SMOOTH	FREQ	F	LEV	BM	D_{CAP}	LnSALES	STDCFO	STDRET	ROA	LnAGE	INST	AUDEXP
|ABACC|														
SMOOTH	−0.217*													
FREQ	−0.264*	0.291*												
F	0.050*	0.038*	−0.168*											
LEV	−0.194*	−0.051*	0.183*	−0.170*										
BM	−0.295*	0.088*	0.207*	−0.220*	0.236*									
D_{CAP}	−0.002	−0.034*	−0.005	−0.014	0.092*	0.060*								
LnSALES	−0.134*	0.062*	0.227*	−0.206*	0.195*	0.047*	−0.115*							
STDCFO	0.148*	0.228*	−0.165*	0.164*	−0.299*	−0.162*	−0.041*	−0.169*						
STDRET	0.289*	−0.109*	−0.321*	0.286*	−0.162*	−0.056*	0.049*	−0.393*	0.347*					
ROA	0.121*	0.033	−0.102*	0.197*	−0.467*	−0.548*	−0.105*	−0.075*	0.148*	−0.050*				
LnAGE	−0.149*	0.032	0.187*	−0.392*	0.196*	0.158*	−0.066*	0.476*	−0.284*	−0.493*	−0.124*			
INST	0.125*	0.022	−0.037*	−0.032	0.040*	−0.075*	0.039*	−0.074*	0.110*	0.079*	0.078*	−0.004		
AUDEXP	−0.065*	−0.035*	0.160	0.049*	0.085*	0.096*	0.002	0.053*	−0.043*	0.050*	−0.061*	0.018	0.066*	

Note: See Table 1 for variable definitions.
*Significance at 5% (Two-tailed).

ownership indicator, F, is negatively correlated with the frequency of positive earnings surprises but is positively correlated with both absolute discretionary accruals and earnings smoothness. The correlations between F and the other control variables are largely consistent with what is reported in Table 2. For example, F is negatively correlated with leverage (LEV), book-to-market (BM), sales (LnSALES), and age (LnAGE) while it is positively correlated with ROA and volatility of cash flows and returns (STDCFO, STDRET). Hence, the univariate correlation results provide mixed results on whether family firms exercise greater reporting discretion compared to non-family firms. Since univariate analysis does not take into account all variables simultaneously, a multivariate analysis is necessary to examine the differential extent of reporting discretion by family firms versus non-family.

4.3. Reporting Discretion Results

Table 4 provides the multivariate analysis on the association between family firms and reporting discretion as proxied by the absolute value of discretionary accruals (|ABACC|), the relative variability of earnings compared to cash flows (SMOOTH) and, frequency of small positive quarterly earnings surprises (FREQ).[14] Results based on |ABACC| and FREQ as dependent variable both show that, Ceteris paribus, family firms engage in significantly less reporting discretion ($t = -2.26$ and $t = -4.49$, respectively). Results using SMOOTH as the dependent variable indicate no significance on the family indicator variable. Hence, the evidence is consistent with family firms exhibiting less absolute discretionary accruals and reporting fewer small positive quarterly earnings compared to non-family firms. Based on prior research that suggests large absolute discretionary accruals and the propensity to avoid negative earnings surprises are associated with accounting manipulation leading to low-quality financial reporting, the evidence in Table 4 provides some support for family firms having higher quality reporting practices than non-family firms (Dechow et al., 1996; Burgstahler & Dichev, 1997; Degeorge et al., 1999).

4.4. Earnings Informativeness Results

Results on earnings informativeness are reported in Table 5. For Eq. (2a), the coefficient on $f \times \text{EARN}$ ($\hat{\alpha}_2$) is highly significantly positive using both

Table 4. OLS regression Results of Tests on Reporting Discretion (White Corrected t-Statistics in Parentheses).

Variable	Predicted Sign	|ABACC|	SMOOTH	FREQ
Intercept	+/–	0.701	2.22	−0.013
		(3.29)*	(2.21)*	(−0.36)
F	+/–	−0.031	0.098	−0.279
		(−2.26)**	(0.87)	(−4.49)*
LEV	+	−0.019	−0.337	0.062
		(−0.37)	(−0.69)	(2.30)*
BM	–	−0.036	−0.201	0.052
		(−1.10)	(−0.88)	(4.00)
D_{CAP}	+	−0.108	−0.238	0.032
		(−1.18)	(−0.86)	(1.34)
LnSALES	+	0.001	−0.095	0.010
		(0.18)	(−1.57)	(73.61)*
STDCFO	+	−0.244	36.843	−0.338
		(−0.99)	(6.69)*	(−2.38)*
STDRET	+/–	−0.062	−4.359	−0.198
		(−0.97)	(−5.89)*	(−6.94)*
ROA	+/–	−0.072	−0.881	−0.001
		(−0.89)	(−1.11)	(−0.02)
LnAGE	+/–	−0.023	0.070	−0.012
		(−1.70)***	(0.68)	(−0.97)
INST	+/–	−0.003	−0.21	0.013
		(−0.16)	(−0.80)	(2.16)**
AUDEXP	–	0.011	−0.155	−0.001
		(0.86)	(−1.22)	(−0.36)
N		3,054	3,054	3,054
Adjusted R^2		0.36	0.17	0.23
F-statistics (p-value)		49.12	147.67	374.30
		(0.00)	(0.00)	(0.00)

Note: See Table 1 for variable definitions. Coefficients on industry and time dummies are suppressed.
*,**,***Significance at the 1, 5, and 10% levels based on White-corrected t-statistics. One-tailed for directional predictions, two-tailed otherwise. I do not note the significance level if the coefficients have the sign opposite to my prediction.

the 12- and 15-month abnormal buy-and-hold returns as dependent variables ($t = 2.24$ and $t = 2.04$), consistent with family firms having more informative earnings than non-family firms. Results from estimating Eq. (2b) show that $\hat{\alpha}_2$ is still significantly positive when using 15-months abnormal buy-and-hold returns as the dependent variable ($t = 1.97$), but insignificant using 12-month returns. The coefficient on $f \times \Delta$EARN ($\hat{\beta}_2$) is

Table 5. OLS Regression Results of Tests on Earnings Informativeness (White Corrected *t*-Statistics in Parentheses).

Variable	Predicted Sign	ABRET_12	ABRET_15	ABRET_12	ABRET_15
Intercept	+/−	−0.867	−0.129	−0.062	−0.087
		(−3.14)*	(−3.54)*	(−2.03)*	(−2.28)*
EARN	+	6.361	7.849	4.732	4.587
		(2.07)**	(2.17)*	(1.51)**	(1.13)
$F \times$ EARN	+/−	1.162	1.513	0.892	1.613
		(2.24)**	(2.04)**	(1.53)	(1.97)**
LOSS $\times$ EARN	−	−2.677	−4.118	−2.372	−3.820
		(−3.37)*	(−4.17)	(−2.83)*	(−3.73)*
LnSALES $\times$ EARN	−	−0.324	−0.356	−0.172	−0.099
		(−1.17)	(−1.00)	(−0.62)	(−0.26)
BM $\times$ EARN	−	−2.232	−2.983	−2.864	−3.428
		(−2.87)*	(−2.70)*	(−3.17)*	(−2.77)*
LEV $\times$ EARN	−	−3.787	−4.846	−2.589	−2.768
		(−1.83)*	(−1.66)**	(−1.21)	(−0.88)
INST $\times$ EARN	+/−	0.758	1.747	0.198	1.179
		(0.85)	(1.60)***	(0.21)	(1.03)
ΔEARN	+			2.785	5.345
				(1.08)	(1.48)***
$F \times \Delta$EARN	+/−			0.534	−0.071
				(0.74)	(−0.07)
LOSS $\times \Delta$EARN	−			−0.333	−0.086
				(−0.63)	(−0.11)
LnSALES $\times \Delta$EARN	−			−0.294	−0.468
				(−1.21)	(−1.38)**
BM $\times \Delta$EARN	−			0.728	0.366
				(0.85)	(0.30)
LEV $\times \Delta$EARN	−			−2.056	−3.648
				(−1.12)	(−1.38)**
INST $\times \Delta$EARN	+/−			1.173	1.269
				(1.45)	(1.20)
N		3,040	3,040	3,040	3,040
Adjusted R^2		0.03	0.05	0.04	0.06
F-statistics(p-value)		9.88	12.63	4.14	5.21
		(0.00)	(0.00)	(0.00)	(0.00)
F-statistics (p-value) for $F \times$ EARN $+ F \times \Delta$EARN$=0$				4.52	2.77
				(0.00)	(0.09)

Note: See Table 1 for variable definitions.Coefficients on industry and time dummies are suppressed.

*,**,***Significance at the 1, 5, and 10% levels based on White-Corrected *t*-statistics. One-tailed for directional predictions, two-tailed otherwise. I do not note the significance level if the coefficients have the sign opposite to my prediction.

not significant in either equation. However, the *F*-statistics for the sum of the coefficients on $f \times$ EARN and $f \times \Delta$EARN ($\alpha_2 + \beta_2 = 0$) is significant at less than the 1% level (10% level) for the regression with ABRET_12 (ABRET_15) as the dependent variable.

The evidence in Table 5 indicates that the earnings of family firms are more informative than those of non-family firms, suggesting that family firms provide relatively more credible and transparent financial reports that are informationally useful to equity market participants. This, together with the results documented in Section 4.3 that family firms exhibit less absolute discretionary accruals and report less small positive quarterly earnings surprises versus non-family, lend support to the conjecture that the reporting practices of family firms are of higher quality than non-family firms.

4.5. Earnings Restatement Results

Table 6 show the results based on earnings restatements. Examining earnings restatements is useful because restatements provide ex post confirmation of violations of GAAP in previous periods. The logit model results in Table 6 show a significantly negative coefficient on the family firm indicator variable after controlling for all other variables. This indicates that family firms are less likely to restate their earnings compared to non-family firms. This result based on ex post confirmation of reporting quality, coupled with the results on reporting discretion and earnings informativeness discussed above, lends strong support to the notion that the reporting practices of family firm are of higher quality compared to non-family firms.

4.6. Sensitivity Analysis

I report the OLS results with White-corrected *t*-statistics on the reporting discretion regressions in Table 4. The measurement of the dependent variables uses overlapping measurement intervals, which potentially creates serial correlation in the error terms. I apply three econometric methods to account for serial correlations in the error term: (1) Newey–West adjustment, (2) full information maximum likelihood (FIML) procedure in SAS, and (3) Fama-MacBeth (1973) procedures based on nine annual regressions. The results obtained using all three methods are qualitatively similar to those reported in Table 4. Similarly, I report OLS results with

Table 6. Logit Regression Results of Tests on Earnings Restatement (Wald χ^2 Statistics in Parentheses).

Variable	Predicted Sign	Restate
Intercept	+/−	−6.580
		(22.76)*
F	+/−	−0.772
		(5.00)**
LEV	+	0285
		(0.07)
BM	−	−0.148
		(0.08)
D_{CAP}	+	−14.659
		(0.00)
LnSALES	+	0.339
		(6.93)*
STDCFO	+	−3.597
		(0.317)
STDRET	+/−	2.735
		(10.33)*
ROA	+/−	−3.970
		(4.69)**
LnAGE	+/−	−0.197
		(0.88)
INST	+/−	0.917
		(2.19)
AUDEXP	−	0.071
		(0.06)
N		1,664
Likelihood ratio(p-value)		29.39
		(0.00)

Note: See Table 1 for variable definitions.
*Significance at 1% (χ^2 test).
**Significance at 5% (χ^2 test).
***Significance at 10% (χ^2 test).

White-corrected t-statistics on the earnings informativeness regressions in Table 5. I also apply the three econometric approaches to the earnings informativeness regressions in Table 5 and obtain qualitatively similar results. Hence, both the reporting discretion and earnings informativeness results are robust.

Table 2 shows the existence of potential outliers in the test and control variables. I use two alternative procedures to test the sensitivity of the results to outliers. First, I drop observations with studentized residuals greater than 3 or less than -3 from the OLS and logistic analyses. Alternatively, I winsorize all continuous variables at the 1st and 99th percentile levels. Both procedures yield results qualitatively similar to those reported in Table 4–6.

While the earnings informativeness tests in Table 5 employ abnormal buy-and-hold returns, I also test the sensitivity of the analysis by using raw buy-and-hold returns. Results obtained are qualitatively similar to those reported in Table 5.

5. CONCLUSION

Family firms provide a unique setting for researchers to study the classical owner–manager conflict and the moral hazard issue between family owners and other atomistic shareholders. Ex ante it is not clear which way the double-edged sword cuts: families' historical presence, undiversified equity position and controlling stake place founding families in a powerful position that allows them to monitor managers more effectively but at the same time makes it easier for founding family members to seek private benefits at the expense of other atomistic shareholders. This paper examines family firms relative to their non-family owned counterparts in the S&P 500 from 1992 to 2003 in order to shed further light on which one of the agency issues dominates and has a greater influence on the financial reporting process of family firms.

I examine five measures of financial reporting practices of family firms. First, I compare the levels of absolute discretionary accruals, the variability of earnings relative to cash flows and the frequency of small positive earnings surprises of family firms compared to non-family firms. Results indicate that family firms exhibit lower absolute discretionary accruals and report fewer small positive earnings surprises. As suggested by prior research, lower levels of discretionary accruals and less frequent positive earnings surprises are indicative of *less* opportunistic accounting manipulation (Warfield et al., 1995; Dechow et al., 1996; Burgstahler & Dichev, 1997; Matsumoto, 2002). I also examine the earnings informativeness and the frequency of earnings restatements in family firms. The results indicate that family firms have more informative earnings and less earnings restatements relative to non-family controlled firms. Taken together, the evidence in its entirety is consistent with family firms having higher quality financial reporting compared to non-family firms. The better financial reporting practices are consistent with a long-run investment horizon, concerns over reputation and better monitoring of professional managers. They also suggest less opportunistic rent-extraction activities by founding families.

The evidence in this paper complements the literature in finance that examines whether family ownership creates or destroys value (Anderson & Reeb, 2003; Villalonga & Amit, 2006). These studies in general document that family firms perform better than non-family firms, and that family management adds value, especially when the founder is still on the scene and serves as the CEO or the Chairman of the firm. This paper also extends the literature on the financial reporting practices of organizational forms with strong insider ownership (e.g., Dhaliwal et al., 1982; Warfield et al., 1995; Fan & Wong, 2002; Francis et al., 2005). In addition, this paper adds to the literature on how governance factors affect the informativeness of earnings (e.g., auditor choice in Teoh & Wong, 1993; managerial ownership in Warfield et al., 1995; institutional holdings in Jiambalvo et al., 2002; concentrated ownership in Fan & Wong, 2002; dual class ownership structure in Francis et al., 2005). Finally, together with the concurrent research by Ali, Chen, and Radhakrishnan's (2005) on corporate disclosure practices of family firms, this study facilitates a better understanding of the financial reporting practices of family-owned firms.

This paper takes only the first step toward a better understanding of the issue of whether, on average, family owners utilize their powerful positions for self-serving purposes. Building on prior research, which indicates that family ownership is a more profitable ownership structure, the evidence in this study suggests that family ownership is better at mitigating managerial opportunism. Future research can delve deeper by examining, among other things, any differences in financial reporting behavior between founder–CEO firms and founder-descendent firms, and how ownership versus control makes a difference in family owners' financial reporting practices.

NOTES

1. In the paper, I do not draw the distinction between ownership, control, and management. Instead, I assume that founding families exert strong influence on the running of family firms whether via direct involvement in management or indirectly via common stock ownership or superior voting rights. This assumption is supported by recent empirical evidence (e.g., see Villalonga & Amit, 2006).

2. An anecdotal example of improper consumption of company's resources is John Rigas's ordering of two Christmas trees flown to New York at $6,000 a piece and his purchase of 3,600 acres of timberland at a cost of $26 million to preserve the pristine view outside is home – all charged to Adelphia Communications.

3. My sample period also differs from both Ali et al. (2005) and Chen and Radhakrishnan (2005). Ali et al. (2005) examine a period from 1998 to 2002 while Chen and Radhakrishnan (2005) examine a period after 2001 when fees paid to auditors are made publicly available. My sample period is from 1992 to 2003.

4. Chen and Radhakrishnan (2005) strictly examine the association between family firms and discretionary accruals and how the association is moderated by fees paid to auditors.

5. Villalonga and Amit (2006) also report evidence of family control in excess of ownership; family firms tend to use some control-enhancing mechanism that entitles them to greater voting rights than reflected by their percentage of share ownership.

6. For example, John Rigas of Adelphia Communications was accused of conspiring to hide $2.3 billion of Adelphia Communications debt, stealing $100 million, and lying to investors about the company's financial situation by aggressively inflating the firm's reported earnings in order to cover up such indiscretion.

7. For a complete list of the 177 family firms, please refer to ''Defining Family: How Did BW Come Up With Its List?'', BW (November 13, 2003). I assume that the 177 firms are classified as family firms throughout the sampling period of 1992–2003. If a firm was not in existence at the beginning of the sampling period (e.g., eBay), such a firm is not included in the analysis in those corresponding years.

8. I eliminate firms in the financial service industry (SIC codes 6000–6999) in forming the samples because accruals in the financial service industry are not comparable with accruals in other industries.

9. See Fields, Lys, and Vincent (2001) for a summary of this literature.

10. Table 1 provides a detailed description of the measurement of each variable.

11. Burgstahler and Dichev (1997) use a range of 0–0.01 for annual earnings. As in BRV (2005), I choose 0.0025 as the outer end of the range because I use quarterly data in my analysis.

12. I use the book-to-market instead of the market-to-book ratio to avoid the small denominator problem.

13. The mean for family ownership is small compared to the mean of 16.0% reported by Anderson and Reeb (2003). Anderson and Reeb (2003) calculation of family ownership includes distant relatives and descendants, whose last names might not be the same as the original founding members. Anderson and Reeb (2003) track down these distant relatives and descendants by examining corporate histories (through sources such as Gale Business Resources) and include their shareholdings in their calculation of the overall family ownership. As I only capture the percentage ownership of founding members using proxy statements, the mean for family ownership is likely to be understated when compared to that reported in Anderson and Reeb (2003).

14. I examine whether multicollineraity among variables might influence the regression results on the association between reporting discretion proxies and the family firm indicator variable reported in Table 4. Kennedy (1992) suggests that a variance inflation factor (VIF) greater than ten is indicative of problematic collinearity. The VIFs are less than two for all variables in the regressions reported in Table 4. Coupled with the evidence in Table 3 that no absolute correlation is >0.5 for the variables in the regressions, multicollinearity is unlikely to significantly affect the results in Table 4.

ACKNOWLEDGMENTS

I thank Shuping Chen and Shiva Rajgopal, the associate editor and two anonymous reviewers for their insightful comments. I also thank Rick Mergenthaler and Jenny Zhang for assistance in data collection.

REFERENCES

Ali, A., Chen, T., & Radhakrishnan, S. (2005). *Corporate disclosures by family firms*. Working paper. University of Texas at Dallas, TX, USA.

Anders, G., Hymowitz, C., Lublin, J. S., & Clark, D. (2005). All in the family – Why should executive posts at public traded companies be passed on like heirlooms? *The Wall Street Journal* (08/01/2005).

Anderson, R. C., & Reeb, D. M. (2003). Founding-family ownership and firm performance: Evidence from the S&P 500. *Journal of Finance*, *58*(3), 1301–1328.

Anderson, R. C., Reeb, D. M., Mansi, S., & Reeb, D. M. (2003). Founding family ownership and the agency cost of debt. *Journal of Financial Economics*, *68*, 263–285.

Barclay, M., & Holderness, C. G. (1989). Private benefits from control of corporations. *Journal of Financial Economics*, *25*, 371–395.

Beaver, W., Kettler, P., & Scholes, M. (1970). The association between market determined and accounting determined risk measures. *The Accounting Review*, *45*, 654–682.

Bowen, R., Rajgopal, S., & Venkatachalam, M. (2005). *Accounting discretion, corporate governance, and firm performance*. Working Paper. University of Washington, Seattle, WA and Duke University, Durham, NC, USA.

Burgstahler, D., & Dichev, I. (1997). Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics*, *24*, 99–126.

Bushee, B. (1998). Institutional investors, long term investment, and earnings management. *The Accounting Review*, *73*(3), 305–333.

Chen, T., & Radhakrishnan, S. (2005). *Family firms, audit and non-audit fees and earnings management*. Working paper. University of Texas at Dallas, TX, USA.

Claessens, S., Djankov, S., & Lang, L. (2000). The separation of ownership and control in East Asian corporations. *Journal of Financial Economics*, *58*, 81–112.

Colvin, G. (2005). The other victims of Bernie Ebbers's Fraud. *Fortune* (07/25/2005).

Craswell, A. T., Francis, J. R., & Taylor, S. L. (1995). Auditor brand name reputations and industry specializations. *Journal of Accounting and Economics*, *20*, 297–322.

Dechow, P., & Skinner, D. (2000). Earnings management: Reconciling the views of accounting academics, practitioners, and regulators. *Accounting Horizons*, *14*(2), 235–250.

Dechow, P., Sloan, R., & Sweeney, A. (1995). Detecting earnings management. *The Accounting Review*, *70*, 193–225.

Dechow, P., Sloan, R., & Sweeney, A. (1996). Causes and consequences of earnings manipulations: An analysis of firms subject to enforcement actions by the SEC. *Contemporary Accounting Research*, *13*(1), 1–36.

DeFond, M., & Jiambalvo, J. (1994). Debt covenant violation and manipulation of accruals. *Journal of Accounting and Economics*, *17*(1–2), 145–176.

Degeorge, F., Patel, J., & Zeckhauser, R. (1999). Earnings management to exceed thresholds. *Journal of Business*, *72*, 1–33.

Demsetz, H. (1983). The structure of ownership and the theory of the firm. *Journal of Law and Economics*, *25*, 375–390.

Demsetz, H., & Lehn, K. (1985). The structure of corporate ownership: Causes and consequences. *Journal of Political Economy*, *93*, 1155–1177.

Dhaliwal, D., Salamon, G., & Smith, E. (1982). The effect of owner versus management control on the choice of accounting methods. *Journal of Accounting and Economics*, *4*, 41–53.

Dunn, K. A., & Mayhew, B. W. (2004). Audit firm industry specialization and client disclosure quality. *Review of Accounting Studies*, *9*, 35–58.

Durtschi, C., & Easton, P. (2005). Earnings management? The shapes of the frequency distributions of earnings metrics are not evidence Ipso Facto. *Journal of Accounting Research*, *43*(3), 557–592.

Faccio, M., & Lang, L. (2002). The ultimate ownership of Western European corporations. *Journal of Financial Economics*, *65*, 365–395.

Faccio, M., Lang, L., & Young, L. (2001). Dividends and expropriation. *American Economic Review*, *91*, 54–78.

Fama, E., & Jenson, M. (1983). Separation of ownership and control. *Journal of Law and Economics*, *26*, 301–325.

Fama, E., & MacBeth, J. (1973). Risk, return and equilibrium: Empirical tests. *Journal of Political Economy*, *81*, 607–636.

Fan, J., & Wong, T. J. (2002). Corporate ownership structure and the informativeness of accounting earnings in East Asia. *Journal of Accounting and Economics*, *33*, 401–425.

Fields, T., Lys, T., & Vincent, L. (2001). Empirical research on accounting choice. *Journal of Accounting and Economics*, *31*, 255–307.

Francis, J., Schipper, K., & Vincent, L. (2005). Earnings and dividend informativeness when cash flow rights are separated from voting rights. *Journal of Accounting and Economics*, *39*, 329–360.

Frankel, R., McNichols, M., & Wilson, P. G. (1995). Discretionary disclosure and external financing. *The Accounting Review*, *70*(1), 135–150.

Graham, J., Harvey, C., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics*, *40*(1–3), 3–73.

Healy, P., & Wahlen, J. (1999). A review of the earnings management literature and its implications for standard setting. *Accounting Horizons*, *13*(4), 365–383.

Himmelberg, C., Hubbard, G., & Palia, D. (1999). Understanding the determinants of managerial ownership and the link between ownership and performance. *Journal of Financial Economics*, *53*, 353–384.

Holthausen, R., & Verrecchia, R. (1990). The effect of informativeness and consensus on price and volume behavior. *The Accounting Review*, *65*(1), 191–209.

Hunt, A., Moyer, S., & Shevlin, T. (1997). *Earnings volatility, earnings management, and equity value*. Working paper. University of Washington, Seattle, WA, USA.

Jiambalvo, J., Rajgopal, S., & Venkatachalam, M. (2002). Institutional ownership and the extent to which stock prices reflect future earnings. *Contemporary Accounting Research*, *19*(1), 117–145.

Jones, J. (1991). Earnings management during import relief investigations. *Journal of Accounting Research*, *29*(Autumn), 193–228.

Kennedy, P. (1992). *A guide to econometrics* (3rd ed.). Oxford, UK: Basil Blackwell Ltd.

Kothari, S. P., Leone, A. J., & Wasley, C. (2005). Performance matched discretionary accruals measures. *Journal of Accounting and Economics*, *39*(1), 163–197.

La Porta, R., López-de-Silanes, F., & Shleifer, A. (1999). Corporate ownership around the world. *Journal of Finance*, *54*, 471–517.

Leuz, C., Nanda, D., & Wysocki, P. (2003). Investor protection and earnings management: An international comparison. *Journal of Financial Economics*, *69*(3), 505–527.

Lev, B. (1988). Toward a theory of equitable and efficient accounting policy. *The Accounting Review*, *63*(1), 1–22.

Matsumoto, D. (2002). Management's incentives to avoid negative earnings surprises. *The Accounting Review*, *77*(3), 483–514.

Minton, B., & Schrand, C. (1999). The impact of cash flow volatility on discretionary investment and the costs of debt and equity financing. *Journal of Financial Economics*, *54*, 423–460.

Pincus, M., & Rajgopal, S. (2002). The interaction between accrual management and hedging: Evidence from oil and gas firms. *The Accounting Review*, *77*(1), 121–160.

Shivdasani, A. (1993). Board composition, ownership structure, and hostile takeovers. *Journal of Accounting and Economics*, *16*, 167–198.

Shleifer, A., & Vishny, R. (1986). Large shareholders and corporate control. *Journal of Political Economy*, *94*, 461–488.

Skinner, D., & Sloan, R. G. (1999). Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio. *Review of Accounting Studies*, *7*, 287–312.

Stein, J. (1988). Takeover threats and managerial myopia. *Journal of Political Economy*, *96*, 61–80.

Stein, J. (1989). Efficient capital markets, inefficient firms: A model of myopic corporate behavior. *Quarterly Journal of Economics*, *103*, 655–669.

Teoh, S. H., & Wong, T. J. (1993). Perceived audit quality and earnings response coefficient. *The Accounting Review*, *68*, 346–367.

Teoh, S. H., Wong, T. J., Welch, I., & Wong, T. J. (1998a). Earnings management and the long-run market performance of initial public offerings. *Journal of Finance*, *53*(6), 1935–1974.

Teoh, S. H., Wong, T. J., Welch, I., & Wong, T. J. (1998b). Earnings management and the underperformance of seasoned equity offerings. *Journal of Financial Economics*, *50*(1), 63–99.

Villalonga, B., & Amit, R. (2006). How do family ownership, control and management affect firm value? *Journal of Financial Economics*, *80*(2), 385–417.

Warfield, T., Wild, J., & Wild, K. (1995). Managerial ownership, accounting choices, and informativeness of earnings. *Journal of Accounting and Economics*, *20*(1), 61–91.

Watts, R., & Zimmerman, J. (1990). Positive accounting theory: A ten-year perspective. *The Accounting Review*, *65*, 131–156.

Weber, J., Lavelle, L., Lowry, T., Zellner, W., & Barrent, A. (2003). Family, Inc. *Business Week* (Nov. 10), 100–114.

This page intentionally left blank

FINANCIAL REPORTING FACTORS AFFECTING DONATIONS TO CHARITABLE ORGANIZATIONS

John M. Trussel and Linda M. Parsons

ABSTRACT

The purpose of this study is to develop a framework to identify factors in nonprofit financial reports that can impact donations. We posit that there are four reporting factors related to donations. The factors are the efficiency of the organization in allocating resources to its programs, the financial stability of the organization, the information available to donors, and the reputation of the organization. We use factor analysis with variables from previous studies and find that the variables align on four components that appear to represent the factors that we conceptualize.

1. INTRODUCTION

An important function of accounting and financial reporting is to assist in the analysis and evaluation of organizations. In the commercial sector, financial statement users look at accounting ratios, such as measures of profitability and leverage, to make judgments about a firm's performance. In the nonprofit sector, there is not a similar profit motive or an ability to reward equity stakeholders. So what, if any, accounting measures do

Advances in Accounting, Volume 23, 263–285

ISSN: 0882-6110/doi:10.1016/S0882-6110(07)23010-X

financial statement readers find useful for evaluating the operations of nonprofit organizations?

It is possible that the various potential user groups, such as donors, grantors, board members, and beneficiaries, focus on different aspects of operational performance. We opt to examine the perspective of donors, who provide a significant portion of the funding for many nonprofit organizations. The Better Business Bureau, a major watchdog agency for charitable entities, suggests that potential donors obtain financial reports of an organization before making a contribution.[1] Bradley, Jansen, and Silverman (2003) raise concerns about the cost of inefficiencies within the nonprofit sector. After several large fundraising efforts in the aftermath of recent disasters such as the September 11, 2001 terrorist attacks, the Asian tsunami, and Hurricane Katrina, donors are increasingly aware of the need to evaluate the way nonprofit organizations use charitable gifts to address the needs of beneficiaries. We therefore seek to determine which accounting measures provide donors with useful information.

Gordon and Khumawala (1999) suggest that several factors influence an individual's decision to make a charitable contribution. These include preference for the cause to which the entity is dedicated, discretionary income, religious affiliation, and personal belief in altruism. Once a donor has decided to give a certain amount to a particular cause, accounting reports may be used to select a particular organization. Accounting researchers explore whether and how donors use accounting information when making donation decisions. Studies to date incorporate a number of accounting data and financial ratios to demonstrate the value-relevance of nonprofit financial statements to donors. Thus, at least in part, donors' decision-making process possibly includes an evaluation of financial statement information. There is no consensus on which variables are related to donations, and several measures are used as proxies for similar constructs.

The purpose of this paper is to develop a framework to identify financial reporting factors that impact donations and to empirically test this framework by incorporating the various independent variables from previous studies. We find there are four factors in current accounting reports that affect donations. The factors are efficiency, stability, reputation, and the amount of available information. Using factor analysis, we find that the independent variables used in previous studies align on these factors as predicted by our proposed framework. An OLS regression analysis demonstrates that direct donations are significantly related to these four factors. This is the first study to demonstrate the relationship between the conceptual constructs and donations. We synthesize the prior literature that examines different proxies for these conceptual constructs.

By identifying the relationship among financial ratios and other information available in financial reports, the framework developed and tested in this study can provide guidance to researchers studying the usefulness of nonprofit accounting and financial reports. Additionally, it assists donors, grantors, and other financial statement users with evaluation of nonprofit reports. Finally, standard setters, regulators, and watchdog groups can use the framework to better determine the benefit of accounting and financial reports to contributors.

The remainder of the paper is organized as follows. Section 2 summarizes the predictor variables common in prior research. The empirical results are presented in Section 3. Section 4 outlines this study's contribution to the literature.

2. FINANCIAL REPORTING FACTORS RELATED TO DONATIONS

Some prior studies examine whether donors value financial information in the donation decision process. For example, Hyndman (1991, 1990) uses surveys to determine that (a) donors view financial reports of performance as somewhat important in the giving process and (b) charity officials and auditors also perceive that contributors regard financial information as important for making a donation decision. Khumawala and Gordon (1997) simulate the financial statement evaluation process in an experiment with students as potential donors. The potential donors rank financial information, especially the program ratio, as a useful part of the donation process. However, the experiment did not give the potential donors the choice of simply ignoring the financial statement items in their decisions. Finally, Parsons (2007) uses a field experiment to observe the actual responses to a fundraising appeal. She finds that certain individuals are more likely to make a charitable donation when they receive financial information as part of their fundraising request.

Much of the previous research examines the relationship between financial variables and donations.[2] These studies use a number of reporting variables that are hypothesized to impact the charitable giving decisions of donors. These variables, many of which are used by charity watchdog agencies to rate the performance of charitable organizations, operationalize underlying factors of nonprofit accounting reports that may impact charitable donations. However, there is no consistency in the choice of accounting ratios included in prior studies or the description of the underlying factors.

Parsons (2003) proposes that the efficiency of operations and the financial stability of the organization are the most important factors used by donors faced with a giving decision. Weisbrod and Dominguez (1986) demonstrate that the quantity of information available to donors influences donations. Tinkelman (1999) provides evidence that the importance of accounting information is dependent on the organization's reputation, what Weisbrod and Dominguez (1986, p. 87) call a "stock of goodwill". We examine measures from the literature cited by Weisbrod and Dominguez (1986), Tinkelman (1999), and Parsons (2003) and used in a variety of nonprofit accounting studies to determine whether these measures are in fact proxies for more general, underlying factors. Our proposed framework includes efficiency, stability, information, and reputation as constructs. Table 1 summarizes the proxies for each of the constructs, which are defined in the following paragraphs of this section. Although it is possible that other constructs exist, we examine only the four constructs available within a single year's financial statements that are prevalent in previous studies.

2.1. Efficiency

Parsons (2003) defines efficiency as the degree to which nonprofits direct their available resources to the organization's mission. These measures indicate the average portion of each contribution that reaches the organization's beneficiaries. Hyndman (1991) and Khumawala and Gordon (1997) report that donors' principal financial concern is the percentage of expenses dedicated to programs. Watchdog agencies, such as the Better Business Bureau's Wise Giving Alliance and the American Institute of Philanthropy, guide donors to focus on this aspect of nonprofit performance and offer suggestions on minimum acceptable levels. Several proxies for efficiency appear in the nonprofit literature.

2.1.1. Price of Output (PRICE)

Weisbrod and Dominguez (1986) define price as the cost to a donor to purchase one dollar of output for an organization's beneficiaries. Two attributes affect PRICE. First, the donor's cost to provide one dollar of charitable output is less than one dollar when contributions are tax deductible. Second, nonprofits use contributions for purposes other than

Table 1. Variables from Prior Studies.

Construct	Variable (Proxy)	Measurement
Efficiency	Price of output (PRICE)[a]	ln (total expense/Program expense)
	Program expense ratio (PROG)[a]	Program expense/Total expense
	Administrative cost ratio (ADMIN)[b]	Administrative expense/Total expense
Stability	Adequacy of equity (EQUITY)	Net assets/Total revenue
	Revenue concentration (CONCEN)	$\Sigma[(\text{Revenue source})/\text{Total revenues}]^2$
	Operating margin (MARGIN)	(Total revenue – total expense)/Total Revenue
	Administrative cost ratio (ADMIN)[b]	Administrative expense/Total expense
Information	Fundraising expense (FUND)	ln (fundraising expense)
	Fundraising efficiency ratio (FUNDCONT)	Fundraising expense/Total contributions
Reputation	Age of the organization (AGE)	ln (number of years tax exempt)
	Size of the organization (SIZE)	ln (total assets)
	Grant revenues (GRANTS)	ln (government grants + indirect contributions)
	Program revenues (PROGREV)	ln (program revenue)
	Other revenues (OTHREV)	ln (total revenue–direct contributions–indirect contributions–government grants–program revenues)

Note: Following previous studies (e.g., Tinkelman, 1999), PRICE, FUND, AGE, SIZE, GRANTS, PROGREV, and OTHREV are measured in natural log form. GRANTS, PROGREV, and OTHREV are measured in the year donations are received as "major changes in government funding or types of programs would become quickly known to the organization's supporters" (Tinkelman, 1999, p. 159). All other variables are measured in the year prior to the receipt of donations because donors would not have access to current year accounting reports when making a current year contribution decision.

[a]PRICE is calculated as the inverse of PROG, so only one of these measures is included in the factor analysis.

[b]Tuckman and Chang (1991) state that ADMIN measures stability, but Greenlee and Brown (1999) and Frumkin and Kim (2001) use similar measures as a proxy for efficiency. Parsons and Trussel (2008) suggest that ADMIN is an efficiency measure. The factor analysis is used to determine which of these constructs ADMIN measures.

providing charitable services, such as general administrative and overhead costs and fundraising. Ignoring tax deductions, the donor must give more than one dollar to the organization to provide one dollar of output because not all of every dollar is used for programs.

Studies employing PRICE ignore the tax effect of donations (since this does not vary across organizations) and measure PRICE as the inverse of the percentage of expenses dedicated to programs (Posnett & Sandler, 1989; Callen, 1994; Tinkelman, 1998). These studies demonstrate that more efficient nonprofits generate more contributions on average than less efficient organizations.

2.1.2. Program Ratio (PROG)

Baber, Roberts, and Visvanathan (2001) and Roberts, Smith, and Taranto (2006) use the program ratio as an alternative to the PRICE variable. This ratio, defined as the percentage of total expenses spent on programs, is the inverse of the PRICE variable described above. Baber et al. (2001) state that the program ratio can indicate a nonprofit's fundraising strategy. Roberts et al. (2006) use the program ratio to judge the efficiency of nonprofit managers who experience changes in available resources.

2.1.3. Administrative Cost Ratio (ADMIN)

The administrative ratio is administrative expense as a percentage of total expenses. Frumkin and Kim (2001) use this measure of efficiency and find no significant relationship with contributions. However, Greenlee and Brown (1999) use a measure similar to this (excluding fundraising costs from total expenses) to examine the relationship between organizational efficiency and donations. Their findings support studies that use PRICE to measure efficiency with evidence that efficient organizations generate greater contributions.

2.2. Stability

Parsons (2003) defines financial stability as a nonprofit's ability to continue operations if faced with a decrease in resources. In addition to knowing that a nonprofit organization works efficiently, donors want to know whether the organization can continue to operate in the future (analogous to a measure of the ability to continue as a going concern). Anthony (1983) asserts that, just like business entities, nonprofits must maintain positive net equity, with assets in excess of obligations, in order to operate. Mautz (1988) and Pallot (1990) contend that donors are interested in nonprofits' future cash commitments and the ability to fulfill their obligations. The Better Business Bureau's Wise Giving Alliance, the American Institute of Philanthropy, and Charity Navigator advise donors to examine nonprofit reserves and operating margins and offer guidelines of acceptable performance.

Tuckman and Chang (1991) recommend four measures to assess an organization's financial condition. Greenlee and Trussel (2000) and Trussel and Greenlee (2004) find that Tuckman and Chang's stability measures are useful for predicting the financial vulnerability of a charitable organization. Parsons and Trussel (2008) find that certain stability measures are positively linked to total donations. Tuckman and Chang's (1991) stability measures are defined below.

2.2.1. Adequacy of Equity (EQUITY)
First, the ratio of net assets to total revenue can be calculated to determine the adequacy of "equity". This ratio provides a measure of the number of periods of revenue a nonprofit currently has on hand. In the event of a temporary decline in revenues, a firm with greater access to funds faces a lower risk of collapse. An organization with a larger measure of net assets to total revenue is more likely to be able to (a) liquidate existing assets or (b) obtain credit in order to meet future needs. Without an adequate reserve of funds, a nonprofit firm will be unable to continue to operate normally when faced with a reduction in revenues.

Trussel and Greenlee (2004) find the adequacy of equity measure is a positive and significant predictor of financial stability. However, Parsons and Trussel (2008) and Marudas (2004) observe that donations are negatively related to EQUITY, implying that donors punish organizations that do not spend donations on programs.

2.2.2. Revenue Concentration (CONCEN)
Second, a firm with a greater number of revenue sources is expected to be less susceptible to financial shocks. A firm that is dependent on one or a few revenue providers is vulnerable to declines in the economic health or changes in the donation preferences of those providers. To capture the extent of revenue dispersion, Tuckman and Chang recommend computing an index of revenue concentration similar to the Herfindahl Index used by economists to measure market concentration. Specifically, Tuckman and Chang define the revenue concentration index as the summation of the squared percentage share that each revenue source represents of total revenue. If a single source of revenue exists, the index equals one. A firm with many sources of revenue has an index closer to zero. Greenlee and Trussel (2000) demonstrate that revenue concentration is a significant predictor of financial vulnerability. Parsons and Trussel (2008) show that nonprofits with greater financial stability (lower revenue concentration) generate more contributions on average.

2.2.3. Operating Margin (MARGIN)

Third, Tuckman and Chang suggest a measure analogous to the gross margin ratio used in a business setting. This ratio, called operating margin, is revenues less expenditures, divided by revenues. A higher operating margin is indicative of a greater potential surplus on which to draw in the event of unexpected financial difficulties. Greenlee and Trussel (2000) show that nonprofits with higher operating margins are less susceptible to financial vulnerability. Parsons and Trussel (2008) find that more financially stable nonprofits, defined as those with a higher operating margin, raise more donations than those with a lower operating margin.

2.2.4. Administrative Costs Ratio (ADMIN)

The last stability measure recommended by Tuckman and Chang is the administrative ratio used as an efficiency measure in some research studies. Tuckman and Chang reason that a firm with high administrative expenses could adjust to revenue reductions by taking steps to cut costs. When faced with a reduction in revenues, an organization with larger overhead costs has the option to cut those costs instead of reducing the overall level of program services offered. By contrast, a leaner, more efficient firm may have less ability to economize without cutting expenditures. According to Tuckman and Chang, nonprofits with lower ratios are the most vulnerable to financial crisis.

Greenlee and Trussel (2000) illustrate that more stable firms (those with high administrative ratios) are less susceptible to financial vulnerability. However, when Greenlee and Brown (1999) use a similar measure, they find that donors prefer (and donate more to) nonprofits with lower administrative ratios. Parsons and Trussel (2008) show that administrative ratio and PRICE are highly correlated and negatively related to donations. They suggest that the administrative ratio is an efficiency measure (instead of a stability measure). Therefore, we suspect the administrative ratio is a measure of efficiency, as suggested by Greenlee and Brown (1999) instead of a stability measure, as posited by Tuckman and Chang (1991).

2.3. Information Available

In order to generate charitable contributions, nonprofits must alert potential donors to the mission of the organization and the plight of its beneficiaries. Fundraising efforts work similar to advertising by providing potential donors with information about organizations and their operations.

Hansmann (1980) states that donors rely on information from nonprofits for assurance that donations are used appropriately (i.e., on the mission). Gordon, Greenlee, and Nitterhouse (1999) claim the widespread availability of accounting reports impacts donors (and thereby total contributions) and other financial statement users. Each of these papers implies that donors are more likely to make contributions to nonprofits when they have adequate information for doing so. The quantity of information donors receive is difficult to measure directly, but two proxies for the amount of information have been suggested in the previous literature.

2.3.1. Fundraising Expense (FUND)

Weisbrod and Dominguez (1986) claim that advertising by for-profit firms is one method for transferring information to potential customers. They argue that fundraising efforts by nonprofits serves the same purpose as advertising. Tinkelman (1999, p. 137) agrees that "fundraising expenses are used as a proxy for the information available to donors". Both studies find that total fundraising expense is positively related to total contributions. Frumkin and Kim (2001) find that fundraising expenditures positively impact contributions in every nonprofit sector.

2.3.2. Fundraising Efficiency Ratio (FUNDCONT)

The fundraising efficiency ratio is fundraising expense as a percentage of direct contributions. This ratio provides an indication of the cost of generating current contributions, thus addressing the efficiency and effectiveness of fundraising instead of the efficiency or effectiveness of operations. Watchdog agencies, such as the Better Business Bureau's Wise Giving Alliance, focus on this ratio when setting performance guidelines for nonprofits. To date, accounting studies have not examined the relationship of this particular fundraising ratio with donations.

2.4. Reputation

Donors are more likely to make contributions to organizations that provide the best service. However, often the donors cannot directly view a nonprofit organization's output (Gordon & Khumawala, 1999) and make a judgment about its quality. Therefore, donors must rely, in part, on organizational reputation to assess output. The following proxies for organizational reputation are included in our analysis.

2.4.1. Organization Age (AGE)

Bennett and DiLorenzo (1994) state that relatively new organizations need time to establish themselves with donors and achieve name recognition. In order to survive in the long term, nonprofits need to produce quality output and succeed in fundraising. AGE may proxy for an organization's ability to establish what Weisbrod and Dominguez (1986) and Posnett and Sandler (1989) call a "stock of goodwill". The findings related to the AGE coefficient are mixed (Weisbrod & Dominguez, 1986; Posnett & Sandler, 1989; Tinkelman, 1999; Parsons & Trussel, 2008). We include AGE in order to determine if it is, in fact, a measure of reputation for quality. Following prior studies, AGE is measured as the natural log of the number of years since the nonprofit received its tax-exempt status.

2.4.2. Organizational Size (SIZE)

Organizational growth can only be achieved when a nonprofit entity is able to continue to generate revenues over a number of years. Size may represent a nonprofit organization's ability to succeed in its mission and attract revenues, including contributions. Tinkelman (1999) uses size as a proxy of reputation. Following prior studies, size is measured as the natural log of total assets.

2.4.3. Government Grants and Indirect Donations (GRANTS)

Government grantors and indirect donors (such as United Way) serve as monitors to nonprofit organizations, subjecting recipient organizations to increased reporting and auditing requirements. Though donors may not have access to the reports nonprofits submit to state and federal government grantors, they may perceive the quality of nonprofit output is improved due to the government oversight. Small donors may look to expert donors (like government grantors) to evaluate a nonprofit's performance the way novice investors look at trading by institutional and other knowledgeable investors to value for-profit firms. Tinkelman (1999) proposes that indirect donations may impact funding from direct donors. Expanding the measure used by Tinkelman (1999), GRANTS are measured as the natural log of total government grants plus indirect donations.

2.4.4. Program (PROGREV) and Other Revenues (OTHREV)

Charities may generate revenues through charging fees for services, which are classified as program revenues. For example, colleges charge tuition for educational services and hospitals charge for healthcare services. The output from these business-like services requires market (product) discipline and the quality of the output can be determined by the recipients. PROGREV is measured as the natural log of program revenues. Other revenues (besides donations, grants, and program revenues) include membership dues, asset sales, rental income, and investment income. Like program revenues these are business-like, which require market discipline and thus send quality signals. OTHREV is the natural log of other revenues.

Posnett and Sandler (1989), Callen (1994), and Tinkelman (1998, 1999) include other revenue sources, such as investment income, as control variables in their extensions to the Weisbrod and Dominguez (1986) model. Though the results from these studies are mixed as to whether and how alternative funding sources are related to direct contributions, we expect a positive relationship, since more other revenues imply an enhanced reputation.

3. EMPIRICAL TESTS

The study uses data from the Internal Revenue Service Statistics of Income (SOI)[3] database developed by the National Center for Charitable Statistics (NCCS) for the tax years 1998 and 1997. This database includes information for all 501(c)3 charitable organizations with at least \$10 million in assets, and a random sample of approximately 4,000 smaller charitable organizations that are required to file a Form 990 with the IRS. Organizations that are not required to file IRS Form 990, such as religious organizations or those with gross receipts less than \$25,000, are not included.

Our initial sample comprised 13,058 organizations that provided financial information for both 1998 and 1997. Organizations with missing data and outliers, defined as those charities with any financial indicator in the extreme top and bottom one-percentile for that indicator, were eliminated from our sample. Outliers are eliminated because they may represent data input errors.[4]

Since our focus is on the value-relevance of financial reporting factors to donors, we limit our investigation to nonprofits that rely on donations as a material source of income.[5] We define donations as material if they represent 10% or more of total revenues.[6] The final sample included 4,727 organizations. Table 2, Panel A shows how the final sample was derived. The sectors reported in Table 2, Panel B are based upon the classification system developed by the NCCS and called the National Taxonomy of Exempt Entities (NTEE). The NTEE classifies nonprofit organizations into 10 major sectors. Five of these sectors combined have fewer than 10% of all

Table 2. Composition of Sample.

Panel A: Sample selection

	Number	Percent
Total organizations	13,058	100.0
Missing data	3,706	28.4
Outliers[a]	297	2.3
Organizations without significant donations[b]	4,328	33.1
Final sample	4,727	36.2

Panel B: Sample by sector

Sector	Number	Percent
Arts	570	12.1
Education	1,411	29.8
Human services	963	20.4
Public benefit	672	14.2
Health	648	13.7
Other	463	9.8
Total	4,727	100.0

[a]Outliers are defined as those charities with any variable (from Table 1) in the extreme one-percentile.
[b]Defined as those charities receiving less than 10% of total revenues from direct contributions.

nonprofit organizations; thus, we combined these sectors into one "other" category.

Table 3 summarizes the univariate statistics for the charities in our sample. The mean of total assets is \$10.3 million and the average age is 25 years. Organizations dedicate, on average, 79% of total spending to programs and spend 15% on administrative costs. Fundraising expenses average 10% of direct contributions.

3.1. Factor Analysis

We hypothesize there are four factors that can impact direct donations to charities. We extract the factors from the variables in Table 1 using the maximum likelihood method, since the tests are based in theory. PRICE is calculated as the inverse of PROG and the anti-image correlation coefficients for these variables are very high (i.e., greater than 90%). Thus, we included only one of these variables (PRICE) in the factor analysis.[7]

Table 3. Descriptive Statistics.

	Mean	Standard Deviation
PRICE	0.26	0.23
PROG	0.79	0.15
ADMIN	0.15	0.13
EQUITY	3.31	8.31
CONCEN	0.53	0.19
MARGIN	0.27	0.29
FUND	8.40	5.83
FUNDCONT	0.10	0.14
AGE	3.22	0.82
SIZE	16.15	2.22
GRANTS	4.91	6.48
PROGREV	8.76	7.05
OTHREV	12.59	4.03

Note: The independent variables are defined as PRICE, ln (total expenses as a percentage of program expenses); PROG, Program expenses as percentage of total expenses; ADMIN, Administrative expenses as percentage of total expenses; EQUITY, Net assets $\div$ total revenues; CONCEN, Σ [(revenue source) $\div$ total revenues]^2; MARGIN, (Total revenues – total expenses) $\div$ total revenues; FUND, ln (fundraising expenses); FUNDCONT, Fundraising expenses as percentage of direct contributions; AGE, ln (number of years since granted tax-exempt status); SIZE, ln (total assets); GRANTS, ln (government grants + indirect contributions); PROGREV, ln (program revenues); OTHREV, ln (revenues from other sources).

We employed the Varimax rotation method with Kaiser normalization. The results, displayed in Table 4, are primarily as anticipated.

The reported results support our prediction that there are four factors represented by the variables commonly used in the prior nonprofit literature. The variables included in this study, along with the factor loadings for each, are presented in Table 4. The highest factor loadings for each construct are shown in bold.

The PRICE and ADMIN variables are highly correlated with the first factor, which we label efficiency. The variables representing MARGIN and EQUITY are highly correlated with the second factor, which we call stability. These results support the findings in Parsons and Trussel (2008) that ADMIN is an indicator of efficiency, not stability as Tuckman and Chang (1991) suggest.

Measures for FUND and FUNDCONT have the highest correlation with the third factor, which we attribute to the availability of information. The variables representing AGE, GRANTS, PROGREV, and OTHREV variables load with the fourth factor, representing reputation, as predicted. The SIZE variable has the highest correlation with the hypothesized factor,

Table 4. Results of Factor Analysis.

	Efficiency	Stability	Information	Reputation
PRICE	**0.955**	0.102	0.269	
ADMIN	**0.898**		−0.135	
EQUITY		**0.188**		
CONCEN	−0.117			**−0.517**
MARGIN		**0.634**		
FUND			**0.669**	0.372
FUNDCONT	0.102	−0.233	**0.710**	
AGE			0.136	**0.512**
SIZE	−0.104	0.462	0.211	**0.708**
GRANTS		−0.230		**0.442**
PROGREV		−0.326		**0.646**
OTHREV		0.286	0.171	**0.608**

Note:

1. This table represents the result of factor analysis with Maximum Likelihood as the extraction method and Varimax with Kaiser Normalization as the rotation method. The factors with the highest correlation with the variables are noted in bold. Loadings less than 0.10 are not shown for ease of analysis.
2. The independent variables are defined as PRICE, ln (total expenses as a percentage of program expenses); ADMIN, Administrative expenses as percentage of total expenses; EQUITY, Net assets ÷ total revenues; CONCEN, Σ[(revenue source$_j$) ÷ total revenues]^2; MARGIN, (Total revenues–total expenses) ÷ total revenues; FUND, ln (fundraising expenses); FUNDCONT, Fundraising expenses as percentage of direct contributions; AGE, ln (number of years since granted tax-exempt status); SIZE, ln (total assets); GRANTS, ln (government grants + indirect contributions); PROGREV, ln (program revenues); OTHREV, ln (revenues from other sources).

reputation; however, this variable also has a relatively high correlation (0.462) with the stability factor. SIZE seems to be primarily a proxy for reputation, but to a lesser extent, it proxies stability. This finding is in line with Trussel and Greenlee (2004), who find that SIZE is a significant control variable in their financial stability model. CONCEN also loads with the reputation factor. We anticipated that CONCEN would be a proxy for stability. The negative sign on CONCEN suggests that a higher concentration of revenues is a signal of lower quality or reputation. Perhaps if revenue is too concentrated in a few sources, then the reputation is diminished.

3.2. Sensitivity Analysis

The above factor analysis is based upon the entire sample of nonprofit organizations. In this section, we perform various analyses to test the sensitivity of our results to alternative specifications. First, there may be differences in results due to the various sectors in which an organization operates. Thus, we explore the factor analysis by the six major sectors, as defined in Table 2, Panel B. The factors with the highest correlation with the variables are presented in Table 5, Panel A. Those variables that load on a factor other than the one hypothesized are noted in bold. In all sectors, the PRICE and ADMIN variables have the highest factor loadings with the efficiency factor, as hypothesized. Also as anticipated, FUND and FUNDCONT, have the highest correlation with the information factor. All of the exceptions to the hypothesized factor loadings are related to the variables in the stability and reputation factors. Similar to the entire sample of organizations, the CONCEN variable has the highest factor loadings with the reputation construct in all sectors except the human services and the public benefit sectors. In the case of the public benefit sector, EQUITY loads highest with the reputation factor. Also, in all sectors except the arts, one or more of the reputation variables load highest with the stability factor. With the exception of AGE and MARGIN, all of the variables that are associated with the stability and reputation factors tend to tradeoff in the various sectors. In most cases (not reported), these variables also have relatively high correlations with the hypothesized factor. For example, in the education sector, the correlation between PROGREV and the stability factor is 0.61. However, the correlation between this variable and its hypothesized factor, reputation, is 0.55.

Second, there may also be differences in the results based on the relative amount of direct contributions that an organization receives, since many nonprofits generate alternative forms of revenues. We divide the entire sample into thirds by ranking according to the percent of total revenues derived from direct contributions. The cutoff of direct contributions as a percent of total revenues for the first group is less than 24% and for the third group is greater than 57%. The second group is between these two percentages. The results of the factor analyses on these three groups appear in Table 5, Panel B. In all three groups, CONCEN has the highest correlation with the reputation factor. Also, in the middle third of contributions, GRANTS and PROGREV load highest on the stability factor. These two results are similar to the previous results in which there is sometimes the variables have high correlations with both the stability and reputation factors. There is one unique result in this testing. The FUND variable has the highest correlation with the reputation factor in the lowest third of contributions. When an organization receives less than 24% of its revenues from contributions, then fundraising expenditures tend to bolster

Table 5. Results of Factor Analysis: Robustness Tests.

Panel A: Sectors

Hypothesized Factor	Variable	Arts	Education	Human Services	Public Benefit	Health	Other
Efficiency (EFF)	PRICE	EFF	EFF	EFF	EFF	EFF	EFF
	ADMIN	EFF	EFF	EFF	EFF	EFF	EFF
Stability (STAB)	EQUITY	STAB	STAB	STAB	**REP**	STAB	STAB
	CONCEN	**REP**	**REP**	STAB	STAB	**REP**	**REP**
	MARGIN	STAB	STAB	STAB	STAB	STAB	STAB
Information (INFO)	FUND	INFO	INFO	INFO	INFO	INFO	INFO
	FUNDCONT	INFO	INFO	INFO	INFO	INFO	INFO
Reputation (REP)	AGE	REP	REP	REP	REP	REP	REP
	SIZE	REP	REP	REP	**STAB**	REP	**STAB**
	GRANTS	REP	REP	**STAB**	REP	**STAB**	REP
	PROGREV	REP	**STAB**	REP	REP	**STAB**	STAB
	OTHREV	REP	REP	REP	**STAB**	REP	REP

Panel B: Direct contributions as a percent of total revenues

Hypothesized Factor	Variable	Lower Third of Contributions	Middle Third of Contributions	Upper Third of Contributions
Efficiency (EFF)	PRICE	EFF	EFF	EFF
	ADMIN	EFF	EFF	EFF
Stability (STAB)	EQUITY	STAB	STAB	STAB
	CONCEN	**REP**	**REP**	**REP**
	MARGIN	STAB	STAB	STAB
Information (INFO)	FUND	**REP**	INFO	INFO
	FUNDCONT	INFO	INFO	INFO
Reputation (REP)	AGE	REP	REP	REP
	SIZE	REP	REP	REP
	GRANTS	REP	**STAB**	REP
	PROGREV	REP	**STAB**	REP
	OTHREV	REP	REP	REP

Note:

1. This table represents the results of factor analysis with Maximum Likelihood as the extraction method and Varimax with Kaiser Normalization as the rotation method. The factors with the highest correlation with the variables are presented. Those variables that do not load with the hypothesized factors are noted in bold. In panel A, the organizations are segregated by the six major sectors. In panel B, the organizations are classified into thirds according to their contributions as a percentage of revenues.
2. The independent variables are defined in Table 4.

the reputation of the organization, rather than acting as information about the organization.

Third, the coefficient associated with EQUITY may not be monotonic in relation to donations. Donors might be concerned when reserves are too large (indicating that a nonprofit is not distributing donated funds to the intended beneficiaries) or too small (evidence that the organization is financially vulnerable to changes in revenues) (Parsons & Trussel, 2008). We have no ex ante definition of high or low reserves. Therefore, following Parsons and Trussel (2008), we use alternative definitions. In the first iteration, we divide the sample population of organization-years into thirds (based on the EQUITY measure). In the second iteration, we use EQUITY less than zero and more than three as the cutoffs for low and high reserves, respectively, following the guidelines of the American Institute of Philanthropy. For both iterations, we estimate the factor analyses separately for each category of EQUITY – low, medium, and high. For these two iterations, the results (not shown) are similar for EQUITY in all three sample groups. Our previous results are robust to these specifications of the model.

Fourth, we reconsider the FUND and FUNDCONT variables. Hager (2003) and Krishnan, Yetman, and Yetman (2006) provide evidence that some organizations that receive contributions incur fundraising costs but report zero fundraising expense, distorting the reported ratios. To determine whether our results are impacted by organizations that underreport fundraising expense, we repeat our analyses after excluding organizations with zero fundraising expense. The results (not shown) are similar to those when these organizations are included in the sample.

Finally, FUNDCONT and FUND may have nonlinear relationships with donations. Although we hypothesize that higher value of these variables are proxies for more information, very high values may be counter to the mission of the organization. For example, the Better Business Bureau's Wise Giving Alliance recommends that FUNDCONT be no more than 35%. The Wise Giving Alliance claims that ratios in excess of the recommended amount may represent inefficiencies. We test this possibility using the same methodology as we did for the EQUITY variable above, since we again have no ex ante definition of high or low ratios. We first divide the sample population of organization-years into thirds (based on the FUNDCONT measure). In the second iteration, we use FUNDCONT equal to zero and more than 35% as the cutoffs for low and high amounts, respectively, following the guidelines of the Wise Giving Alliance. For both iterations, we estimate the factor analyses separately for each category of

FUNDCONT – low, medium, and high. For these two iterations, the results (not shown) are similar for FUNDCONT in all three sample groups. We also test FUND in a similar manner with similar results. Our previous results are robust to these specifications of the model.

3.3. Regression Analysis

The variables we examine in the factor analysis appear to represent four distinct factors of operational performance. Next, we include the factors identified in our proposed framework in an OLS regression model to determine if these constructs are determinants of donations.[8] Our regression model, which uses the factor scores obtained from the factor analysis, follows.

$$\begin{aligned}\mathrm{DON}_{it} = {} & \beta_0 + \beta_1 \mathrm{EFFICIENCY}_{i(t-1)} + \beta_2 \mathrm{STABILITY}_{i(t-1)} \\ & + \beta_3 \mathrm{INFORMATION} + \beta_4 \mathrm{REPUTATION}_{i(t-1)} + \varepsilon_i\end{aligned}$$

The dependent variable is the natural log of direct donations in 1998. The independent variables are measured as the factor scores obtained from the factor analysis summarized in Table 4. The factor score is a composite measure of the underlying construct that is extracted from the observed variables (including all variables, not just those with the highest loadings).

Since we focus on the impact of the factors on donations, we test the model only on the charities with a material level of contributions relative to total revenues (at least 10%).[9] The results are presented in Table 6. The overall model is significant at the 0.001 level and the adjusted R^2 indicates that the model explains over 57% of the variations in donations in our sample. The results indicate the financial reporting factors suggested by our framework explain a significant amount of the variation in donations.

Each of the independent variables is significant at the 0.001 level, indicating that donations are related to each of the constructs suggested by our proposed framework. The negative sign on the efficiency variable is expected, since a higher price or more administrative costs are predicted to be inversely related to contributions. The other factors show positive signs, as predicted. Though several of the individual variables have been shown to be related to donations, this study is the first to demonstrate the relationship of the constructs to charitable contributions.

Table 6. Results from OLS Regression Analysis.

$$\mathrm{DON}_{it} = \beta_0 + \beta_1\mathrm{EFFICIENCY}_{i(t-1)} + \beta_2\mathrm{STABILITY}_{i(t-1)} + \beta_3\mathrm{INFORMATION}_{i(t-1)} + \beta_4\mathrm{REPUTATION}_{i(t-1)} + \varepsilon_i$$

	Coefficient	*t*-Statistic	*p*-Value
Intercept	14.106	708.93	0.000
EFFICIENCY	−0.355	−17.28	0.000
STABILITY	0.855	35.58	0.000
INFORMATION	0.730	31.64	0.000
REPUTATION	1.253	54.71	0.000
Model			
Adjusted R^2			0.571
F-statistic			1,489.789
p-value			0.000

Note: DON = ln (direct donations) in 1998 and the independent variables are the factor scores from the factor analysis measured in 1997.

4. SUMMARY AND CONCLUSIONS

Parsons (2003) identifies conceptual factors that have been consistently used in previous nonprofit accounting studies. However, the factors she discusses have been operationalized using a variety of measures in prior research. This study establishes a framework that identifies the constructs represented by the variables used in previous accounting studies. Our framework demonstrates that accounting measures and other information from nonprofit financial reports can be categorized into four major constructs, each of which is an important determinant of donations.

Based on the variety of measures commonly used in prior studies, we hypothesize that there are four factors that affect donations – efficiency, stability, information, and reputation. We measure 12 independent variables used or suggested in previous studies to model donations on a sample of 4,727 nonprofit organizations. Using factor analysis, we find support for our hypothesis that the 12 variables load on the four conceptual factor scores. Using OLS regression with these four factors as predictor variables, we find that donations are a function of efficiency, financial stability, the amount of information provided by the organization and the reputation of the organization.

By identifying the relationship among financial ratios and other information available in financial reports, the framework developed and tested in this study can provide guidance to researchers studying the usefulness of nonprofit accounting and financial reports. Additionally, it assists donors, grantors, and other financial statement users with evaluation of nonprofit reports. Finally, standard setters, regulators, and watchdog groups can use the framework to better determine the benefit of accounting and financial reports to contributors.

NOTES

1. See www.give.org for the Better Business Bureau's Wise Giving Alliance and its "Tips on Giving".

2. Though these studies do not claim that all donors, especially those making small donations, directly evaluate financial information, they find a positive relationship between certain financial ratios and contributions. It is possible that even if donors do not rely directly on financial reports, they may use recommendations from watchdog agencies or large donors to judge organizational performance.

3. See discussions of the IRS 990 data and the NCCS database in Gordon et al. (1999) and Froelich and Knoepfle (1996).

4. The National Center for Charitable Statistics (NCCS), which compiled the database, suggests that outliers are likely to represent either (1) errors in the data input process, (2) errors in the 990 forms, or (3) influential organizations that can mask financial trends demonstrated in other organizations. Therefore, the NCSS suggests excluding outliers when analyzing aggregated data. See a discussion of the database at http://nccs.urban.org/

5. Tinkelman stated that the relevance of certain ratios are "uninformative for organizations that depend primarily on program fees or other revenues" (1999, p. 139). Gordon and Khumawala (1999) note that financial statements are more relevant for organizations that serve as a broker between donors and beneficiaries (e.g., homeless shelter) than for organizations that provide services for a fee (e.g., colleges and universities).

6. The results are robust for materiality levels of 5–20%.

7. The results are similar to those reported when we include PROG rather than PRICE.

8. One could argue that organizations reporting zero fundraising costs are not actively seeking donations and should be excluded from testing. Excluding those organizations does not affect the tenor of our results.

9. The results are robust for materiality levels of 5–20%.

ACKNOWLEDGMENTS

We thank Jeff Callen, Rita Cheng, Karen Kitching, Dan Tinkelman, the associate editor, two anonymous reviewers, and workshop participants at the 2004 annual meeting of the Association for Research on Nonprofit Organizations and Voluntary Action (ARNOVA) and the 2004 AAA Annual and Mid-Atlantic regional meetings for their helpful comments on earlier drafts of this paper.

REFERENCES

Anthony, R. N. (1983). *Tell it like it was: A conceptual framework for financial accounting*. Homewood, IL: Richard D. Irwin, Inc.

Baber, W. R., Roberts, A. A., & Visvanathan, G. (2001). Charitable organizations' strategies and program-spending ratios. *Accounting Horizons*, *15*(4), 329–343.

Bennett, J. T., & DiLorenzo, T. J. (1994). *Unhealthy charities – hazardous to your health and wealth*. New York: Basic Books.

Bradley, B., Jansen, P., & Silverman, L. (2003). The nonprofit sector's $100 Billion opportunity. *Harvard Business Review*, *81*(5), 94–103.

Callen, J. L. (1994). Money donations, volunteering and organizational efficiency. *The Journal of Productivity Analysis*, *5*, 215–228.

Froelich, K. A., & Knoepfle, T. W. (1996). Internal revenue service 990 data: Fact or fiction? *Nonprofit and Voluntary Sector Quarterly*, *25*(1), 40–52.

Frumkin, P., & Kim, M. T. (2001). Strategic positioning and the financing of nonprofit organizations: Is efficiency rewarded in the contributions marketplace? *Public Administration Review*, *61*(3), 266–275.

Gordon, T. P., Greenlee, J. S., & Nitterhouse, D. (1999). Tax-exempt organization financial data: Availability and limitations. *Accounting Horizons*, *13*(2), 113–128.

Gordon, T. P., & Khumawala, S. B. (1999). The demand for nonprofit financial statements: A model for individual giving. *Journal of Accounting Literature*, *18*, 31–56.

Greenlee, J. S., & Brown, K. L. (1999). The impact of accounting information on contributions to charitable organizations. *Research in Accounting Regulation*, *13*, 111–125.

Greenlee, J. S., & Trussel, J. M. (2000). Predicting the financial vulnerability of charitable organizations. *Nonprofit Management and Leadership*, *11*(2), 199–210.

Hager, M. A. (2003). Current practices in allocation of fundraising expenditures. *New Directions in Philanthropic Fundraising*, *41*, 39–52.

Hansmann, H. B. (1980). The role of nonprofit enterprise. *The Yale Law Journal*, *89*(5), 835–901.

Hyndman, N. (1990). Charity accounting – an empirical study of the information needs of contributors to UK fund raising charities. *Financial Accountability and Management*, *6*(4), 295–307.

Hyndman, N. (1991). Contributions to charities – a comparison of their information needs and the perceptions of such by the providers of information. *Financial Accountability and Management*, *7*(2), 69–82.

Khumawala, S. B., & Gordon, T. P. (1997). Bridging the credibility of GAAP: Individual donors and the new accounting standards for nonprofit organizations. *Accounting Horizons*, *11*(3), 45–68.

Krishnan, R., Yetman, M. H., & Yetman, R. J. (2006). Expense misreporting in nonprofit organizations. *Accounting Review*, *81*(2), 399–420.

Marudas, N. P. (2004). Effects of nonprofit organization wealth and efficiency on private donations to large nonprofit organizations. *Research in Governmental and Nonprofit Accounting*, *11*, 71–91.

Mautz, R. K. (1988). Monuments, mistakes, and opportunities. *Accounting Horizons*, *2*(2), 123–128.

Pallot, J. (1990). The nature of public assets: A response to Mautz. *Accounting Horizons*, *4*(2), 79–85.

Parsons, L. M. (2003). Is accounting information from nonprofit organizations useful to donors? A review of charitable giving and value-relevance. *Journal of Accounting Literature*, *22*, 104–129.

Parsons, L. M. (2007). The impact of financial information and voluntary disclosures on contributions to not-for-profit organizations. *Behavioral Research in Accounting*, *19*, 179–196.

Parsons, L. M., & Trussel, J. M. (2008). Fundamental analysis of not-for-profit financial statements: An examination of financial vulnerability measures. *Research in Governmental and Nonprofit Accounting*, *12*(forthcoming).

Posnett, J., & Sandler, T. (1989). Demand for charity donations in private non-profit markets: The case of the U.K. *Journal of Public Economics*, *40*, 187–200.

Roberts, A. A., Smith, P., & Taranto, K. (2006). *Evaluating charitable spending: Marginal versus average performance measures*. Working paper. Boston College, University of Texas at San Antonio and Georgetown University.

Tinkelman, D. (1998). Differences in sensitivity of financial statement users to joint cost allocations: The case of nonprofit organizations. *Journal of Accounting, Auditing and Finance*, *13*(4), 377–393.

Tinkelman, D. (1999). Factors affecting the relation between donations to nonprofit organizations and an efficiency ratio. *Research in Government and Nonprofit Accounting*, *10*, 135–161.

Trussel, J. M., & Greenlee, J. S. (2004). A financial rating system for charitable nonprofit organizations. *Research in Government and Nonprofit Accounting*, *11*, 93–116.

Tuckman, H. P., & Chang, C. F. (1991). A methodology for measuring the financial vulnerability of charitable nonprofit organizations. *Nonprofit and Voluntary Sector Quarterly*, *20*(4), 445–460.

Weisbrod, B. A., & Dominguez, N. D. (1986). Demand for collective goods in private nonprofit markets: Can fundraising expenditures help overcome free-rider behavior? *Journal of Public Economics*, *30*, 83–96.

This page intentionally left blank

THE VALUE-RELEVANCE OF NONFINANCIAL INFORMATION: THE BIOTECHNOLOGY INDUSTRY

Ya-wen Yang

ABSTRACT

This study examines whether nonfinancial patent information is useful to investors in assessing and valuing biotech firm's long-term financial performance. Using six patent variables measuring both quantity and quality aspects of patents, I find that patent information is associated with biotech firm's subsequent financial performance. In addition, I derive a return model to test whether patent information adds incremental value-relevance over traditional accounting measures to the market valuation in the biotech industry. The result suggests that nonfinancial patent information captures the biotech firms' value not valued by traditional accounting measures.

INTRODUCTION

This study examines whether nonfinancial information, particularly patent information, is useful to investors in assessing and valuing biotech firms'

Advances in Accounting, Volume 23, 287–314

ISSN: 0882-6110/doi:10.1016/S0882-6110(07)23011-1

long-term financial performance. The emphasis on nonfinancial patent information is motivated by the current FASB study that highlights the importance of measurement and recognition of internally developed intangible assets in financial statements (FASB, 2001). Current accounting practice requires companies immediately expense their significant value enhancing investments in internally developed intangible assets, such as research and development (R&D), for financial statement purposes. As a result, accounting assets do not fully reflect a company's valuable intangible assets, and financial variables, such as earnings and book values of equity, are often negative and appear to be unrelated to market values.

The periodic R&D expenditure is the only innovation-relevant financial information required to be disclosed in corporate financial reports. As full expensing of R&D cost fails to sufficiently inform market participants about a firm's R&D activities and potential future earnings power, several prior studies examine the relationship between intangible asset proxies and firm performance (Shortridge, 2004; Rajgopal, Venkatachalam, & Kotha, 2003; Ely, Simko, & Thomas, 2003; Trueman, Wang, & Zhang, 2001 among others). These studies suggest that financial and nonfinancial information are complimentary and both should be included when assessing financial and market performance. This study extends this literature and focuses on the value-relevance of nonfinancial patent information in the biotech industry.

The biotech industry is characterized by short-cycle technological developments requiring large investments with very uncertain payoffs. R&D productivity (i.e., the quality and quantity of inventive output) is a vital determinant of long-term success in high-tech companies, particularly for the biotech industry because, as a percentage of revenues, R&D spending in the biotech industry is among the highest of any U.S. industry group.[1] Economic literature usually considers patent information as an indicator of inventive output, which measures the productivity of R&D spending (i.e., an indicator of inventive input) (Hall, Jaffe, & Trajtenberg, 2000; Harhoff, Narin, Scherer, & Vopel, 1999; Griliches, 1990). A patent protects a new process or product from competition and allows a firm to recoup R&D costs while earning a good return on its investment. Because patents are among the most important benchmarks of progress in developing new biotechnology products, this study uses publicly available patent information to determine whether market participants use this nonfinancial information in assessing future cash flows.

Although patents create a significant competitive advantage in biotech companies, patent information may not appear value relevant in empirical work given the measurement difficulties in these variables. As Grilliches (1990) points out, patents differ greatly in their technical and economic significance. Many of them reflect minor improvements of little economic value, while some of them prove extremely valuable. To overcome the measurement problem, this study incorporates patent variables in prior research (Deng, Lev, & Narin, 1999; Hirschey, Richardson, & Scholz, 2001), including the number of patents granted, the number of citations subsequently made to those patents, the age of patents cited, and number of references cited. It also adds previously excluded variables such as number of claims (CLAIM) and the extent to which patents are DNA type (DNA%). These patent variables measure both quantity and various quality aspects in patents and thus provide an insight into what patent information translates to firm value.

My analysis unfolds in two steps. First, I examine whether patent information is associated with and can be useful in predicting financial performance in the U.S. biotech industry. Financial performance is measured as operating income, before depreciation and the expensing of R&D, scaled by sales. The performance variable is regressed on previous 5 years' six patent variables measuring patent quantity and various aspects of patent quality in OLS model, controlling for the beginning-of-year tangible assets and previous 5 years' R&D spending. The results suggest that patent information is associated with and can be useful in predicting a biotech firm's financial performance.

Second, I investigate whether patent information adds incremental value-relevance over traditional financial information to the market valuation in biotech companies. Specifically, I use a return model derived from Ohlson (1995) method and incorporate the six patent variables to test the incremental value-relevance of patent information over traditional accounting measures. Two-way random or fixed effects models instead of OLS are used to estimate coefficients in the return model to avoid possible heteroskedasticity (see the Research Design section for discussion). Using a sample of 231 biotech firms over the years 1990–2001, I present evidence that patent information captures the biotech firms' value not currently valued by financial statement items.

There has been a growing concern among academics and practitioners about the declining value-relevance of financial statement information (e.g., Wallman, 1995, 1996; Amir & Lev, 1996; Brown, Lo, & Lys, 1999). To enhance financial reporting, SEC Commissioner Wallman (1996) suggests an accounting model that deemphasizes recognition and focuses on providing information that is ''highly relevant and consistently measurable with a high degree of reliability'' but does not meet the accounting definition of an asset, liability, or component of equity. This study illustrates the role of patents in supplementing recognized financial statement values and contributes to the stream of recent accounting literature that seeks to understand the link between nonfinancial leading indicators and future earnings (Ittner & Larcker, 1998; Behn & Riley, 1999; Rajgopal et al., 2003).

This study further contributes to the accounting literature by focusing on the biotech industry instead of the high tech sectors in general. Previous value-relevance studies address the relation between nonfinancial patent measures and financial performance in the high-tech sectors (Deng et al., 1999; Hirschey et al., 2001). However, not all high-tech industries extensively participate in patenting activities. Software development and production companies, for example, rely heavily on copyright and trademarks instead of on patenting activities. Therefore, the evidence in prior research documenting the value-relevance of patents in the high-tech sectors in general may not be sufficient and could just indicate that firms in industries that are more likely to seek and be granted patents tend to have higher value. By focusing on the biotech industry, this paper examines the value-relevance of patents on a set of firms that are equally likely (or more equally likely than those in different high-tech industries) to participate in patenting activities, and thus provides more direct and reliable evidence on the underlying issue.

This study should be of interest to both investors and standard setters. From the investor's perspective, the examined nonfinancial patent information provides a useful tool to assess biotech firms' potential profits and future cash flows. From the standard setters' perspective, it sheds light on the nonfinancial disclosure issue. Although greater disclosure is generally believed to be preferable, accounting authorities are concerned that the risk of such disclosure may outweigh the benefits. For example, inaccurate measurements or surprise write-downs of intangible assets may result in federal securities lawsuits.[2] By demonstrating the value-relevance of a set of objective and publicly available patent measures in the biotech industry, this study contributes to the discussion of what information biotech firms should disclose to improve current financial reporting.

This paper proceeds as follows. The next section contains an overview of the biotechnology industry and develops the hypotheses. The third section reviews prior research on patent measures and defines six patent variables. The research design is described in the fourth section. The fifth section outlines data sources and sample selection procedures. The sixth section presents research findings, and the final section concludes.

THE BIOTECHNOLOGY INDUSTRY AND HYPOTHESIS DEVELOPMENT

The Biotechnology Industry

In the context of current industrial practice, biotechnology commonly refers to the application of biological and biochemical science to large-scale production for the purpose of modifying human, health, food supplies, or the environment (Standard & Poor's, 2002). The biotech industry today comprises many different practices, some of which involve the alteration of genetic material. Although people recognize its potential to cure diseases, many also fear that genetic research might result in the accidental creation and release of deadly new pathogens into the environment. In the early 1980s, the Supreme Court recognized patent rights on genetically altered life forms. This ruling means that U.S. biotech firms could continue to invest in costly research projects knowing that patents would protect their discoveries and ultimately maintain financial incentives.

The biotech industry consists of more than 1,400 public and private entities with over 194,000 employees (Ernst & Young, 2003). Biotech companies range in size from small start-ups to multibillion-dollar firms. In 2001, the top 10 publicly owned U.S. biotech firms had over $13 billion in revenue, much higher than the $6 billion obtained by the top 10 in 1996. In the future, the disparity in revenue between the big firms and the emerging concerns is likely to grow primarily due to small biotech firms' lack of potential blockbuster research and products in the pipelines.

The analysis of a biotech firm, like that of any company, includes a thorough study of both business strategy and financial health. However, in contrast to companies in more mature industries, many biotech firms do not have commercial track records. The usefulness of looking at a biotech company's financial statements depends largely on whether the firm has an earnings history.[3] Because the majority of biotech companies are young and in the developmental stages, traditional analytical techniques are of limited value. For these companies, analysts and investors tend to focus on the future earnings potential of products in development and on whether the company has the researches to fully develop those products. In the absence of any explicit market information on the value of a company's research pipeline or technology, patents can serve as a proxy for the firm's knowledge base and future earning potential.

Current patents can lead to royalties if a company decides to license its technology to other firms. Patents also protect companies by preventing potential competitors from entering certain markets.[4] As the information content of financial statements may be limited, this study tests whether nonfinancial patent information fills this gap by providing important signals of financial performance in the biotech industry.

Hypotheses Development

The first part of this study examines the link between nonfinancial leading indicators and future financial performance. Advocates argue that nonfinancial indicators of investments in intangible assets may be better predictors of future financial performance than are historical accounting measures. Ittner and Larcker (1998), for example, examine the relation between customer satisfaction and financial performance and conclude that nonfinancial indicators should supplement financial measures in internal accounting systems and executive compensation plans. Behn and Riley (1999) provide empirical evidence that timely nonfinancial information can be useful in predicting financial performance in the U.S. airline industry, and they suggest that nonfinancial information disclosure may enhance traditional financial reporting. For a sample of e-commerce firms, Rajgopal et al. (2003) show that network advantages are positively associated with 1- and 2-year-ahead earnings forecasts provided by equity analysts. This previous research leads to the first hypothesis, stated in the alternative form as follows:

> **Hypothesis 1.** Nonfinancial patent information is associated with and can be useful in predicting financial performance in the U.S. biotech industry.

Another important purpose of this study is to understand the value-relevance of reported financial information and that of nonfinancial intangible knowledge capital for asset valuation in the biotech industry. Brown et al. (1999) document a declining value-relevance of financial statement information as an important determinant of the market value of the firm. Lev and Zarowin (1999) further suggest that financial statement information has less value-relevance for research-intensive firms. To enhance financial reporting, academic research turns its attention to capitalization of research and development (Lev & Sougiannis, 1996).

However, as discussed in Ely et al. (2003), capitalization is not the only way to provide investors with information. Disclosure of nonfinancial information provides an alternative that allows investors to assess the asset potential of intangibles without requiring their recognition. The Jenkins

Committee report (AICPA, 1994) stimulated a number of recent studies that examine the value-relevance of nonfinancial information. Examples include market size and market penetration in the wireless industry (Amir & Lev, 1996), customer satisfaction (Ittner & Larcker, 1998), patents in high-tech firms (Deng et al., 1999), Web traffic measures in the Internet industry (Trueman et al., 2001), and network advantages in the e-commerce sector (Rajgopal et al., 2003). The current study extends this line of research by turning attention to innovative, science-based biotech companies.

Biotech firms compete with others in an R&D intensive and technologically innovative environment. Consequently, frequent breakthrough innovations based on the firms' knowledge capital result in significant increases in the firms' asset values. The annual R&D expenditures of a firm are considered to be investments that add to a firm's knowledge asset. This knowledge asset depreciates over time so that the older R&D investment becomes less valuable as time passes. To supplement the information content, this study uses patent information as an indicator of the value of the additions to a biotech firm's underlying knowledge capital and future earnings potential. A maintained assumption of this study is that patents are an indicator of the output or "success" of R&D rather than an input of R&D. A patent grants the property right to the inventor to exclude others from making, using, offering for sale, or selling the invention. Thus, I predict that patent information captures the biotech firms' value not currently valued by traditional financial indicators. This reasoning leads to the second hypothesis, stated in the alternative form as follows:

Hypothesis 2. Nonfinancial patent information adds incremental value-relevance to the market valuation in biotech companies.

PRIOR RESEARCH AND PATENT VARIABLE DEFINITIONS

Prior Research on Patent Measures

A patent, by definition, is a temporary legal monopoly granted to inventors for the commercial use of an invention. Pakes (1985) was among the first to examine the relationship among the number of successful patent application of a firm, the firm's investment in inventive input (its R&D expenditures), and its inventive output (the stock market value of the firm).

He finds that unexpected changes in patents and in R&D are associated with large changes in the market value of the firm. When included in a market value equation, patents typically do not have as much explanatory power as an R&D measure. However, they do appear to add explanatory power above and beyond R&D (Hall, 1998). One reason patents may not exhibit much correlation with dollar denominated measures such as R&D or market value is that they are an extremely noisy measure of the underlying economic value of the innovations with which they are associated (Griliches, Pakes, & Hall, 1987; Griliches, 1990). The distribution of the value of patented innovations is known to be extremely skewed toward the low end, with a long and thin tail into the high-value side. Therefore, the number of patents held by a firm may be a poor proxy for the value of knowledge assets.[5] Some studies suggest that the number of citations received by a patent may be correlated with its economic value, so that weighting patents by the number of citations received may improve the measure (Harhoff et al., 1999).

Patent citations identify the number of times each patent has been cited in subsequent patents. Harhoff et al. (1999) survey the German patent holders of 962 U.S. invention patents that also were filed in Germany, asking them to estimate at what price they would have been willing to sell the patent right in 1980, about 3 years after the date at which they filed the German patent. The results show that the most highly cited patents are very valuable, "with a single U.S. citation implying on average more than $1 million of economic value." The citation indicators, therefore, are expected to have a high positive correlation with market value.

Various studies have shown that patent citations capture important aspects of R&D value. For example, Trajtenberg (1990) reports a positive association between citation counts and consumer welfare measures for CAT scanners; Shane (1993) finds that patent counts weighted by citations contribute to the explanation of differences in Tobin's q measures (market value over replacement cost of assets) across semiconductor companies; and Hall et al. (2000) report that citation-weighted patent counts are positively associated with firms' market values (after controlling for R&D capital). Patent counts and citations thus reflect technological elements used by investors to value companies.[6]

Other potentially informative patent measures used in prior studies are claim-weighted patents (Darby, Liu, & Zucker, 2000), science linkage, and technology cycle time (Deng et al., 1999; Hirschey et al., 2001). Patent claims define the scope of the patent protection and describe what the patented invention does that has never been done before. Although simple patents have only a few claims, broader patents may cover separable inventions, each spelled out in a separate claim. Science linkage is measured as the average number of references cited on the front page of the patent, including academic journal articles and papers presented at scientific meetings. Technology cycle time is defined as the median age in years of earlier U.S. patents referenced by a patent. It shows how quickly a technology is evolving. Empirical analysis indicates that these patent-related measures are statistically associated with subsequent stock returns (Deng et al., 1999; Hirschey et al., 2001) and market-to-book ratios (Deng et al., 1999), suggesting that patent-related measures provide a useful tool for the investment analysis of technology and science-based firms.

Patent Variable Definitions

Financial statements do not report patent information under current U.S. accounting standards. Testing the hypotheses requires identifying a set of patent variables. Based on prior research on patent measures, I develop six patent variables, including PATNUM, CLAIM, CITATION, REFAGE, REFNUM, and DNA%. PATNUM indicates the total number of U.S. patents granted to the company during a given year. CLAIM indicates the average number of claims in a firm's granted patents in a given year. Patent claims define the scope of the patent protection and describe what the patented invention does that has never been done before. CITATION, a citation intensity indicator, provides the average number of citations to the company's patents issued in a given year, divided by the average number of citations to all patents in the sample granted in the same years.[7] The percentage, rather than citation counts, is used to construct CITATION to avoid age bias caused by patents issued in earlier years receiving more citations than newly granted patents. The fundamental idea underlying the economic analysis of patent citations is that a large number of citations to a patent indicates that the examined patent represents an important invention.[8]

REFAGE is based on the average median age of the U.S. patents cited on the front page of a patent. A tendency to cite mature patents indicates that the firm engages in old technology. REFNUM indicates the average number of references to scientific journal papers and conference proceedings cited by a patent. This variable shows how strongly a patent is linked to scientific research. DNA% is the percentage of genetic patents in a firm's total granted patents in a given year. This paper is the first to identify DNA%

and use it as a patent variable as I am aware of. The percentage, rather than genetic patent counts, is used to construct the variable to avoid the possible high collinearity between patent counts and genetic patent counts. A higher DNA% indicates that the firm is more strongly linked to genetic research with high future earning potential. Of the six patent variables, PATNUM measures a biotech firm's patent quantity and CLAIM, CITATION, REFAGE, REFNUM, and DNA% assess various aspects of quality of the firm's patents. Panel A of Table 1 lists definitions for the financial and nonfinancial variables used in the empirical tests.

RESEARCH DESIGN

Performance Model

Performance model empirically tests the ability of nonfinancial measures in year *t–k* to predict future annual financial performance in year *t* (Hypothesis 1). Financial performance is defined as operating income, before depreciation and the expensing of R&D, scaled by sales. I estimate the following model modified from work done by Lev and Sougiannis (1996):

$$\begin{aligned}\left(\frac{\mathrm{OI}}{S}\right)_{jt} &= \gamma_0 + \gamma_1\left(\frac{\mathrm{TA}}{S}\right)_{j,t-1} + \sum_k \gamma_{2,k}\left(\frac{\mathrm{R\&D}}{S}\right)_{j,t-k} + \sum_k \gamma_{3,k}\left(\frac{\mathrm{PATNUM}}{S}\right)_{j,t-k} \\ &\quad + \sum_k \gamma_{4,k}\mathrm{CLAIM}_{j,t-k} + \sum_k \gamma_{5,k}\mathrm{CITATION}_{j,t-k} + \sum_k \gamma_{6,k}\mathrm{REFAGE}_{j,t-k} \\ &\quad + \sum_k \gamma_{7,k}\mathrm{REFNUM}_{j,t-k} + \sum_k \gamma_{8,k}\mathrm{DNA\%}_{j,t-k} + \varepsilon_{jt}, \quad k = 1, 2, \ldots 5. \end{aligned} \tag{1}$$

where OI is annual operating income, before depreciation and R&D expenses, of firm *j* in year *t*; *S* the annual sales; TA the value of plant and equipment and inventory measured at the beginning-of-year values; R&D the annual R&D expenditures; and PATNUM the total number of U.S. patents granted. The five patent quality variables (defined in the previous section) represent averaged patent attributes (CLAIM, REFAGE, and REFNUM) or indices (CITATION and DNA%) and can measure the potential of a biotech firm's patents in turning to marketable and quality products. If coefficients $\gamma_{3,k}$ to $\gamma_{8,k}$ in Eq. (1) appear to be jointly statistically significant, one can conclude that current and lagged patent information have incremental explanatory power in future earnings.

Table 1. Variable Definitions and Descriptive Statistics.

Panel A: Variable definitions

Variables	Definition
Dependent	
RETURN	Change in market value of a firm in year t divided by market value of the firm at the beginning of year t
OI/S	Operating income, before depreciation and the expensing of R&D, scaled by sales
Financial	
$1/MV_{t-1}$	Inverse of the market value of a firm at the beginning of year t
ΔE	Change in earnings before R&D expenditures of a firm in year t
E	Earnings before R&D expenditures at the end of year t
R&D	R&D expenditures
Patent	
PATNUM	The total number of US patents granted to the company during a given year
CLAIM	The average number of claims on a firm's granted patents in a given year
CITATION	The average number of citations to the firm's patents issued in a given year, divided by the average number of citations to all patents in the sample granted in the same year
REFAGE	The average median age of the US patents cited on the front page of a patent in a given year
REFNUM	The average number of references to scientific journal papers and conference proceedings cited by a patent in a given year
DNA%	The percentage of genetic patents in a firm's total granted patents in a given year
PATNUM × CLAIM	PATNUM and CLAIM interaction
PATNUM × CITATION	PATNUM and CITATION interaction
PATNUM × REFAGE	PATNUM and REFAGE interaction
PATNUM × REFNUM	PATNUM and REFNUM interaction
PATNUM × DNA%	PATNUM and DNA% interaction
Control	
BM	Book-to-market ratio
MV	Total market value of a firm at the end of a given year
BETA	CAPM-beta of each firm

Panel B: Descriptive statistics

Variable	Mean	Standard Deviation	Minimum	Median	Maximum
OI/S	−2.68	34.68	−492.45	0.08	806.37
RETURN	0.51	2.37	−0.98	−0.09	40.25
Financial					
TA/S	4.73	39.07	0	0.68	1151
RND/S	10.89	51.52	0	1.51	1,001.37
$\Delta E/MV_{t-1}$	0.02	0.22	−2.23	0.00	4.69

Table 1. (*Continued*)

Panel B: Descriptive statistics

Variable	Mean	Standard Deviation	Minimum	Median	Maximum
E/MV_{t-1}	0.01	0.18	−2.27	0.00	1.08
$\mathrm{RND/MV}_{t-1}$	0.15	0.26	0	0.09	4.68
Patent					
PATNUM/S	1.80	12.41	0	0.14	333.33
$\mathrm{PATNUM/MV}_{t-1}$	0.05	0.16	0	0.01	2.23
CLAIM	13.20	14.60	0	11.71	176
CITATION	0.71	1.63	0	0.21	30.61
REFAGE	4.43	4.29	0	4.41	27.08
REFNUM	16.30	23.39	0	8.11	230
DNA%	0.26	0.38	0	0	1.00
$(\mathrm{PATNUM} \times \mathrm{CLAIM})/\mathrm{MV}_{t-1}$	0.89	3.37	0	0.14	69.82
$(\mathrm{PATNUM} \times \mathrm{CITATION})/\mathrm{MV}_{t-1}$	0.03	0.09	0	0	1.24
$(\mathrm{PATNUM} \times \mathrm{REFAGE})/\mathrm{MV}_{t-1}$	0.29	1.13	0	0.05	18.52
$(\mathrm{PATNUM} \times \mathrm{REFNUM})/\mathrm{MV}_{t-1}$	0.94	3.42	0	0.10	56.08
$(\mathrm{PATNUM} \times \mathrm{DNA\%})/\mathrm{MV}_{t-1}$	0.02	0.10	0	0	1.89
Control					
BM	0.37	0.46	0.00	0.25	5.80
MV	999.48	5,973.75	0.34	128.93	87,878.60
BETA	1.68	0.82	0.00	1.62	6.44

Observations = 1,183 (872 for BETA; 1,080 for OI/S, TA/S, and RND/S)

Value-Relevance Model

To test whether patent measures capture the biotech firms' value not currently valued by traditional financial indicators (Hypothesis 2), I use a return model as a baseline model and examine the relation between financial variables and returns of biotech firms.[9] Motivated by the work of Fama and French (1992), the return model in Eq. (2) adds book-to-market ratio, market value, and CAPM-beta as control variables.

$$\begin{aligned} \mathrm{RETURN}_{jt} = {} & \alpha_0 + \sum_{yr=1990}^{2001} \alpha_{yr+1}\mathrm{YR} + \sum_{n=1}^{292} \alpha_{n+1}\mathrm{FIRM} + \alpha_1 \frac{E_{jt}}{\mathrm{MV}_{jt-1}} \\ & + \alpha_2 \frac{\Delta E_{jt}}{\mathrm{MV}_{jt-1}} + \alpha_3 \frac{\mathrm{R\&D}_{jt}}{\mathrm{MV}_{jt-1}} + \alpha_4 \mathrm{BM}_{jt} + \alpha_5 \mathrm{MV}_{jt} \\ & + \alpha_6 \mathrm{BETA}_{jt} + \varepsilon_{jt}, \end{aligned} \tag{2}$$

where RETURN_{jt} is measured as the change in market value of firm j from year $t-1$ to year t divided by market value at the end of year $t-1$; YR and FIRM are year and firm dummy variables included to control for time and firm variation; MV_{jt-1} and MV_{jt} are total market value of firm j at the end of year $t-1$ and year t, respectively; E_{jt} is earnings before R&D expenditures of firm j at the end of year t; ΔE_{jt} the change in earnings before R&D expenditures of firm j in year t; $\mathrm{R\&D}_{jt}$ the R&D expenditures of firm j at the end of year t; BM_{jt} the book-to-market ratio of firm j at the end of year t; and BETA_{jt} the CAPM-beta of firm j, estimated from 60 monthly stock returns (minimum of 24) to the end of year t.

To test the incremental value-relevance of patent information, I then regress returns on financial information along with six patent variables using the following model:

$$
\begin{aligned}
\mathrm{RETURN}_{jt} = {} & \beta_0 + \sum_{yr=1990}^{2001} \beta_{yr+1}\mathrm{YR} + \sum_{n=1}^{292} \beta_{n+1}\mathrm{FIRM} + \beta_1 \frac{E_{jt}}{\mathrm{MV}_{jt-1}} + \beta_2 \frac{\Delta E_{jt}}{\mathrm{MV}_{jt-1}} \\
& + \beta_3 \frac{R\&D_{jt}}{\mathrm{MV}_{jt-1}} + \beta_4 \frac{\mathrm{PATNUM}_{jt}}{\mathrm{MV}_{jt-1}} + \beta_5 \mathrm{CLAIM}_{jt} + \beta_6 \mathrm{CITATION}_{jt} \\
& + \beta_7 \mathrm{REFAGE}_{jt} + \beta_8 \mathrm{REFNUM}_{jt} + \beta_9 \mathrm{DNA\%}_{jt} \\
& + \beta_{10}\left(\frac{\mathrm{PATNUM}_{jt}}{\mathrm{MV}_{jt-1}} \times \mathrm{CLAIM}_{jt}\right) + \beta_{11}\left(\frac{\mathrm{PATNUM}_{jt}}{\mathrm{MV}_{jt-1}} \times \mathrm{CITATION}_{jt}\right) \\
& + \beta_{12}\left(\frac{\mathrm{PATNUM}_{jt}}{\mathrm{MV}_{jt-1}} \times \mathrm{REFAGE}_{jt}\right) + \beta_{13}\left(\frac{\mathrm{PATNUM}_{jt}}{\mathrm{MV}_{jt-1}} \times \mathrm{REFNUM}_{jt}\right) \\
& + \beta_{14}\left(\frac{\mathrm{PATNUM}_{jt}}{\mathrm{MV}_{jt-1}} \times \mathrm{DNA\%}_{jt}\right) + \beta_{15}\mathrm{BM}_{jt} + \beta_{16}\mathrm{MV}_{jt} \\
& + \beta_{17}\mathrm{BETA}_{jt} + \varepsilon_{jt} \qquad (3)
\end{aligned}
$$

where $\mathrm{PATNUM}_{jt}/\mathrm{MV}_{jt-1}$ is the total number of U.S. patents granted to firm j in a given year t, deflated by the firm's market value at the end of year $t-1$; CLAIM, REFAGE, REFNUM CITATION, and DNA% are the same set of patent quality variables used in Eq. (1); $(\mathrm{PATNUM}_{jt}/\mathrm{MV}_{jt-1}) \times \mathrm{CLAIM}_{jt}$, $(\mathrm{PATNUM}_{jt}/\mathrm{MV}_{jt-1}) \times \mathrm{CITATION}_{jt}$, $(\mathrm{PATNUM}_{jt}/\mathrm{MV}_{jt-1}) \times \mathrm{REFAGE}_{jt}$, $(\mathrm{PATNUM}_{jt}/\mathrm{MV}_{jt-1}) \times \mathrm{REFNUM}_{jt}$, and $(\mathrm{PATNUM}_{jt}/\mathrm{MV}_{jt-1}) \times \mathrm{DNA\%}_{jt}$ are interaction terms of deflated PATNUM and each of the five patent quality variables. In Eq. (3) each patent quality variable may influence returns directly (in coefficients β_5–β_9) or indirectly through the interaction with PATNUM (in coefficients β_{10}–β_{14}). If the coefficients of patent variables (β_4–β_{14}) are jointly statistically significant and the R^2 of regression (3) exceeds that of regression (2), one can conclude that patent information adds incremental value-relevance to the market value of the biotech companies.

One concern about Eqs. (2) and (3) is that the error term is proportional to $1/(\mathrm{MV}_{t-1})$ and may result in heteroskedasticity.[10] Therefore, this study uses two-way random or fixed effects models instead of OLS to estimate coefficients in the return model. Two-way random or fixed effects models are designed to handle various types of heteroskedasticity. They are more general approaches than a more structural GLS approach, which would ignore other unknown sources or forms of heteroskedasticity.

A potential problem in the above equations incorporating patent variables is the degree of multicollinearity, which results in a higher standard error of estimate. I apply tests to examine the degree of multicollinearity, including using an F-test for the full model and checking variance inflation factor (VIF). Multicollinearity is not an issue in testing the hypotheses if the patent variables in the equation are jointly statistically significant. However, if the patent variables are not jointly statistically significant and the multicollinearity among them appears to be high, a patent index will be created to alleviate this problem.

Another concern is that the above models might contain omitted variables. Factors not incorporated in this study, such as human capital, strategic alliances with major pharmaceutical firms, FDA approvals, and technology platform, also could drive a biotech firm's value. These potential value drivers are not incorporated in this study because they either are difficult to quantify or do not apply to the entire biotech sample, e.g., the number of FDA approvals is not a valid value drivers to a pure biotech firm without commercialized products or drugs in the pipeline.

DATA SOURCES AND SAMPLE SELECTION

Financial analysts and investors have various definitions for the biotech industry. This study adopts Standard & Poor's Market Insight industry classification, which defines the biotechnology industry as companies primarily involved in the development, manufacturing, or marketing of products based on advanced biotechnology research. Preliminary investigation reveals an initial sample contains 292 U.S. biotechnology companies on Market Insight that generate $27,090 million in combined annual sales (based on 12-month moving data). The analysis is based on these companies' financial and nonfinancial data for the years 1990–2001. A search of Compustat results in 255 possible sample companies with 1,551 firm-year observations available. Of those 1,551 firm-year observations, 273 are lost when constructing 1-year lagged data, and 95 firm-year observations are deleted because of negative book values. Firms with negative book values are eliminated from the analysis because, practically, those firms are bankrupt and the normal assumption of the earnings-returns or patent-returns relation may not hold. The final sample consists of 231 companies with 1,183 firm-year observations.

Nonfinancial patent data on five patent attributes (PATNUM, CLAIM, CITATION, REFAGE, and REFNUM) for the years 1990–2001 are from the United States Patent and Trademark Office's (USPTO) Web page. If a patent found on the USPTO's Web page matches with one on the DNA Patent Database (DPD), the patent is identified as a DNA patent. The DPD is a joint project of the Georgetown University's Kennedy Institute of Ethics and the Foundation for Genetic Medicine. The DPD is being created to enable relevant empirical studies of DNA-based patents issued in the United States. Patents included in the DPD were identified by virtue of their USPTO classification numbers and the presence of keywords such as "DNA," "nucleotide," or "polynucleotide" in one or more claims. All financial data are obtained from the Compustat database.

EMPIRICAL RESULTS

Descriptive Statistics

Panel B of Table 1 reports descriptive statistics for the variables used in regression models. Table 2 provides a correlation table for financial and patent variables. The lower left-hand side of the table reports Pearson correlations, and the higher right-hand side reports Spearman rank correlations. The Spearman rank correlations show significant positive correlations among patent variables, and Pearson correlations are generally consistent with the Spearman results. The high correlations among patent variables might impair the models' ability to explain the variation in returns. However, the low VIFs and the joint statistical significance of patent variables in the next section suggest that the high correlations among the patent variables do not affect the regression results.

Table 3 provides reports the SIC composition and the patenting activities of the sample. Approximately 81% of the sample observations are in the drugs and pharmaceuticals segment (SIC codes beginning with 283), and, on average, 69% of those in drugs and pharmaceuticals segment have patents.

Table 2. Correlation Matrix.

Variables	$\Delta E/\mathrm{MV}_{t-1}$	E/MV_{t-1}	$\mathrm{RND/MV}_{t-1}$	$\mathrm{PATNUM/\ MV}_{t-1}$	CLAIM	CITATION	REFAGE	REFNUM	DNA%
$\Delta E/\mathrm{MV}_{t-1}$	1	**0.52**	0.12	0.02	−0.02	−0.02	−0.02	−0.02	0.02
		(0.00)	(0.00)	(0.41)	(0.57)	(0.60)	(0.47)	(0.57)	(0.49)
E/MV_{t-1}	**−0.70**	1	0.17	0.18	0.05	0.05	0.03	0.13	0.15
	(0.00)		(0.00)	(0.00)	(0.11)	(0.09)	(0.23)	(0.00)	(0.00)
$\mathrm{RND/MV}_{t-1}$	0.07	0.14	1	0.14	0.05	0.09	0.02	0.12	0.18
	(0.04)	(0.00)		(0.00)	(0.09)	(0.00)	(0.43)	(0.00)	(0.00)
$\mathrm{PATNUM/MV}_{t-1}$	0.11	−0.11	0.11	1	**0.72**	**0.67**	**0.65**	**0.74**	**0.58**
	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
CLAIM	−0.03	−0.00	0.01	0.08	1	**0.65**	**0.70**	**0.69**	**0.46**
	(0.39)	(0.93)	(0.74)	(0.0I)		(0.00)	(0.00)	(0.00)	(0.00)
CITATION	−0.01	0.00	−0.05	−0.01	0.25	1	**0.55**	**0.60**	**0.42**
	(0.67)	(0.89)	(0.11)	(0.83)	(0.00)		(0.00)	(0.00)	(0.00)
REFAGE	−0.02	−0.01	−0.07	0.11	**0.52**	0.25	1	**0.65**	0.30
	(0.61)	(0.66)	(0.05)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)
REFNUM	−0.01	0.02	−0.02	0.05	**0.40**	0.27	**0.41**	1	**0.55**
	(0.79)	(0.49)	(0.62)	(0.16)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
DNA%	−0.02	0.08	0.03	0.16	0.29	0.19	0.17	0.32	1
	(0.54	(0.02)	(0.36)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	

Note: The lower left-hand side of the matrix reports Pearson correlations, and the upper right-hand side reports Spearman rank correlations. Correlations greater than 0.40 are in bold, and probability $>|r|$under H_0: Rho = 0 is in parenthesis. See Panel A of Table 1 for variable definitions.

Table 3. SIC Composition for Firms With and Without Patents.

SIC Composition		Firm: Year Observations				
		Without patent	With patent	Subtotal (number of observations and %)		With-patent observations as a % of sub-total
2836	Biological products, except diagnostic substances	149	326	475	40.15	68.63
2834	Pharmaceutical preparations	53	176	229	19.36	76.86
2835	In vitro and in vivo diagnostic substances	84	136	220	18.60	61.82
8731	Commercial physical and biological research	20	28	48	4.06	58.33
2833	Medicinal chemicals and botanical products	7	24	31	2.62	77.42
3841	Surgical and medical instrument and apparatus	11	14	25	2.11	56.00
3842	Orthopedic, prosthetic, and surgical appliances and supplies	7	13	20	1.69	65.00
Others (0100, 2810, 2820, 2821, 2840, 2860, 2844, 2870, 2890, 3559, 3580, 3826, 3845, 3829, 7370, 5160, 6552, 6794, 7372, 8071, 9995)		50	85	135	11.41	62.96
Total		381	802	1,183	100.00	67.79

Note: Other SICs list the SICs with less than 1% of the total sample observations (12 firm-year observations).

Similarly, about 68% of overall sample observations are in the with-patent group, suggesting a general tendency of biotech firms to engage in patenting activities.

Generally, the mean values of the variables for firms with patents are significantly different than those for firms without patents at conventional levels (not tabulated). Compared with the without-patent group, the with-patent group has larger numbers in MV, natural log of assests, and R&D. Apparently, large biotech firms with intensive R&D investment are more likely to be successful in knowledge assets development. The statistics of performance variable reflect the importance of knowledge assets in the biotech industry, given that, on average, firms with patents have better performance than those without patents in terms of operating income (adjusted for depreciation and expensing for R&D), scaled by sales. The with-patent group also has higher mean and median values in RETURN, which implies that patents add to biotech firms' future earning potential and that investors value patents in market valuation.

Multivariate Results

Table 4 reports the OLS results of Eq. (1), where financial performance is regressed on tangible assets, lagged R&D, and lagged patent variables. Because the sample covers 12-year data and most sample firms are younger than 12 years, this study did not examine the explanatory power of lagged patent information beyond 5 years. Lagged patent variables (up to 5 years) are jointly significantly associated with firm's adjusted operating income scaled by sales ($F = 1.69$, $p < 0.01$), suggesting that lagged patent information is useful in predicting a biotech firm's adjusted operating income over sales.[11] The result supports Hypothesis 1.

The coefficient of deflated tangible assets is positive in the performance regression but the coefficients of the 1- and 2-year lagged deflated R&D expenditures are negative at the conventional significance level. In other words, a biotech firm's operating income improves with an increase in tangible asset investment but deteriorates with increases in 1- or 2-year lagged R&D expenditures with zeros in all patent attributes. Interestingly, the positive and statistically significant coefficients of $\gamma_{3,2}$ and $\gamma_{5,2}$ indicate that deflated PATNUM and CITATION, with a 2-year lag, are associated with a biotech firm's financial performance. Currently, the term of a new patent is 20 years from the date on which the application for the patent was filed in the United States. However, benefits from a patent usually last less

Table 4. OLS Regression of Performance Variables on Lagged R&D and Patent Variables.

Variables	Coefficient	Variables	Coefficient	Variables	Coefficient
γ_0	−1.29	$\gamma_{4,1}$	0.01	$\gamma_{6,3}$	−0.05
	(−0.98)		(0.23)		(−0.23)
γ_1	0.42**	$\gamma_{4,2}$	0.00	$\gamma_{6,4}$	0.12
	(2.32)		(−0.02)		−0.56
$\gamma_{2,1}$	−0.21**	$\gamma_{4,3}$	0.03	$\gamma_{6,5}$	−0.28
	(−2.18)		(0.47)		(−1.24)
$\gamma_{2,2}$	−0.26***	$\gamma_{4,4}$	0.05	$\gamma_{7,1}$	0.02
	(−8.26)		(0.65)		−0.41
$\gamma_{2,3}$	0.00	$\gamma_{4,5}$	0.03	$\gamma_{7,2}$	−0.05
	(0.11)		(0.44)		(−1.03)
$\gamma_{2,4}$	0.01	$\gamma_{5,1}$	−0.21	$\gamma_{7,3}$	−0.02
	(0.49)		(−0.29)		(−0.32)
$\gamma_{2,5}$	0.01	$\gamma_{5,2}$	1.84*	$\gamma_{7,4}$	0.04
	(0.32)		(1.67)		(0.66)
$\gamma_{3,1}$	0.02	$\gamma_{5,3}$	0.53	$\gamma_{7,5}$	0.02
	(0.09)		(0.63)		(0.50)
$\gamma_{3,2}$	0.91***	$\gamma_{5,4}$	−1.30	$\gamma_{8,1}$	−0.25
	(6.12)		(−1.55)		(−0.09)
$\gamma_{3,3}$	0.02	$\gamma_{5,5}$	−0.42	$\gamma_{8,2}$	−1.95
	(0.21)		(−0.58)		(−0.71)
$\gamma_{3,4}$	0.00	$\gamma_{6,1}$	0.13	$\gamma_{8,3}$	−0.18
	(0.01)		(0.65)		(−0.06)
$\gamma_{3,5}$	0.19	$\gamma_{6,2}$	−0.12	$\gamma_{8,4}$	−1.80
	(0.69)		−(0.59)		(−0.63)
				$\gamma_{8,5}$	2.39
					(0.86)
Adj. R^2				0.11	
F-statistics for overall model				2.30***	
F-statistics for joint significance of patent variables				1.69**	

Note: Refer Eq. (1). *t*-statistics are in the parenthesis.
*Significance 10% levels.
**Significance 5% levels.
***Significance 1% levels.

than the patent period because competition takes away the competitive advantage. Therefore, it is not surprising that patent information is not associated with future profitability measures beyond 2 years.

Table 5 presents results of regressing returns on financial, patent, and control variables, using two-way random or fixed effects models. The results of fixed effects models are omitted because the low H values (reported in the

Table 5. Regression of Returns on Financial, Patent and Control Variables Using Two-Way Random or Fixed Effects Models.

Variables	Coefficients			Mean[a]
	A	B	C	
Constant	0.32	0.23	0.31	
	(0.96)	(0.71)	−(0.93)	
$\Delta E/\mathrm{MV}_{t-1}$	1.62***	1.43***	1.08***	
	(4.18)	(3.65)	(2.79)	
E/MV_{t-1}	−1.13**	−0.95*	0.19	
	(−1.99)	(−1.68)	−(0.33)	
$\mathrm{R\&D/MV}_{t-1}$	2.17***	2.06***	1.54***	
	(6.87)	(6.51)	(4.88)	
$\mathrm{PATNUM/MV}_{t-1}$		1.69***	−10.77***	0.05
		(3.13)	(−4.37)	
CLAIM			−0.02***	13.91
			(−3.24)	
CITATION			−0.11	0.68
			(−1.48)	
REFAGE			0.01	4.70
			(0.35)	
REFNUM			−0.005	17.52
			(−0.99)	
DNA%			0.70**	0.28
			(2.51)	
$(\mathrm{PATNUM} \times \mathrm{CLAIM})/\mathrm{MV}_{t-1}$			0.41***	0.90
			(5.16)	
$(\mathrm{PATNUM} \times \mathrm{CITATION})/\mathrm{MV}_{t-1}$			8.73***	0.03
			(5.37)	
$(\mathrm{PATNUM} \times \mathrm{REFAGE})/\mathrm{MV}_{t-1}$			0.59***	0.31
			(2.79)	
$(\mathrm{PATNUM} \times \mathrm{REFNUM})/\mathrm{MV}_{t-1}$			0.04	1.05
			(0.78)	
$(\mathrm{PATNUM} \times \mathrm{DNA\%})/\mathrm{MV}_{t-1}$			−1.26	0.02
			(−0.52)	
BM	−1.03***	−1.04***	−0.94***	
MV	0.00	0.00	0.00	
	(−0.32)	(−0.22)	(−0.05)	
BETA	0.14	0.16	0.16	
	(1.18)	(1.36)	(1.36)	
R^2	0.12	0.13	0.22	
Hausman test	3.76	3.69	9.04	
F-statistics for overall model	18.59***	17.63***	13.84***	
F-statistics for joint significance of patent variables			8.12***	
Observations	860	860	860	

Note: Refer Eq. (3). *t*-statistics are in the parenthesis.
*Significance 10% levels.
**Significance 5% levels.
***Significance 1% levels.
[a]The average value of each variable in the sample.

Hausman test) favors random effects models. As shown in column C, the R^2 of the full model is ten percentage points higher than that of the regression with only financial and control variables (column A) ($R^2 = 0.22$ vs. 0.12). The *F*-statistic to test the incremental value-relevance of patent information is 8.12, rejecting the null hypothesis that the coefficients of all patent variables (β_4–β_{14}) are jointly zero in the full model at the 1% significance level. These results support Hypothesis 2 that patent information adds incremental value-relevance to the market valuation of the biotech companies.

The negative coefficient estimate of deflated PATNUM in column C of Table 5 indicates that a patent with zeros in all patent attributes reduces a biotech firm's return. However, a biotech firm's patent with average patent attributes contribute to a 0.07 percentage point increase in its return,[12] consistent with the positive coefficient of deflated PATNUM shown in column B. In addition, to test whether each patent variable is individually significantly associated with returns in the full model, I use the *F* test to determine whether the coefficients of all regressors involving the underlying patent variable are jointly zero (e.g., CLAIM is value relevant if the null hypothesis that both β_5 and β_{10} equal to zero is rejected). Results reveal that PATNUM, CLAIM, CITATION, REFAGE, and DNA% each has influence on a biotech firm's returns in the full model.[13]

Despite the high correlations among patent variables as shown in Table 2, the VIFs (not reported) suggest that no notable multicollinearity exists. I also test an alternative model specification of Eq. (3) without presence of interaction terms. The results agree with the findings presented in Table 5 and support the incremental value-relevance of patent information. Particularly, the *t* test shows that PATNUM and DNA% are significantly positively associated with returns in this model specification.

Sensitivity Tests

The sensitivity test examines whether patent information is value relevant in a subsample of biotech firms with losses before R&D expenditures. Hayn (1995) reports an unusual earnings–returns relation when earnings are negative. Ertimur (2003) shows that firms reporting accounting losses experience higher levels of information asymmetry among investors than do those reporting profits. In this situation, one should expect to see that patent information mitigates the information asymmetry and provides more explanatory power in firms with net losses. Table 6 reports the regression

Table 6. Regression of Returns on Financial, Patent and Control Variables in a Subsample of Firms with Losses Before R&D Expenditures.

Variables	Coefficients A	Coefficients B	Mean[a]
Constant	−0.35	−0.09	
	(−0.64)	(−0.18)	
ΔE_t/MVn	3.87***	3.44***	
	(4.28)	(3.58)	
E_t/MV_{t-1}	−5.76***	−4.22***	
	(−4.78)	(−3.15)	
R&D/MV_{t-1}	1.52*	−0.30	
	(1.64)	(−0.34)	
PATNUM/MV_{t-1}		−31.50***	0.05
		(−4.83)	
CLAIM		−0.06***	13.60
		(−3.57)	
CITATION		0.05	0.67
		(0.26)	
REFAGE		0.03	4.72
		(0.28)	
REFNUM		−0.01**	16.02
		(−2.17)	
DNA%		1.46***	0.22
		(2.78)	
(PATNUM × CLAIM)/MV_{t-1}		0.73***	0.94
		(4.54)	
(PATNUM × CITATION)/MV_{t-1}		13.92***	0.04
		(5.44)	
(PATNUM × REFAGE)/MV_{t-1}		2.42***	0.31
		(3.06)	
(PATNUM × REFNUM)/MV_{t-1}		0.29**	0.91
		(2.43)	
(PATNUM × DNA%)/MV_{t-1}		−5.23**	0.02
		(−0.83)	
BM	−0.74*	−0.47	
	(−1.90)	(−1.36)	
MV	0.00	0.00**	
	(1.08)	(1. 95)	
BETA	0.28	0.38**	
	(1.30)	(1.92)	
Adj. R^2	0.13	0.35	
Hausman test	4.48	19.51	
F-statistics for overall model	9.59***	12.07***	
F-statistics for joint significance of patent variables		9.66***	
Observations	391	391	

Note: t-statistics are in the parenthesis.
*Significance 10% levels.
**Significance 5% levels.
***Significance 1% levels.
[a]The average value of each variable in the sample.

results supporting this line of reasoning. Compared with the ten percentage points increase in the R^2 in the full sample (columns C vs. A in Table 5), the R^2s in the regressions incorporating both financial and patent variables are 22 percentage points higher than the R^2 in the regression with only financial variables (columns B vs. A in Table 6). The results indicate that patent variables provide more explanatory power for firms with losses before R&D expenditures than for those in the full sample. Patent information help mitigates the information asymmetry in biotech firms with losses before R&D expenditures. In addition, a patent with average value of all attributes results in 0.09 percentage point increase in a biotech firm's return, consistent to the findings in Table 5.

CONCLUSION

In this study, I investigate whether nonfinancial patent information is useful to investors in predicting biotech firms' future financial performance and examine whether nonfinancial patent information adds incremental value-relevance over financial information. Using a sample of 231 biotech firms over the years 1990–2001, I found evidence consistent with the idea that patent information is associated with and can be useful in predicting a biotech firm's long-term financial performance. In addition, patent information captures the biotech firms' value not currently formally valued by traditional financial indicators and adds incremental value-relevance to the market valuation of the biotech companies.

These results enhance our understanding of nonfinancial patent information in supplementing recognized financial statement values. FASB promotes the importance of measurement and recognition of internally developed intangible assets in financial statements (FASB, 2001). As full expensing of R&D cost fails to sufficiently inform market participants about a firm's R&D activities and potential future earnings power, this paper illustrates the role of nonfinancial patent information as an indicator of inventive output and provides insight into what patent variables translate to firm value.

This research is important because both academics and standard setters have expressed concerns about the declining importance of financial reporting and disclosure and have suggested that nonfinancial leading indicators showing how key business processes are performing may enhance financial statement users' ability to evaluate and predict financial performance. Given the current debate over what information should be disclosed and audited, this study contributes to the existing literature by providing empirical evidence that the disclosure of biotech companies could be improved by using all the value drivers in the business, including both financial results and value-enhancing nonfinancial patent measures.

NOTES

1. According to Ernst & Young, R&D expenditures by public biotech firms reached \$16.3 billion in 2002, up from \$11.6 billion in 2001 and \$9.9 billion in 2000.
2. According to Halsey Bullen, senior project manager at FASB (Business Week, Auguest 26, 2002, p. 110).
3. Amir and Lev (1996) suggest that while negative earnings may have no value implications, the change of such earnings appear to be relevant for securities pricing in the predominantly negative earnings biotech sample.
4. A recent high-profile legal case involving Transkaryotic Therapies, Inc. and Amgen underscores the value of patents. Transkaryotic developed a version of Amgen's Epogen anemia drug by a different manufacturing process than Amgen used. In January 2001, the court ruled that Transkaryotic's process did infringe upon a patent held by Amgen and enjoined Transkaryotic from entering the market that Epogen serves (Standard & Poor's, 2002).
5. Knowledge assets include rights to future benefits emanating from discovery and development activities (e.g., patents, know-how); brands, franchises, and other customer-related assets; and unique organizational designs of corporations (Lev, 2000).
6. The research using patent counts and citations as R&D output measures is summarized in Griliches (1990) and Hall et al. (2000).
7. For example, if a biotech firm's 1996 patents on average received two citations from later patents up to the end of year 2001 and all the 1996 patents of the sample firms on average received 0.5 citations from later patents during the same time period, then the firm's CITATION in 1996 is 4, calculated as 2 divided by 0.5.
8. The CITATION measure in this study includes ''self citations,'' citations to a company's patent in subsequent patents of the same company. Self-citation may indicate that the company continues to build on its earlier inventions. This interpretation implies that self-citations are more valuable than citations from others. Hall et al. (2000) find that ''the self-citation effect is small and positive: if the 'self' share of citations is higher, the market value is higher, other things equal.'' Because the self-citation effect is not considered to be significant, this study does not adjust its influence on the CITATION measure.
9. Model development is in the Appendix.
10. See Eq. (A3) in the Appendix.
11. The results of regressing financial performance on tangible assets and lagged R&D variables are not statistically significant ($F=1.33$, $p=0.24$) and are not reported in Table 4.
12. It is calculated as sum of the multiplicative results of the mean and coefficient estimate of each patent variable. That is, $0.05\times(-10.77)+13.91\times(-0.02)+0.68\times(-0.11)+4.70\times0.01+17.52\times(-0.005)+0.28\times(0.70)+0.90\times0.41+0.03\times8.73+0.31\times0.59+1.05\times0.04+0.02\times(-1.26)\approx0.07$.

13. The *F*-statistics are 11.52 for CLAIM, 15.37 for CITATION, 5.20 for REFAGE, and 4.87 for DNA%. Each is significant at the 1% level.

ACKNOWLEDGMENTS

This paper is based on my dissertation at The University of Tennessee at Knoxville. I thank Bruce Behn (my dissertation chair) for his guidance and advice. I also thank Ken Anderson, Don Bruce, Ron Shrieves, Ran Barniv, Tom Robinson, Paul Gao, and workshop participants at Kent State University, Saint Louis University, Southern Illinois University Edwardsville, Illinois State University, and University of Miami for helpful comments and discussions.

REFERENCES

American Institute of Certified Public Accountants (AICPA). (1994). *Improving business reporting – A customer focus: Meeting the information needs of investors and creditors*. Jenkins Committee Report. New York, NY.

Amir, E., & Lev, B. (1996). Value-relevance of nonfinancial information: The wireless communications industry. *Journal of Accounting and Economics*, *22*, 3–30.

Behn, B., & Riley, R. (1999). Using nonfinancial information to predict financial performance: The case of the U.S. airline industry. *Journal of Accounting, Auditing, and Finance* (Winter), 29–56.

Brown, S., Lo, K., & Lys, T. (1999). Use of R^2 in accounting research: Measuring changes in value relevance over the last four decades. *Journal of Accounting and Economics*, *28*, 83–115.

Business Week. August 26, 2002: 110. New York, NY.

Darby M. R., Liu, Q., and Zucker, L. G. (2000). *High stakes in high technology: High-tech market values as real options*. Working Paper. UCLA.

Deng, Z., Lev, B., & Narin, F. (1999). Science and technology as predictors of stock performance. *Financial Analysts Journal*, *55*(May/June), 20–32.

Easton, P. (1999). Security returns and the value relevance of accounting data. *Accounting Horizons*, *13*, 399–412.

Ely, K., Simko, P. J., & Thomas, L. G. (2003). The usefulness of biotechnology firms' drug development status in the evaluation of research and development costs. *Journal of Accounting, Auditing and Finance*, *18*(1), 163–195.

Ernst & Young. (2003). Resilience: Americas Biotechnology Report 2003.

Ertimur, Y. (2003). *Financial information environment of loss firms*. Working Paper. NYU.

Fama, E., & French, K. (1992). The cross section of expected stock returns. *Journal of Finance*, *47*, 427–465.

Financial Accounting Standards Board. (2001). *Business and financial reporting: Challenges from the new economy*. FASB: Norwalk, CT.

Griliches, Z., Pakes, A., & Hall, B. (1987). The value of patents as indicators of inventive activity. In: P. Dasgupta & P. Stoneman (Eds), *Economic policy and technological performance* (pp. 97–124). Cambridge, England: Cambridge University Press.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, *28*, 1661–1707.

Hall, B. (1998). Innovation and market value, prepared for the NIESR conference on Productivity and Competitiveness, London, England.

Hall, B., Jaffe, A., & Trajtenberg, M. (2000). *Market value and patent citations: A first look*. NBER Working Paper 7741. Cambridge, MA.

Hand, J. (2001). *The market valuation of biotechnology firms and biotechnology R&D*. Working Paper. UNC Chapel Hill.

Harhoff, D., Narin, F., Scherer, F., & Vopel, K. (1999). Citation frequency and the value of patented innovations. *The Review of Economics and Statistics*, *81*(3), 511–515.

Hayn, C. (1995). The information content of losses. *Journal of Accounting and Economics*, *20*, 125–153.

Hirschey, M., Richardson, V., & Scholz, S. (2001). Value relevance of nonfinancial information: The case of patent data. *Review of Quantitative Finance and Accounting*, *17*, 223–235.

Ittner, C., & Larcker, D. (1998). Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research*, *36*(Supplement), 1–35.

Lev, B. (2000). *Knowledge and shareholder value*. Working Paper. NYU.

Lev, B., & Sougiannis, T. (1996). The capitalization, amortization, and value-relevance of R&D. *Journal of Accounting and Economics*, *21*, 107–138.

Lev, B., & Zarowin, P. (1999). The boundaries of financial reporting and how to extend them. *Journal of Accounting Research* (Autumn), 353–385.

Ohlson, J. (1995). Earnings, book values and dividends in security valuation. *Contemporary Accounting Research*, *11*, 661–687.

Pakes, A. (1985). On patents, R&D, and the stock market rate of return. *The Journal of Political Economy*, *93*(2), 390–409.

Rajgopal, S., Venkatachalam, M., & Kotha, S. (2003). The value relevance of network advantages: The case of e-commerce firms. *Journal of Accounting Research*, *41*, 135–162.

Shane, H. (1993). *Patent citations as an indicator of the value of intangible assets in the semiconductor industry*. Working Paper. University of Pennsylvania.

Shortridge, R. (2004). Market valuation of successful versus non-successful R&D efforts in the pharmaceutical industry. *Journal of Business Finance and Accounting*, *31*, 1301–1325.

Standard & Poor's Biotechnology Industry Survey. May (2002).

Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *Rand Journal of Economics*, *21*, 172–187.

Trueman, B., Wang, F., & Zhang, X. (2001). The eyeballs have it: Searching for the value in internet stocks. *Journal of Accounting Research*, *38*, 137–162.

Wallman, S. (1995). The future of accounting and disclosure in an evolving world: The need for dramatic change. *Accounting Horizons*, *9*, 81–91.

Wallman, S. (1996). The future of accounting and financial reporting part II: The colorized approach. *Accounting Horizons*, *10*, 138–148.

APPENDIX. MODEL DEVELOPMENT

To examine whether nonfinancial patent information supplement the information content of financial information in market valuation, I follow recent theoretical work on valuation models developed by Ohlson (1995) who modeled the market value of the firm as a function of book value, earnings, and other relevant information. Knowledge asset has a crucial role in biotech firms' value creation. Therefore, to apply the Ohlson method in the biotech setting, this paper includes knowledge asset as other relevant information along with accounting data, such as book value and earnings, in the valuation model.

For a biotech firm, the largest and most important components of knowledge asset are its R&D expenditures and the discoveries made by its R&D activities. When successfully combined, these intangibles produce the intellectual property and legal patents that can rapidly translate into annual sales, profits, and/or large equity market value (Hand, 2001). Because cumulative R&D expenditures and cumulative patent information measure the inventive input and inventive output that closely tie to a biotech firm's future earning potential, they are used to proxy for the knowledge asset in the biotech industry.

A biotech firm's market value can then be modeled as a function of book value, earnings, cumulative R&D spending, and cumulative patent information. That is, a biotech firm's market value at the end of year t can be written as:

$$\begin{aligned}\mathrm{MV}_t = f(&\mathrm{BV}_t, \mathrm{E}_t, \mathrm{R\&D}_t,\ \mathrm{R\&D}_{t-1}, \ldots,\ \mathrm{R\&D}_{t-m},\\ &\mathrm{PATENT}_t,\ \mathrm{PATENT}_{t-1}, \ldots,\ \mathrm{PATENT}_{t-n}),\end{aligned} \tag{A1}$$

where MV is market value, BV is book value, E represents earnings, m the number of years in the economic life of R&D spending, and n the number of years in the economic life of patents. Three problems arise in estimating Eq. (A1), however. First, the appropriate economic life of R&D and patent information (i.e., m and n in Eq. (A1)) in the biotechnology industry are unknown, and prior research on the lagged effects of R&D on patents are inconclusive. Second, $\mathrm{R\&D}_t$, $\mathrm{R\&D}_{t-1}, \ldots$, $\mathrm{R\&D}_{t-m}$ and PATENT_t, $\mathrm{PATENT}_{t-1}, \ldots$, PATENT_{t-n} tend to move together and may result in multicollinearity problem. Third, when using time-series data over a given period, each lag included causes the loss of one data point. To avoid these problems, this study chooses to use a return model.

To derive the return model, first, express the market price at the end of year $t-1$ in functional form as follows:

$$\begin{aligned}\mathrm{MV}_{t-1} = f(&\mathrm{BV}_{t-1},\ \mathrm{E}_{t-1},\ \mathrm{R\&D}_{t-1},\ \mathrm{R\&D}_{t-2},\ldots,\mathrm{R\&D}_{t-m-1},\\ &\mathrm{PATENT}_{t-1},\ \mathrm{PATENT}_{t-2},\ldots,\ \mathrm{PATENT}_{t-n-1}).\end{aligned} \tag{A2}$$

Subtracting Eq. (A2) from Eq. (A1) and deflating both sides by MV_{t-1} yields the following specified form:

$$\begin{aligned}\frac{\Delta\mathrm{MV}_t}{\mathrm{MV}_{t-1}} &= \alpha_0 + \alpha_1\frac{E_t}{\mathrm{MV}_{t-1}} + \alpha_2\frac{\Delta E_t}{\mathrm{MV}_{t-1}}\\ &\quad + \alpha_3\frac{\mathrm{R\&D}_t}{\mathrm{MV}_{t-1}} + \alpha_4\frac{\mathrm{PATENT}_t}{\mathrm{MV}_{t-1}} + \frac{\varepsilon_{jt}}{\mathrm{MV}_{t-1}}\end{aligned} \tag{A3}$$

where E_t and ΔE_t are earnings and change in earnings in year t, respectively. The change in BV from year $t-1$ to year t was replaced by earnings in year t (E_t) because of the change in book value equals earnings, assuming no dividends. The terms $\mathrm{R\&D}_{t-m-1}$ and PATENT_{t-n-1} are omitted in Eq. (A3) because, as time passes, the knowledge asset depreciates, and the older R&D investments and patents become less valuable. The prior years' R&D expenditures and patent measures cancel out, and the lagged effects of R&D on patents are then eliminated. Compared with Eq. (A1), the return model does not require use of the economic life of R&D and patent information nor the lagged effects of R&D on patents. In addition, the return model provides evidence regarding the timeliness of investors' use of financial and nonfinancial patent information (Easton, 1999). Therefore, this study will use the return model as a baseline model.