

THE COMPLEXITY OF QUALITATIVE ACCOUNTING DISCLOSURES:
MANAGERS' CHOICES AND INVESTORS' REACTIONS

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Kristina Marie Rennekamp

May 2012

© 2012 Kristina Marie Rennekamp

THE COMPLEXITY OF QUALITATIVE ACCOUNTING DISCLOSURES:
MANAGERS' CHOICES AND INVESTORS' REACTIONS

Kristina Marie Rennekamp, Ph. D.

Cornell University 2012

This dissertation presents two studies on the complexity of qualitative accounting disclosures. Reporting is considered to be more complex as it becomes longer and/or less readable.

The first study examines investors' reactions to disclosure readability. The SEC's emphasis on the use of plain English is designed to make disclosures more readable and more informative. Using an experiment, I find that more readable disclosures lead to stronger reactions from small investors, so that changes in valuation judgments are more positive when news is good and more negative when news is bad. Drawing on research in psychology to explain this result, I predict and find that processing fluency from a more readable disclosure acts as a subconscious heuristic cue and increases investors' beliefs that they can rely on the disclosure. While I do not find that more readable disclosures directly increase perceptions of management credibility, I do find evidence of an indirect effect operating through feelings of processing fluency. In supplemental analyses, I find that investors who receive more readable disclosures revise their valuation judgments to be less extreme when they are explicitly made aware of the potential for variation in readability. I discuss potential explanations for these revised valuation judgments.

The second study examines managers' reporting choices with respect to disclosure complexity. Prior research finds that the reporting complexity of qualitative disclosures increases as firm performance deteriorates, and that reporting complexity may dampen investors' short-term reactions to news. While some argue that managers intentionally increase reporting complexity to hide bad performance, others suggest that bad performance is instead simply more difficult to describe in fewer and more readable words. I use a controlled experiment to investigate these explanations, and find greater support for the idea that bad performance is inherently more difficult to describe than good performance. Counter to arguments made in the archival literature, I find that individuals provide longer, but *more* readable, disclosures when performance is bad and they are given a goal of reporting strategically rather than reporting accurately. This contradicts the claim that managers intentionally obfuscate poor performance in order to mitigate negative reactions to bad news. My study also provides evidence in a controlled setting on how other linguistic choices vary with performance and goals.

BIOGRAPHICAL SKETCH

Kristi Rennekamp received a B.B.A. in Finance from the University of Iowa in 2002. She worked as a Financial Analyst for the United States Department of the Interior and completed the three levels of the Chartered Financial Analyst exams before returning to Iowa for graduate studies. She received an M.B.A. in 2006 from the University of Iowa, with concentrations in Finance and Accounting. She received an M.S. in Management from Cornell University in 2010, and a Ph.D. in Management from Cornell University in 2012. As of August 2011, Kristi is an Assistant Professor of Accountancy at the University of Illinois at Urbana-Champaign. She teaches in the area of Accounting Measurement and Disclosure. Her research examines the judgment and decision-making of managers, investors, and other financial reporting intermediaries.

Outside of the office, Kristi enjoys spending time traveling, hiking, kayaking and cooking with her husband, Michael.

TO THOSE WHO MADE THIS POSSIBLE:

MY PARENTS, ROSE AND STEVE, AND MY BROTHER, BRYAN

for being my first and best teachers

MY ADVISER, BOB LIBBY

for his patience, guidance, and sense of humor

MY HUSBAND, MICHAEL

for his love and support through everything

ACKNOWLEDGMENTS

I thank my dissertation committee members: Robert Libby (Chair), Robert Bloomfield, Jay Russo, and Steven Schwager for their valuable guidance and advice. I also thank Chris Agoglia, Scott Asay, Sanjeev Bhojraj, Tim Brown, Brooke Elliott, Jeffrey Hales, Vicky Hoffman, Kevin Jackson, Karim Jamal, Steve Kachelmeier, Bill Kinney, Lisa Koonce, Tamara Lambert, Don Moser, Mark Nelson, Mark Peecher, David Piercey, Kathy Rugar, Nick Seybert, Abbie Smith, Shankar Venkataraman, Michael Williamson, Holly Yang, an anonymous reviewer, and workshop participants at the University of Alberta, University of Chicago, Cornell University, Georgia Institute of Technology, University of Illinois at Urbana-Champaign, University of Iowa, University of Massachusetts-Amherst, University of Pittsburgh and the University of Texas at Austin for helpful comments and discussions. I gratefully acknowledge financial support from Cornell University's Johnson Graduate School of Management, the University of Illinois at Urbana-Champaign, and the Deloitte Foundation. I am indebted to Ramji Balakrishnan, Dan Collins, Gary Gaeth and Tom Gruca for encouraging me to enter a doctoral program.

TABLE OF CONTENTS

Biographical Sketch.....	iii
Dedications	iv
Acknowledgements	v
Table of Contents	vi
List of Figures.....	ix
List of Tables	x
Chapter 1: Processing Fluency and Investors’ Reactions to Disclosure Readability.....	1
1.1 Introduction	1
1.2 Background and Development of Hypotheses	7
1.2.1 The Effect of Disclosure Readability on Investors.....	7
1.2.2 The Role of Processing Fluency	9
1.3 Experiment	14
1.3.1 Participants	14
1.3.2 Design.....	17
1.3.3 Manipulations	18
1.3.4 Task and Procedure	19
1.3.5 Primary Dependent Variables.....	21
1.4 Results	22
1.4.1 Effects of Readability on Valuation Judgments	22

1.4.2 Effects of Readability on Processing Fluency and Reliance on a Disclosure	26
1.4.3 Effects of Readability on Judgments of Management Credibility	29
1.4.4 Supplemental Analyses	32
1.5 Conclusion	36
Chapter 2: The Influence of Performance and Reporting Goals on Managers' Choice of Reporting Complexity in Disclosures	42
2.1 Introduction	42
2.2 Background and Development of Hypotheses	47
2.2.1 Strategic Obfuscation by Management	47
2.2.2 The Inherent Complexity of Explaining Poor Performance	50
2.3 Experiment	54
2.3.1 Participants	54
2.3.2 Design and Manipulations	54
2.3.3 Task and Procedure	55
2.3.4 Primary Dependent Variables	58
2.4 Results	61
2.4.1 The Main Effect of Performance on Reporting Complexity	61
2.4.2 The Interaction between Performance and Goals on Reporting Complexity	70

2.4.3 Supplemental Analyses	74
2.5 Conclusion.....	75
Appendix A: Examples of <i>Plain English Handbook</i> Suggestions	78
Appendix B: Press Releases from More Readable Conditions	80
Appendix C: Press Releases from Less Readable Conditions.....	82
References	84

LIST OF FIGURES

Figure 1: Sources of Financial Information for Participants	17
Figure 2: 2x2 Between-Subjects Design	18
Figure 3: Timeline of Events in Stages 1 and 2 of the Experiment.....	18
Figure 4: Plot of News x Readability Interaction on Change in Valuation Judgments – Stage 1	25
Figure 5: Mediation Analyses for Testing H2	28
Figure 6: Plot of News x Readability Interaction on Change in Valuation Judgments – Stage 2	36
Figure 7: Expected Pattern of Results for the Length Measure of Reporting Complexity	53
Figure 8: Expected Pattern of Results for the Readability Measures of Reporting Complexity	53
Figure 9: Screen Shot Showing the Four Ambiguous Facts Provided to Participants for the Preparation of their Report	57
Figure 10: Sample Report from the Bottom 10% (Most Readable) of both the FOG and Flesch-Kincaid Measures	67
Figure 11: Sample Report from the Top 10% (Least Readable) of both the FOG and Flesch-Kincaid Measures.....	67
Figure 12: Actual Pattern of Results for the Primary Dependent Variables	72

LIST OF TABLES

Table 1: Descriptive Statistics and Analysis of Variance – Valuation Judgments.....	23
Table 2: Descriptive Statistics and Two-Sample T-Tests – Credibility Judgments	30
Table 3: Descriptive Statistics, T-Tests and Tests for Simple Main Effects of Readability on Stage 2 Valuation Judgments	34
Table 4: Descriptive Statistics and Analysis of Variance – Report Length	62
Table 5: Descriptive Statistics and Analysis of Variance – FOG Measure of Readability	64
Table 6: Descriptive Statistics and Analysis of Variance – Flesch-Kincaid Measure of Readability	65
Table 7: Descriptive Statistics and Analysis of Variance – BOG Measure of Readability	69
Table 8: Descriptive Statistics and Analysis of Variance – READscore Measure of Readability	70

CHAPTER 1
PROCESSING FLUENCY AND INVESTORS' REACTIONS TO DISCLOSURE
READABILITY

1.1 Introduction

The Securities and Exchange Commission has placed increasing emphasis on providing clear and readable financial disclosures. The Plain English Rule (421(d)), passed in 1998, requires that issuers adhere to plain English principles in the design of firm prospectuses. The rule is accompanied by a *Plain English Handbook* (SEC [1998]) that provides both linguistic and formatting suggestions for preparing plain English disclosures. In its *Plain English Handbook*, the SEC encourages firms to adopt the suggestions in all of their communications (SEC [1998]).

In response to the SEC's agenda, recent work in the archival accounting literature uses techniques from computational linguistics to investigate disclosure readability (Li [2008], You and Zhang [2009]) and its effects on the behavior of small investors (Miller [2010]). These archival studies argue that less readable disclosures may limit investors' ability or willingness to extract information from relatively long and complex documents, such as 10-K filings (Grossman and Stiglitz [1980], Bloomfield [2002], Hirshleifer and Teoh [2003]). I use a controlled experiment to provide complementary evidence on a different effect of disclosure readability that may not be as intuitive to managers, investors and regulators. My findings show that disclosure readability can affect investors' reactions, even when it does not affect the amount of information that they acquire.

Research in the psychology literature suggests that processing fluency from a more readable disclosure will increase investors' beliefs that they can rely on the

information in the disclosure. Processing fluency is an individual's subjective feeling about how easy it is for them to process information. Prior research has shown that feelings of processing fluency are subconsciously treated by individuals as a heuristic cue that information can be relied upon in making related judgments (Shah and Oppenheimer [2007], Hafner and Stapel [2010]). My experiment holds disclosure length and total information constant, and isolates the effects of disclosure readability on investors' reactions to both good and bad news. I focus on small investors in my study, both to be consistent with prior archival literature (Miller [2010]) and also because of the SEC's assertion that clear writing is primarily intended to assist the "least-sophisticated investors" (SEC [1998]). The use of an experiment allows me to capture measures of processing fluency and disclosure reliance that are not directly observable in archival data.

Prior research on processing fluency also suggests that a fluent message can lead to more favorable evaluations of the messenger (Oppenheimer [2006]). This implies that more readable disclosures may increase investors' perceptions of management credibility, regardless of whether a disclosure conveys good or bad news. My experiment allows me to isolate judgments about the manager from judgments about the firm, and to tie my research to the larger literature in accounting on the determinants of management credibility.

Finally, the use of an experiment allows me to provide evidence on whether investors' judgments change when the potential for variation in disclosure readability is made salient. I use a two-stage experiment to examine whether investors' valuation judgments change once they explicitly consider that the same information could have been presented in a more or less readable way.

In Stage 1 of my experiment I collect judgments on my primary measures of interest. I present 234 participants with background information on a fictitious soft drink company and ask them to provide initial valuation judgments. Participants then receive a press release about the financial performance of the firm. Between-subjects manipulations alter the readability of the press release and whether the press release conveys good or bad news. For the manipulation of readability, I vary characteristics of the press release so that they violate (less readable condition) or conform to (more readable condition) linguistic and formatting suggestions outlined in the SEC's *Plain English Handbook* (SEC [1998]). After reading the release, participants provide a revised judgment on the appropriate valuation of the firm. Participants also provide judgments about the extent to which they feel that they can rely on the information in the press release, their feelings about processing fluency, and their perceptions of management credibility. The difference between the revised and initial valuation judgments serves as a measure of the strength of investors' reactions to the press release containing my manipulations.

As predicted by theories of processing fluency, I find that more readable disclosures lead to stronger reactions to both good and bad news. Changes in participants' valuation judgments in response to a more readable press release are more positive when news is good and more negative when news is bad. These results arise despite the fact that, in my setting, variation in readability does not lead to detectable differences in the number of questions that participants answer correctly about the information in the press release. This suggests that my results are not driven by differences in the amount of information that participants acquire. Instead, greater readability increases participants' reliance on the information in the press release.

Mediation analysis confirms that participants' reliance on the release is driven by increased processing fluency (i.e., participants' personal feelings about disclosure readability, or processing ease). Counter to my expectations, I do not find that more readable press releases directly increase participants' judgments about the credibility of management. However, I do find evidence that readability has a positive and significant indirect effect on management credibility, and that this indirect effect operates through increased feelings of processing fluency. A significant indirect effect in the absence of a direct or total effect could mean that there are one or more offsetting indirect effects that work in the opposite direction of fluency (MacKinnon, Krull and Lockwood [2000], Shrout and Bolger [2002], Zhao, Lynch Jr., and Chen [2010]), or that the sample size is not large enough to detect a moderate direct or total effect, should it exist (Shrout and Bolger [2002]).

Stage 2 of my experiment provides supplementary evidence on whether investors' valuation judgments change once they explicitly consider the potential for variation in disclosure readability. In Stage 2, I present participants with the press release that they were initially shown, as well as an additional press release from the other readability condition (still containing only good or bad news). The press releases are presented side-by-side to facilitate their comparison. Providing access to press releases at both levels of readability makes salient the fact that managers have a choice when it comes to disclosure readability.

Analysis of Stage 2 valuation judgments shows that participants who initially received the more readable disclosures provide valuation judgments in Stage 2 that are less extreme than the valuation judgments they provide in Stage 1, to the point where the effect of readability on investors' reactions is no longer significant for either good

or bad news. In addition, participants strongly agree that the actual information contained in the more and less readable press releases is identical. Combined, these results suggest that the stronger reactions of participants found in Stage 1 in response to more readable disclosures may be unintentional (Kahneman and Tversky [1996], Tan, Libby, and Hunton [2002]), which is consistent with findings in psychology showing that individuals are not influenced by processing fluency once they are aware of its source (Schwarz [2004], Alter and Oppenheimer [2009]). It is also possible that revisions to valuation judgments in Stage 2 are driven by consideration of managers' strategic choices when it comes to disclosure readability. I examine both possibilities in supplemental analyses included in Section 1.4.4. My results provide more support for the former explanation than the latter.

My results call into question the idea that more readable disclosures are always beneficial. The fact that participants who received the more readable press releases in Stage 1 provide less extreme valuation judgments in Stage 2 suggests that, upon further consideration, participants feel as though they reacted too strongly to the information that was presented in a more attractive package (i.e., with clearer formatting and language). In other words, more readable disclosures may cause investors to overreact to information, particularly those who are the least sophisticated. This does not imply that more readable disclosures are necessarily bad, but it does indicate that the benefits of more readable disclosures are less clear-cut than has been argued by the SEC [SEC 1998] and in prior literature (e.g., Li [2008], You and Zhang [2009], Miller [2010]).

My findings also suggest that readability will affect investors even if it does not affect their willingness or ability to acquire information from long and complex

financial disclosures (e.g., Li [2008], You and Zhang [2009], Miller [2010]). I find that the clarity with which information is conveyed will affect readers' feelings of processing fluency, and have important (and perhaps unintentional) consequences for related judgments and decisions. Thus, despite the focus on the readability of 10-K filings in the prior literature, my results suggest that it is also important to consider the readability of shorter disclosures as well (like press releases).

My results are likely to generalize beyond press releases. SEC filings and verbal (e.g., conference call) disclosures vary substantially in the ease with which they can be processed. In addition, subjective perceptions of processing fluency are likely to vary with knowledge and experience, leading to potential differences in how experienced and inexperienced individuals react to a given disclosure. My findings also suggest that other parties (e.g., auditors, lenders, employees within the firm, etc.) may be influenced by characteristics affecting the processing fluency of the information presented to them.

My study adds to the growing body of literature investigating the style of disclosures as opposed to their content. Content is the literal meaning of the information conveyed, or the concrete facts contained in a disclosure, whereas style captures the methods used to convey meaning to the audience (through, for example, optimistic vs. pessimistic tone, vividness, vocal cues, etc.). Experiments are well-suited to studying disclosure style, but research in the area is relatively sparse. At least one exception is Hales, Kuang and Venkataraman [2011], which uses an experiment to investigate the effects of vivid vs. pallid language, above and beyond actual information content. Future experimental research could provide additional evidence to complement the archival literature by investigating how specific stylistic choices

affect investors' judgments, and how managers make stylistic choices in the first place.

The rest of this chapter proceeds as follows. Section 1.2 provides background information and develops my hypotheses. Sections 1.3 and 1.4 discuss my experiment and results, respectively. Section 1.5 concludes.

1.2 Background and Development of Hypotheses

1.2.1 The Effect of Disclosure Readability on Investors

To address concerns about disclosure readability, the SEC's Plain English Rule (421(d)) went into effect on October 1st, 1998, and requires that issuers adhere to plain English principles in the design of firm prospectuses.¹ Documentation surrounding the rule's release encourages the use of plain English in all disclosure documents, and argues that plain English allows companies to communicate and build relationships with their investors (SEC [1998]). More recently, the SEC has expanded plain English rules. Rules 13a-20 and 15d-20 were adopted in 2006 and amend the 1934 Securities Exchange Act to require plain English in disclosures about issues related to corporate governance, related-party transactions, executive and director compensation, and beneficial ownership. In March of 2009, the SEC began requiring that mutual funds provide a plain English summary prospectus of key investment information at the

¹ In general, definitions of "readability" in prior literature are imprecise, given that the concept of readability has evolved primarily through the evaluation of grade-school textbooks (see Dubay [2004]). Some studies use disclosure length as a measure of readability (e.g., Li [2008], You and Zhang [2009], Miller [2010], Loughran and McDonald [2011]). Other studies have used measures like the Fog Index or Flesch Reading Ease Score (both of which calculate readability based on sentence length and syllable counts) (e.g., Li [2008], Biddle, Hilary and Verdi [2009], Miller [2010], Lehavy, Li and Merkley [2011], Loughran and McDonald [2011]). However, Brennan, Guillamon-Saorin and Pierce [2009] argue for more precise measures of readability in business communications. As a result, my study manipulates readability through linguistic and formatting choices, as outlined in the SEC's *Plain English Handbook* [SEC 1998].

front of their full prospectus.² In July of 2010, the SEC voted unanimously to amend the Investment Advisers Act to require plain English in Form ADV Part 2, the narrative brochure that SEC-registered investment advisers are required to provide to current and prospective clients.

In its *Plain English Handbook*, the SEC outlines a number of practical linguistic and formatting suggestions for preparing plain English disclosures (SEC [1998]). Both linguistic and formatting choices can affect the extent to which investors find a disclosure to be readable. Linguistic suggestions relate to the choice of words and how they are organized. Formatting suggestions relate to the use of features like bullet points, tables and line spacing, which increase the legibility of a disclosure (see Appendix A for examples).

There is evidence in the archival literature to support the idea that disclosure readability affects investors. You and Zhang [2009] show that market underreaction to 10-K filings is more severe as the length of the report (and presumably its complexity) increases. Miller [2010] uses linguistic suggestions in the *Plain English Handbook* to develop a measure of 10-K readability, and provides evidence on the relationship between enhanced readability of 10-K disclosures and the trading behavior of small investors. Miller [2010] finds that more readable disclosures are associated with greater trading activity among small investors (defined as those with trades below \$5000) around the 10-K filing date. This holds even after controlling for information content, firm performance and earnings persistence. Miller [2010] also finds that 10-K complexity reduces consensus among small investors, but not large investors (defined

² See Release No. 33-8998.

as those with trades above \$50,000).³

It is possible that the trading volume of small investors is affected by disclosure readability because a less readable disclosure limits the amount of information that they are willing or able to extract (Grossman and Stiglitz [1980], Bloomfield [2002], Hirshleifer and Teoh [2003]). However, I focus on a separate effect of readability that is likely to be less intuitive to managers, investors and regulators. Research in the psychology literature on processing fluency suggests that even if less readable disclosures do not limit investors' ability to acquire information in the first place, disclosure readability can still affect investors' judgments and decisions. More specifically, enhanced processing fluency from a more readable disclosure may increase investors' perceptions that they can rely on the information in the disclosure.

1.2.2 The Role of Processing Fluency

Processing fluency is subjective, and represents how easy it feels to process information. Prior research has manipulated processing fluency in a variety of ways. Studies have altered visual processing fluency by manipulating whether materials are presented in a difficult- or easy-to-read font (e.g., **Haettenschweiler** vs. Cambria) or by manipulating font size and color (see, e.g., Novemsky, Dhar, Schwarz and Simonson [2007]). Other studies have manipulated linguistic processing fluency through, for example, rhyming (McGlone and Tofiqbakhsh [2000]), the use of easy- or hard-to-pronounce words (Alter and Oppenheimer [2006]) or the use of simple vs. complex

³ Although not the focus of my study, there is also evidence to suggest that disclosure readability affects analysts. Leavy, Li and Merkley [2011] find that firms with less-readable 10-Ks have greater analyst following, suggesting that less readable disclosures increase the need for information intermediaries. However, less readable disclosures also appear to increase the effort expended by analysts, as measured by the time it takes them to provide revised forecasts. Loughran and McDonald [2011] find that less readable 10-Ks are associated with greater dispersion in analysts' forecasts.

synonyms (Oppenheimer [2006]).

Despite the range of techniques that have been used to manipulate processing fluency, the corresponding responses from individuals are remarkably similar across different settings. In general, individuals like messages that feel easy to process. Because individuals act as though feelings experienced while thinking about a target bear on the target itself, the feelings of subjective ease associated with processing fluency are typically treated as a subconscious cue that something about the message or messenger is good. Processing fluency has been associated with higher ratings of truth, preference for the message and the messenger, willingness to rely on information, and confidence in judgments (see Alter and Oppenheimer [2009] for a review).⁴ For example, McGlone and Tofiqbakhsh [2000] find that individuals experience higher processing fluency and judge truthfulness to be higher with rhyming aphorisms than with informationally-equivalent non-rhyming aphorisms (e.g., “What sobriety conceals, alcohol reveals” vs. “What sobriety conceals, alcohol unmasks”). Alter and Oppenheimer [2006] find that, in the short term, stocks with names and ticker codes that are easier to pronounce outperform those that are more difficult to pronounce.

With respect to disclosure readability and processing fluency, the *Plain English Handbook* suggests that readability affects processing ease by arguing that its plain English suggestions correspond with how individuals naturally process information (SEC [1998]). Greater readability in a disclosure should therefore increase

⁴ However there are some exceptions showing that the meaning of processing fluency can be affected by context (Schwarz [2004]). Briñol, Petty and Tormala [2006] find that greater processing fluency can lead to more negative evaluations if ease is framed negatively. Similarly, Labroo and Kim [2009] find that individuals with explicit goals evaluate objects associated with pursuing their goals more negatively when they are high in processing fluency, presumably because individuals adopt a heuristic belief that achieving something of value will require greater effort (and thus ease is viewed negatively).

feelings of processing fluency and increase investors' beliefs that they can rely on the disclosure. Consistent with this prediction, Shah and Oppenheimer [2007] find that when both more and less fluent information is presented, individuals weight the more fluent information more heavily in their judgments. Similarly, Hafner and Stapel [2010] argue that feelings of processing fluency serve as a cue regarding the usability of information.

If investors increase their reliance on disclosures that are more readable, then more readable disclosures should also lead to stronger reactions to the news contained in those disclosures. Miller [2010] finds positive associations between disclosure readability and the trading behavior of small investors, but also finds that length subsumes the effects of readability from the use of plain English. However, archival studies are unable to hold constant firm characteristics or the amount of information in the disclosure. Thus it is still an open question whether disclosure readability from the use of plain English affects small investors above and beyond disclosure length. Furthermore, Miller [2010] looks at small investors' responses to overall readability, but does not separately investigate responses to good and bad news. It is important to understand how investors react to bad news disclosures that are low in readability given prior claims that managers strategically obfuscate negative news [Li 2008].

In Stage 1 of my experiment, participants receive a single disclosure that contains either good or bad news and is either more or less readable. The use of an experiment allows me to hold disclosure length and total information constant while varying disclosure readability, thereby isolating the effects of disclosure readability on small investors' reactions to both good and bad news. I measure investors' reactions by capturing how their valuation judgments change in response to more or less

readable disclosures. My first hypothesis is:

H1: More readable disclosures lead to more positive changes in investors' valuation judgments when news is good, but more negative changes when news is bad.

As discussed above, I do not expect the stronger reactions to more readable disclosures in my study to be driven by differences in the actual information that investors acquire from the press release, but instead by how processing fluency subconsciously affects their belief that they can rely on the disclosure. Existing techniques in the archival literature cannot address processing fluency as a mechanism through which readability influences investors. An experiment allows me to elicit processing fluency and reliance measures that are not available in archival data. Holding disclosure length and information constant, my second hypothesis is:

H2: More readable disclosures increase investors' reliance on the information in the disclosure, and this effect operates through increased feelings of processing fluency.

Processing fluency from a more readable disclosure may also lead to more favorable evaluations of management, even when a disclosure conveys negative news. This is consistent with prior literature finding that fluent messages are judged as coming from a more likeable or intelligent source (Oppenheimer [2006]). Furthermore, if processing fluency from a more readable disclosure increases investors' willingness to rely on the information (Shah and Oppenheimer [2007], Hafner and Stapel [2010]), this may carry over to feelings of being able to rely on the manager as well.

Instead of asking for ratings of intelligence, likeability, or willingness to rely

on the manager, I focus on attitudes about managers' credibility in order to be consistent with prior work in accounting (see e.g., Mercer [2005], Barton and Mercer [2005], Venkataraman [2008], Clor-Proell [2009], Koonce and Lipe [2010], Rupa [2011], Tan, Wang and Zhou [2011]).⁵ H3 focuses on evaluations of management credibility, and predicts:

H3: Investors' judgments of management credibility will be higher when news is more readable, regardless of whether the disclosure conveys good or bad news about the firm.

My experiment identifies a circumstance where processing fluency is expected to simultaneously decrease one favorability judgment when news is bad (valuation) but increase another (management credibility). However, it is also possible that processing fluency from more readable negative news, and the resulting negative reaction predicted by H1, will carry over into negative impressions of management. This could lead to lower evaluations of management credibility when negative news is more readable, contrary to my prediction in H3.

To date, only a few studies in the accounting literature have relied on theories related to processing fluency. Research suggests that fluency feelings can arise from processing new, external information, but also from the ease of generating thoughts and accessing past memories (Novemsky et al. [2007], Alter and Oppenheimer [2009]). In accounting, Sedor [2002] shows that disclosures provided in "scenario

⁵ As discussed by Mercer [2004], managers' credibility is distinct from disclosure credibility, although the two are related. Substantial work in the archival literature has investigated disclosure credibility, particularly with respect to management forecasts. However, these studies typically do not distinguish between the credibility of the firm and the credibility of management, or investigate whether managers benefit from developing a personal reputation for credibility. One recent exception is Yang [2012], which finds that manager-specific forecasting reputations are particularly important when forecast news is extreme or analyst dispersion is high.

form” (rather than a list) lead to unintentional optimism among analysts. Both Sedor [2002] and Kadous, Krische and Sedor [2006] argue that information presented in scenario form reduces cognitive effort and facilitates analysts’ ability to envision how managers’ plans will be achieved. In turn, the ease of envisioning the means of achieving the outcome increases analyst optimism about future performance.⁶ Kadous et al. [2006] show that analyst optimism is reduced when analysts are asked to generate few rather than many counter-explanations as to why managers’ plans might fail. Drawing on the work of Heiman [1990], Koonce [1992] and Kennedy [1995] on the use of counter-explanations and Schwarz et al. [1991] on the availability heuristic, Kadous et al. [2006] argue that their results are driven by the fact that analysts find the generation of few counter-explanations to be subjectively easy, but the generation of many counter-explanations to be difficult. This subjective ease (or processing fluency) serves as a heuristic cue affecting judgments. A more recent paper by Koonce and Lipe [2010] also touches on theories of processing fluency, but in a different context. They find that when information about both earnings trend and earnings performance is available, investors react to each measure only when it is consistent over time (e.g., consistently increasing earnings or consistently meeting performance benchmarks). Koonce and Lipe [2010] argue that inconsistent measures are more difficult to process, and are therefore ignored when consistent information is available.

1.3 Experiment

1.3.1 Participants

Participants are 234 individuals recruited from Amazon’s Mechanical Turk

⁶ Sedor [2002] emphasizes the effects of connecting ideas in a causal narrative, whereas my study holds constant the causal narrative that is actually provided by managers, but manipulates how easy it is to process the story that it conveyed.

platform in exchange for \$0.75. Launched in 2005, Amazon's Mechanical Turk (AMT) is an internet labor market that allows "Requesters" to pay individuals to complete "Human Intelligence Tasks" (HITs). AMT is an increasingly popular source of experimental data for social scientists because the AMT subject pool is large, readily accessible, and at least as representative of the U.S. population as more traditional subject pools (Paolacci, Chandler and Ipeirotis [2010]). Furthermore, studies run on AMT have been shown to reliably replicate a wide range of prior JDM findings (Paolacci et al. [2010], Horton, Rand and Zeckhauser [2011]). Most participants take approximately 12 minutes to complete the study, meaning that their effective hourly wage for participating is \$3.75/hour. This is well above the median reservation wage reported in Horton and Chilton [2010] of approximately \$1.38/hour. These relatively low wages highlight the fact that laboratory experiments are more costly to run than online experiments in part because participants demand additional compensation for physically showing up to the lab (Bloomfield and Rennekamp [2009]).

Libby, Bloomfield and Nelson [2002] suggest that the appropriateness of a particular group of participants can be judged based on whether their knowledge is sufficient for the task. As a baseline requirement, I specifically recruit participants who (1) live in the United States and (2) consider English to be their native language. I collect additional background information on participants to support their use in my experiment. The average participant is 34.13 years old, with an average of 13.29 years of full-time work experience. Participants have completed an average of 1.23 accounting and 1.12 finance courses. Overall, 89.74% of participants indicate that they have at least some experience with investing, with 45.73% saying that they have

invested in individual stocks in the past and 66.24% saying that they plan on investing in individual stocks in the future.⁷ This particular sample of participants should have sufficient knowledge to act as small investors, read a press release and provide simple judgments of a firm and its management.⁸

My focus on small investors is appropriate because Miller [2010] finds that small (but not large) investors are influenced by the readability of disclosures. Furthermore, the SEC emphasizes that clear writing is primarily intended to assist the “least-sophisticated investors” (SEC [1998]). Figure 1 presents the sources that participants indicate using for their investment information. The most common source of investment information (with 43.59% of participants indicating it as a source) is company websites. This suggests that companies may have a substantial amount of control over the presentation of at least some of the information that is accessed by the investing public.

⁷ Elliott, Hodge, Kennedy and Pronk [2007] investigate whether, and when, MBA students are appropriate proxies for nonprofessional investors. They conclude that first-year MBA students are appropriate for tasks that are low in integrative complexity and require little specialized knowledge. My participants have demographic characteristics similar to the first-year MBA students used in Elliott et al.’s [2007] studies. The “early-MBA” participants from their first (second) experiment have an average of 5.2 (5.8) years of full-time work experience, have completed an average of 1.8 (1.6) accounting courses, 1.0 (0.6) finance courses, and 52% (52%) have invested in individual stocks or debt securities in the past.

⁸ AMT provides additional quality-control features beyond allowing researchers to screen for certain types of participants. For instance, unique worker ID’s associated with each HIT and the collection of IP addresses mitigates the likelihood that participants complete the task more than once. I confirm that none of the 234 participants in my study have identical IP addresses or worker ID’s. Researchers also have the ability to deny payment to participants for poor quality work (such as failing too many attention checks). Thus, AMT overcomes some of the potential concerns with conducting accounting experiments online that are raised by Bloomfield and Rennekamp [2009].

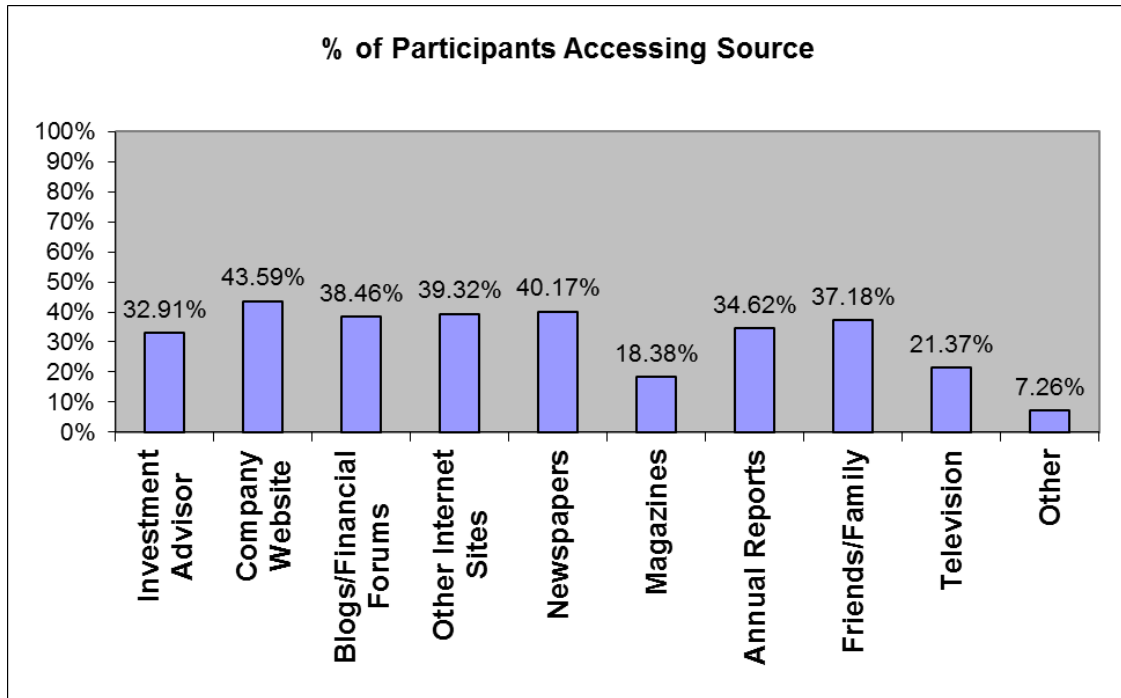


Figure 1. Sources of Financial Information for Participants

1.3.2 Design

My experiment uses a 2x2 between-subjects design manipulating whether an abbreviated press release for a fictitious soft drink company (1) conveys good or bad news about the firm, and (2) is more or less readable (see Figure 2). The experiment is conducted in two stages (see Figure 3). In Stage 1, participants receive a single initial press release. In Stage 2, participants have access to the initial press release but also receive another press release (conveying the same news), but at the other level of disclosure readability. They are told to consider this second press release as an alternative way in which the firm could have presented the information, making salient the potential for variation in disclosure readability.

		Readability of Initial Press Release Reviewed	
		Less	More
News	Good	Good News/ Less Readable	Good News/ More Readable
	Bad	Bad News/ Less Readable	Bad News/ More Readable

Figure 2. 2x2 Between-Subjects Design

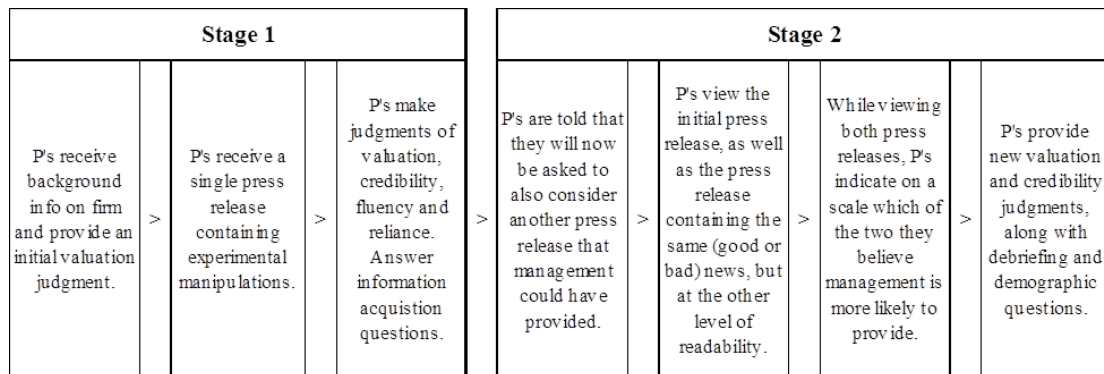


Figure 3. Timeline of Events in Stages 1 and 2 of the Experiment

1.3.3 Manipulations

Materials are adapted from those of a real soft drink company (Jones Soda) but are disguised in order to ensure that participants respond to my experimental materials rather than to other factors (for example, the reputation of the actual firm).

To manipulate readability, I start by considering the linguistic and formatting suggestions made in the SEC's *Plain English Handbook* [SEC 1998]. I then pick the

features that can be varied without changing information content.⁹ My more readable disclosures conform to the SEC's suggestions for the chosen features, whereas my less readable disclosures violate the SEC's suggestions. The specific linguistic features I manipulate are the use of short sentences; active voice; no hidden verbs; no superfluous words; language written in the positive; simple synonyms; personal pronouns; and sentences that keep 'subject', 'verb' and 'object' close together. The formatting features I manipulate are the use of clear headings/hierarchy; appropriate layout; tables; and bullet points. Panels A and B of Appendix A provide examples of the types of linguistic and formatting suggestions, respectively, that I use in my readability manipulations.

To manipulate news I hold revenue and net profit for the firm constant, but change performance in the same quarter for the prior year to alter whether the current quarter represents better (good news) or worse (bad news) performance. I also change the language in the disclosure where appropriate, so that the good news (bad news) disclosures discuss increases (decreases) in performance for the quarter and expectations of further improvement (deterioration) for the upcoming quarter and annual results. Panels A and B of Appendix B provide examples of the abbreviated press releases provided in my more readable conditions for good and bad news, respectively. Appendix C provides similar examples for my less readable conditions.

1.3.4 Task and Procedure

In Stage 1 of my experiment, participants begin by reading background information about the company and providing an initial judgment about the

⁹ Concrete Language and Jargon/Legalese are not included in my readability manipulation because they are difficult to manipulate without changing the information provided to participants.

appropriate common stock valuation for the firm. Following initial valuation judgments, participants are randomly presented with one of the four abbreviated press releases that contain my manipulations (see Appendices B and C). After participants review the press release, I ask them to indicate once again their beliefs about the appropriate common stock valuation of the firm, as well as provide their ratings of management competence and trustworthiness (the two components of credibility used in testing H3).¹⁰ After answering questions about firm valuation and management credibility, participants indicate the extent to which they felt that they could rely on the information in the press release and also provide ratings of their feelings about disclosure readability to capture processing fluency. Again, H2 predicts that a more readable disclosure will lead to higher ratings of processing fluency which, in turn, will increase participants' beliefs that they can rely on the information in the disclosure. Participants conclude Stage 1 of the task by answering three multiple-choice questions about the information in the press release. These are designed to rule out that differences in participants' responses are due to differences in information acquisition rather than differences in subjective perceptions associated with information processing.

In Stage 2 of my experiment, participants are told that they will also be asked to consider another press release that management could have provided (in addition to the one that they initially received). The second press release that they receive contains the same news (good or bad), but is presented at the other level of readability. This presentation is intended to make salient to participants that managers can vary the

¹⁰ I randomly assign participants to evaluate either firm valuation or management credibility first. Ordering of questions does not interact with other manipulations to influence my dependent variables, and is therefore not discussed further.

level of readability in their disclosures. To reinforce the difference between the two press releases and increase the likelihood that participants carefully examine them, participants rate the extent to which the actual information in the two disclosures is identical, rate which of the two disclosures is more readable, and indicate the extent to which they think management is more likely to choose one press release versus the other if the goal is to elicit a favorable reaction from the market.

In Stage 2 of the experiment, participants provide new valuation, competence and trustworthiness judgments on the same scales as in Stage 1. These judgments allow me to provide some supplemental evidence on whether responses in Stage 1 of the experiment appear to be intentional (that is, before the potential for variation in disclosure readability is made salient). Prior papers have successfully used similar designs to disentangle unintentional biases from intentional judgments (see Libby, Bloomfield and Nelson [2002] for a discussion).

1.3.5 Primary Dependent Variables

Change in Valuation Judgments. Following prior literature, I elicit valuation judgments by asking participants to indicate on a 101-point scale what they believe to be an appropriate common stock valuation for the firm, ranging from 0 = low to 100 = high (Koonce and Lipe [2010]). Recall that participants make a valuation judgment both before and after viewing the press release containing my manipulations. The actual measure I analyze in testing H1 is the difference between these two valuation judgments. This change measure captures the strength of investors' reactions to the disclosure.

Reliance on the Disclosure. I measure reliance through a direct question asking participants to indicate their agreement with the statement "I felt like I could

rely on the information in the press release” (1 = strongly disagree, 7 = strongly agree). I also investigate a measure that should reflect reliance, although it is likely to be a noisier proxy. My second measure of reliance is the absolute value of the change in valuation judgments discussed above. Greater reliance should be reflected as a larger value on this measure. I use both of these measures to provide evidence on H2.

Processing Fluency. To measure processing fluency, participants are asked how easy or difficult it *felt* to read the press release that they received (1 = very difficult, 7 = very easy). My measure therefore captures participants’ subjective feelings, and this measure serves as the mediating variable in my process analysis related to H2 (Oppenheimer [2006]).

Management Credibility. I use 101-point scales to capture participants’ judgments about the competence and trustworthiness of management, where 0 = very incompetent or very untrustworthy, and 100 = very competent or very trustworthy (Koonce and Lipe [2010]). Consistent with prior studies, I collapse these two ratings into a single measure of management credibility for testing H3.

1.4 Results

1.4.1 Effects of Readability on Valuation Judgments

My expectation from H1 is that more readable disclosures lead to more positive changes in valuation judgments when news is good, but more negative changes in valuation judgments when news is bad. Descriptive statistics (means, standard deviations and medians) are presented in Panel A of Table 1 for my variable capturing changes in valuation judgments.¹¹

¹¹ Throughout Section 1.4, results are presented based on 234 participants. Originally 250 participants were recruited. Of the 250 participants that completed the task through AMT, 15 erroneously participated even though they were not native English speakers, as requested in my recruiting materials.

Table 1. Descriptive Statistics and Analysis of Variance –Valuation Judgments**Panel A. Descriptive Statistics - Mean, (Standard Deviation), and [Median]**

News	Readability		
	Less	More	Overall
Good	18.153 (12.635) [20.000] n=59	23.500 (13.345) [22.000] n=62	20.893 (13.225) [20.000] n=121
Bad	-10.259 (11.371) [-11.000] n=54	-14.431 (14.357) [-15.000] n=58	-12.420 (13.113) [-13.500] n=112
Overall	4.575 (18.630) [10.000] n=113	5.167 (23.502) [10.500] n=120	

Panel B. Change in Valuation Judgments - Contrast-Coded Analysis of Variance

Source	S.S.	d.f.	M.S.	F-statistic	p-value
News x Readability	65829.425	1	65829.425	389.295	<0.001
Residual	65.485	2	32.743	0.194	0.824
Error	38723.720	229	169.099		

Panel C. Change in Valuation Judgments - Simple Main Effects of Readability

Source	S.S.	d.f.	M.S.	F-statistic	p-value
Good News	864.476	1	864.476	5.112	0.012
Bad News	486.682	1	486.682	2.878	0.046

Notes to Table 1. This table presents descriptive statistics, contrast-coded ANOVA, and simple main effects tests for changes in investors' valuation judgments after viewing a disclosure. The cells of the experiment receive contrast weights as follows: Good News/More Readable = +3, Bad News/More Readable = -3, Good News/Less Readable = +2, Bad News/Less Readable = -2. All p-values are one-tailed equivalents

One additional participant is excluded because a glitch in the Qualtrics software used to administer the task failed to show the page containing my manipulated press release. Finally, one participant provided credibility judgments but not valuation judgments, and is therefore excluded from the tests involving valuation judgments.

given my directional predictions, with the exception of the residual reported in Panel B.

To test my predicted interaction, I derive contrast weights using the procedure outlined in Buckless and Ravenscroft [1990]. The specific contrast weights are as follows: +2 in the Good News/Less Readable Condition, -2 in the Bad News/Less Readable Condition, +3 in the Good News/More Readable Condition, and -3 in the Bad News/More Readable Condition. The first two contrast weights reflect that I expect a relatively strong main effect of good vs. bad news, even at low levels of readability. The third and fourth weights reflect that I expect more readable disclosures to cause changes in valuation judgments to be larger than their less readable counterparts in the same news condition. Results of this planned contrast are presented in Panel B of Table 1.¹² As expected, I find support for my predicted interaction ($p < 0.001$, one-tailed).¹³ Figure 4 presents the plot of the interaction between news and readability. I also confirm that readability has the predicted effect for both good and bad news by conducting tests of simple main effects. Panel C of Table 1 shows that when news is good (bad), a more readable disclosure leads to significantly more positive (negative) changes in valuation judgments ($p = 0.012$ and 0.046 , respectively, both one-tailed).¹⁴

¹² The interaction of news with readability is also significant without planned contrast weights ($p = 0.003$, one-tailed), or with alternative contrast weights reflecting the (1) main effect of news and (2) the predicted interaction.

¹³ All p -values for analyses using F -statistics to test directional predictions are reported as one-tailed equivalents.

¹⁴ The normality assumption is violated in three of the four cells of my design for the variable capturing changes in valuation, but my large sample size ($n > 54$ in each treatment) suggests that parametric tests are robust to this deviation from the normality assumption (Scheffé [1959]). Furthermore, performing Levene's test fails to reject the assumption of homogeneity of variance ($p = 0.579$). Therefore, for my analyses of changes in valuation judgments I focus on means and perform parametric tests. However, inferences are the same when I use the nonparametric Wilcoxon Rank Sum test. When news is good, a

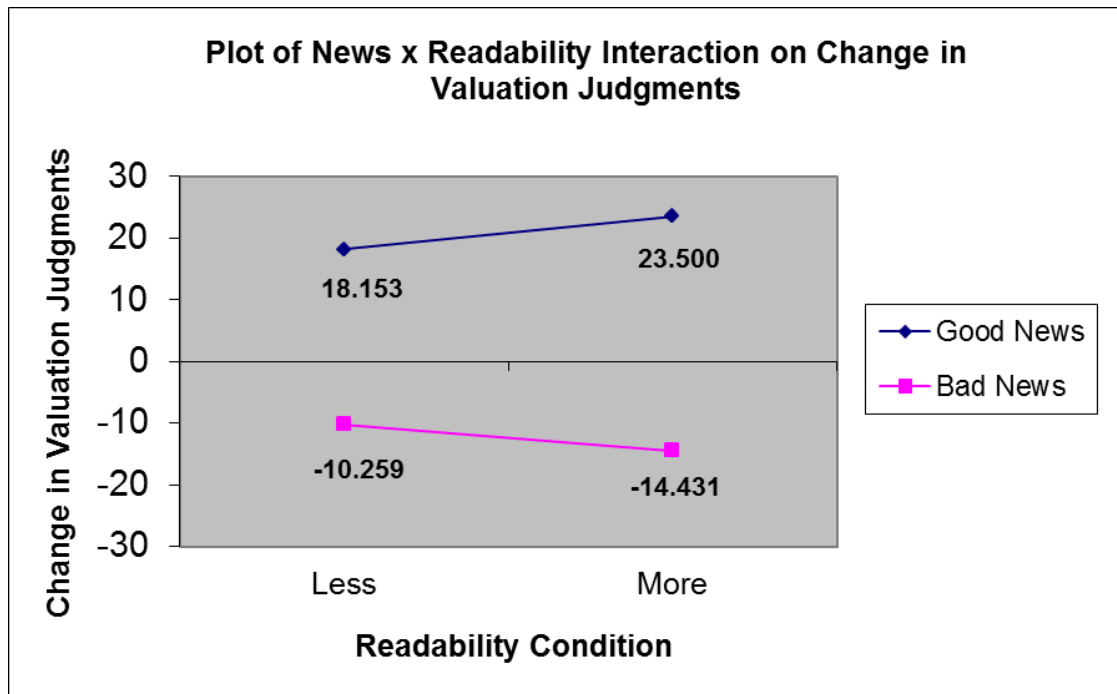


Figure 4. Plot of News x Readability Interaction on Change in Valuation Judgments – Stage 1

To verify that the stronger reactions to more readable disclosures are not driven by differences in information acquisition, participants respond to three multiple-choice questions about the press release. These questions are designed to be somewhat difficult in order to ensure that differences in information acquisition can be detected. I ask them to respond to questions asking them (1) by how much the firm’s revenues changed in the quarter compared to the same quarter in the prior year, (2) whether the CEO expected deteriorating, improving, or steady performance in the upcoming quarter and annual results, and (3) to name the two countries in which the firm primarily operates. The first two questions are designed to detect whether

more readable disclosure leads to significantly more positive changes in valuation judgments ($Z=2.33$, $p=0.010$, one-tailed). When news is bad, a more readable disclosure leads to significantly more negative changes in valuation judgments ($Z=1.86$, $p=0.031$, one-tailed).

participants have acquired information about the firm's past performance and future expectations, both of which may be particularly relevant to making valuation judgments. The last question is designed to detect whether they attended to the detailed information about the firm. There are no significant differences in the overall proportion of responses that are correct across the four conditions ($p=0.539$, two-tailed). This supports the claim that a more readable disclosure does not lead to greater information acquisition in my experiment, and therefore suggests that reactions to actual differences in participants' information are not driving my results.¹⁵ For the first two questions, the proportion of participants answering correctly is 73.93% and 82.91%, respectively. For the last question, 98.72% of participants correctly identify at least one of the two countries in which the firm operates, while 84.19% successfully identify both countries.¹⁶

1.4.2 Effects of Readability on Processing Fluency and Reliance on a Disclosure

H2 predicts that a more readable disclosure will increase feelings of processing fluency and, in turn, increase investors' beliefs that they can rely on the disclosure. As discussed in Section 1.3.5, I measure reliance in two ways. Panel A of Figure 5 summarizes results of a mediation analysis using my first measure of reliance. The mediation analysis is conducted according to the 4-step procedure specified by Baron and Kenny [1986]. In this analysis, the dependent variable is participants' agreement

¹⁵ Responses in Stage 2 further support the claim that differences in information acquisition are not driving my results. Participants who received the less readable disclosures in Stage 1 did not revise their valuation judgments in Stage 2, suggesting that they did not feel as though the more readable disclosure provided in Stage 2 gave them access to additional information.

¹⁶ For the individual questions, the only difference detected is that participants are more likely to answer the second question correctly with a more readable disclosure. However, this difference only occurs for bad news ($p=0.036$, one-tailed) rather than good news ($p=0.481$, one-tailed), so it cannot explain the good news results. Furthermore, if I analyze only those participants that answer the second question correctly, my contrast testing H1 is still significant ($p<0.001$), as are the main effects for good news ($p=0.021$, one-tailed) and bad news ($p=0.020$, one-tailed).

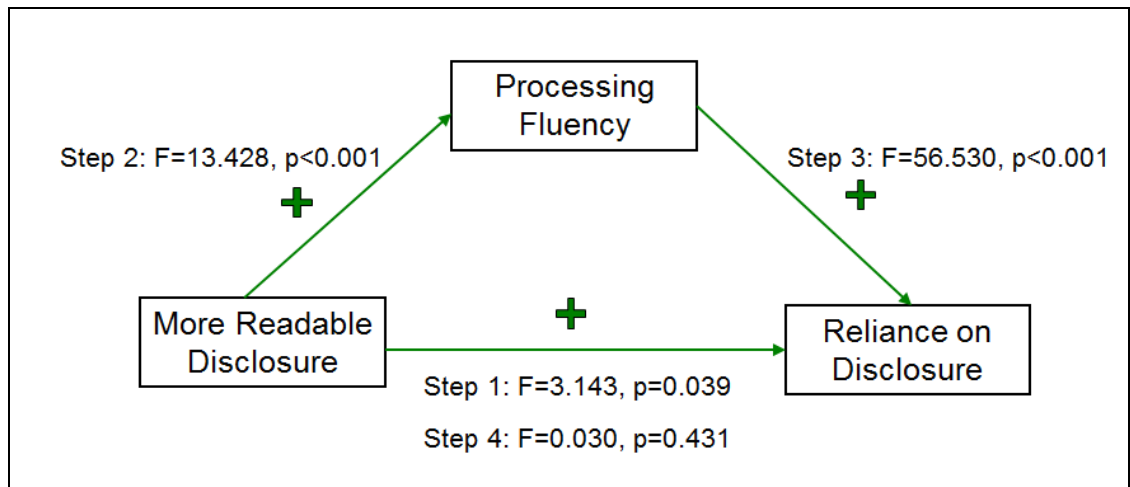
with the statement “I felt like I could rely on the information in the press release” (1 = strongly disagree, 7 = strongly agree). With respect to Step 1, readability is significantly positively associated with participants’ beliefs that they can rely on the press release ($p=0.039$, one-tailed). Step 2 confirms that readability is significantly positively associated with participants’ perceptions of processing fluency ($p<0.001$, one-tailed). Step 3 confirms that the mediating variable, processing fluency, is significantly associated with participants’ reliance on the press release, even after controlling for readability ($p<0.001$, one-tailed). Finally, Step 4 confirms that processing fluency fully mediates the relationship between readability and participants’ beliefs that they can rely on the press release, as the effect of readability on reliance is no longer significant when included in the model with processing fluency ($p=0.431$, one-tailed).

Panel B of Figure 5 presents the mediation analysis for H2 with a different measure of reliance as the dependent variable: the unsigned measure capturing the change in valuation judgments. While noisier than a direct measure of reliance, the unsigned measure reflects that changes in valuation judgments should be larger (as demonstrated in H1) if reliance on a disclosure is greater. Step 1 confirms that readability is significantly positively associated with participants’ reliance on the press release ($p<0.001$, one-tailed). Step 2 again confirms that readability is significantly positively associated with participants’ perceptions of processing fluency ($p<0.001$, one-tailed). Step 3 confirms that processing fluency is significantly associated with participants’ reliance on the press release, after controlling for disclosure readability ($p=0.016$, one-tailed). Finally, Step 4 supports partial mediation. Readability is still significant, but the significance is reduced when processing fluency is included in the

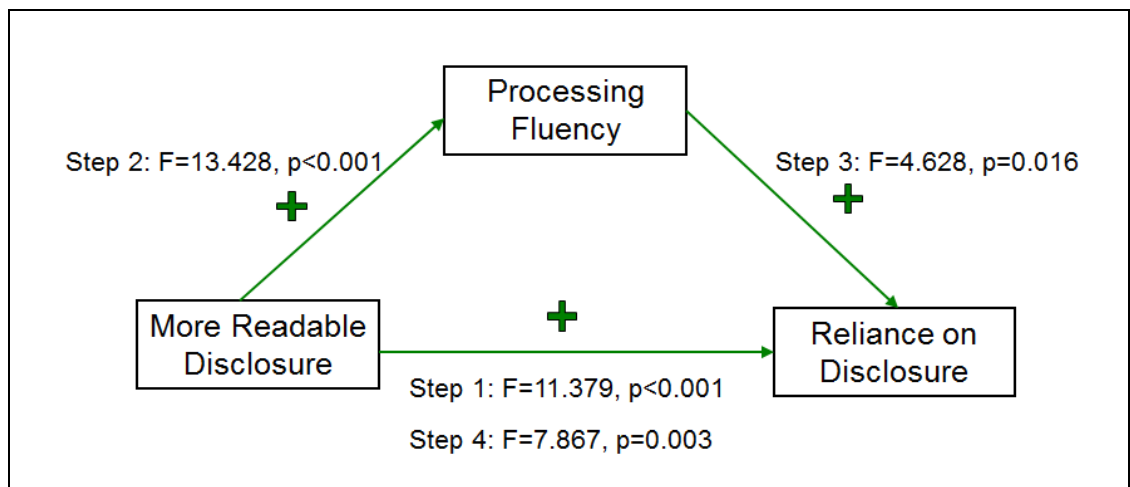
regression (<0.001 to 0.003). I use the Sobel test to confirm that the direct effect is reduced between Steps 1 and 4 (that is, once the mediator is included ($p=0.032$, one-tailed)). Results in Figure 5 support H2 and show that a more readable disclosure increases participants' perceptions that they can rely on the information in the disclosure, as a result of increased processing fluency.

Figure 5. Mediation Analyses for testing H2

Panel A. Beliefs about Ability to Rely on the Disclosure as the Dependent Variable



Panel B. Unsigned Change in Valuation Judgments as the Dependent Variable



1.4.3 Effects of Readability on Judgments of Management Credibility

To test H3, I begin by confirming that the trustworthiness and competence responses represent the same underlying credibility construct by performing a reliability analysis. The resulting Cronbach's alpha is 0.78, which exceeds the recommended threshold of 0.70 (Nunnally [1978]). Therefore, I combine participants' judgments of management trustworthiness and competence into one management credibility measure by averaging the two responses. Panel A of Table 2 presents descriptive statistics for the credibility measure. Panel B presents results of two-sample t-tests of the main effect of readability on credibility. However, the effect is not significant for the full sample ($p=0.299$, one-tailed) or in either the good news ($p=0.141$, one-tailed) or bad news ($p=0.531$, one-tailed) subsamples. Thus, the effect predicted in H3 is not supported, and press releases that are more readable do not directly lead to higher ratings of management credibility. Furthermore, there are no significant effects of readability on either the management competence or management trustworthiness judgments when analyzed separately.

A contemporaneous working paper by Tan et al. [2011] may help to explain why I do not find a direct effect of readability on credibility. Tan et al. [2011] focus on the credibility-enhancing role of disclosure readability, but in a setting where their experimental manipulations affect the likelihood that investors spontaneously consider that managers might choose levels of disclosure readability strategically. The authors find that disclosure readability has a stronger effect on credibility judgments when a disclosure contains inconsistent information about a firm's performance (e.g., favorable information in the opening paragraphs of the disclosure, but disappointing news in a latter paragraph). Tan et al. [2011] argue that this occurs because the

inconsistent performance information makes investors more sensitive to managers' reporting incentives, so that investors' judgments of disclosure credibility are higher (lower) when managers make the incentive-inconsistent (incentive-consistent) decision to provide a more (less) readable disclosure. In my study, the press releases provided to participants convey primarily either good or bad news, and do not contain stark inconsistencies that might specifically trigger strong concerns about credibility.

Table 2. Descriptive Statistics and Two-Sample T-Tests – Credibility Judgments

Panel A. Descriptive Statistics - Mean, (Standard Deviation), and [Median]

News	Readability		Overall
	Less	More	
Good	77.415 (13.481) [80.500] n=59	79.806 (10.800) [80.750] n=62	78.640 (12.189) [80.500] n=121
Bad	61.418 (13.688) [61.000] n=55	61.198 (16.329) [60.750] n=58	61.305 (15.035) [61.000] n=113
Overall	69.697 (15.725) [70.000] n=114	70.813 (16.574) [72.500] n=120	

Panel B. Credibility Judgments - Two-Sample T-Tests

Sample	Difference Between		
	More and Less Readable	t-statistic	p-value
Full Sample	1.116	0.528	0.299
Good News	2.391	1.079	0.141
Bad News	-0.220	-0.078	0.531

Notes to Table 2. Panel A presents descriptive statistics on the credibility measure in my experiment, in which 234 participants read a press release and provide associated competence and trustworthiness judgments on a 101-point scale (0=low, 100=high). Credibility is measured as the average of the Trustworthiness and Competence judgments after Cronbach's alpha confirms that the two measures are capturing the same underlying construct. Panel B presents results for testing the effects of readability on credibility in the full sample as well as the good and bad news subsamples. All p-values are one-tailed given my directional predictions.

I do, however, find evidence that readability has an indirect effect on management credibility through its influence on processing fluency. Hayes [2009] cautions that failure to test for an indirect effect in the absence of a direct or total effect may cause researchers to miss “potentially interesting, important or useful mechanisms by which [the independent variable] exerts some kind of effect on [the dependent variable].” Others have pointed out that an indirect effect can be established even if the direct or total effect of the independent variable on the dependent variable is not significant (Kenny, Kashy and Bolger [1998], MacKinnon et al. [2000]; Shrout and Bolger [2002], Zhao et al. [2010]).

Following the advice of Zhao et al. [2010], I use a bootstrap analysis to test whether readability indirectly increases judgments of management credibility through increased feelings of processing fluency. The bootstrapping method generates an empirical sampling distribution of the indirect effect and generates a confidence interval for its value. Drawing a bootstrap sample of $N=5000$, I find that the mean indirect effect is positive (mean = 1.553, untabulated), with a 99% confidence interval excluding zero (0.135 to 3.727, untabulated). My analysis also separately confirms the significance of the two components of the indirect effect: readability increases feelings of processing fluency ($p<0.001$, one-tailed, untabulated), and feelings of processing

fluency increase judgments of management credibility ($p=0.001$, one-tailed, untabulated). A significant indirect effect in the absence of a direct or total effect could mean that there are one or more effects that work in the opposite direction of fluency (MacKinnon et al. [2000], Shrout and Bolger [2002] Zhao et al. [2010]). For example, more readable bad news may lead to lower credibility assessments through an alternative indirect path, offsetting the positive effects of fluency and canceling out the direct effect, should it exist. Alternatively, an indirect effect in the absence of a significant direct or total effect could also mean that a moderate direct effect exists but is not detectable without a larger sample size (Shrout and Bolger [2002]).

Finally, to provide some evidence on the potential economic consequences of management credibility judgments, I investigate their relationship with participants' valuation judgments used in testing H1. I find that the correlation between valuation judgments and judgments of management credibility is positive and significant ($r=0.561$, $p<0.001$, untabulated). Furthermore, if I run the ANOVA testing for a News x Readability interaction again but add credibility as a covariate, the effect of credibility judgments on valuation judgments is positive and significant ($t=4.413$, $p<0.001$, untabulated), and the News x Readability interaction remains significant as well.

1.4.4 Supplemental Analyses

Stage 2 of my experiment presents participants with both the more and less readable press releases simultaneously for their performance condition (i.e., good or bad news), and asks them to respond to a variety of questions, as described in Section 1.3. Of particular interest in Stage 2 is that participants provide new valuation

judgments (on the same 101-point scales as in Stage 1). This allows a test of whether and how the responses documented in testing H1 differ once it is made salient that disclosure readability can vary.¹⁷

Panel A of Table 3 presents, by condition, the mean difference between Stage 2 valuation judgments and judgments made in Stage 1 after viewing the press release. I find that participants in the good news condition with a more readable disclosure provide valuation judgments in Stage 2 that are significantly lower than their valuation judgments in Stage 1 ($p=0.095$, two-tailed).¹⁸ At the same time, participants in the bad news condition with a more readable disclosure provide valuation judgments in Stage 2 that are significantly higher than their valuation judgments in Stage 1 ($p<0.001$, two-tailed). Participants who initially received less readable disclosures (good or bad news), do not provide valuation judgments in Stage 2 that significantly differ from their valuation judgments in Stage 1 (both p -values are greater than 0.212, two-tailed).

Similar to the measure of participants' reactions used in testing H1, I also calculate a measure for each participant of the difference between valuation judgments made in Stage 2 and initial valuation judgments made at the beginning of Stage 1 (i.e., before receiving the press release). Analyzing this measure shows that there is no significant difference between reactions to more and less readable disclosures for either good news ($p=0.723$, two-tailed) or bad news ($p=0.112$, two-tailed) (see Panel B of Table 3). Figure 6 presents the plot of this measure, by condition. Comparing Figure 4 to Figure 6 shows how participants' valuation judgments differ depending on

¹⁷ As described in Section 1.3, participants also provide new credibility judgments on the same 101-point scales as in Stage 1. However, I do not compare Stage 1 and Stage 2 credibility judgments due to the lack of a significant effect in testing H3.

¹⁸ All p -values presented in the supplemental analyses are two-tailed, given that I make no *ex ante* directional predictions.

whether they are elicited either before or after participants are explicitly made aware of the potential for variation in disclosure readability.

Table 3. Descriptive Statistics, T-Tests and Tests for Simple Main Effects of Readability on Stage 2 Valuation Judgments

Panel A. Tests of Whether Participants Change Valuation Judgments between Stages 1 and 2

Condition	Mean Change	t-statistic	p-value	Change (if any)
Good News - High Readability	-2.261	-1.698	0.095	Significant decrease.
Good News - Low Readability	1.336	1.183	0.212	No change.
Bad News - High Readability	8.410	3.787	<0.001	Significant increase.
Bad News - Low Readability	-1.306	-0.704	0.485	No change.

Panel B. Simple Main Effects of Readability on Stage 2 Valuation Judgments

Source	S.S.	d.f.	M.S.	F-statistic	p-value
Good News	25.059	1	25.059	0.126	0.723
Bad News	495.143	1	495.143	2.483	0.116

Notes to Table 3. This table presents descriptive statistics, t-tests, and simple main effects tests for the measures used in my supplemental analyses to examine investors' valuation judgments after they explicitly consider the potential for variation in disclosure readability. Panel A presents the mean change, by condition, between the Stage 1 valuation judgments made after receiving the press release and the Stage 2 valuation judgments. Panel B presents results of testing for simple main effects of readability on valuation judgments in Stage 2, for both the good and bad news conditions. All p-values in Table 3 are two-tailed given that I have no ex ante directional predictions.

This pattern of revised valuation judgments suggests that the stronger reactions to more readable disclosures found in H1 may be unintentional (Kahneman and Tversky [1996], Tan, Libby, and Hunton [2002]). This is also consistent with findings in psychology showing that people are not affected by processing fluency once they realize its source (Schwarz [2004], Alter and Oppenheimer [2009]). An alternative

possibility is that participants are not revising their valuation judgments to correct for overreaction to a more readable disclosure, but are instead revising their judgments in response to consideration of managers' strategic readability choices. Unfortunately, it is difficult to make investors aware of the potential for variation in disclosure readability without also making them aware of managers' role in that variation (and therefore the potential for strategic behavior). As a result, I cannot clearly disentangle these two possibilities, and Stage 2 valuation judgments should be interpreted with caution. However, I believe that the results support a "correction for overreaction" explanation more than a "reaction to managers' strategic behavior" explanation. Correction for overreaction would more clearly predict that participants in the more readable conditions would revise their valuation judgments to be less extreme (as I observe). If participants are instead reacting to consideration of managers' strategic behavior, I would predict that participants would revise their valuation judgments in the two conditions where levels of readability were more consistent with managers' incentives. More specifically, I would expect those who are in the less readable/bad news condition and the more readable/good news condition to both revise their valuation judgments downward (a pattern which I do not observe).

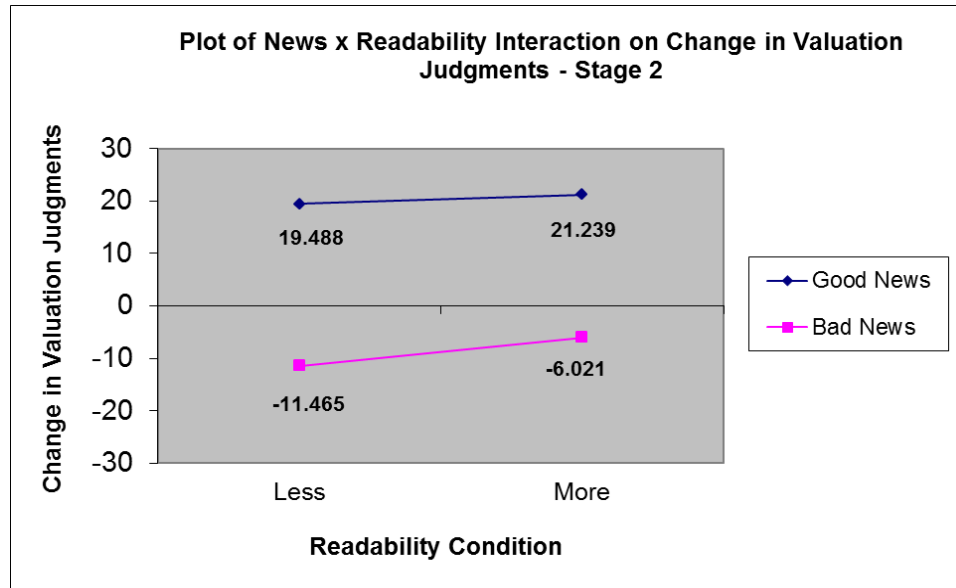


Figure 6. Plot of News x Readability Interaction on Change in Valuation Judgments – Stage 2

1.5 Conclusion

Recent work in the archival accounting literature investigates disclosure readability (Li [2008], You and Zhang [2009]) and its effects on the behavior of small investors (Miller [2010]). I use a controlled experiment to provide complementary evidence and to address questions that cannot be answered with archival data. I find that, holding constant length and information content, more readable disclosures lead to stronger reactions from investors, so that changes in investors' valuation judgments are more positive when news is good and more negative when news is bad. Consistent with prior literature in psychology, I find that participants are more likely to feel as though they can rely on a disclosure that is more readable, and that this effect is mediated by processing fluency. This greater reliance on news (be it good or bad) helps to explain the stronger reactions to more readable disclosures that I observe,

even though readability does not lead to significant differences in the actual information that participants are able to gather from the press release. Counter to my predictions, I do not find that disclosure readability directly affects perceptions of management credibility. However, I do find evidence that readability has a positive and significant indirect effect on management credibility, and that this indirect effect operates through increased feelings of processing fluency.

In Stage 2 of my experiment, I find that participants who initially received the more readable disclosures provide valuation judgments in Stage 2 that are less extreme than the valuation judgments they provide in Stage 1, to the point where the effect of readability on investors' reactions is no longer significant. These results suggest that the stronger reactions in response to more readable disclosures in Stage 1 of my experiment may have been unintentional (Kahneman and Tversky [1996], Tan, Libby and Hunton [2002]). This is consistent with prior research showing that individuals do not treat processing fluency as a cue in their judgments once they realize its source (Schwarz [2004], Alter and Oppenheimer [2009]). This result has potentially important implications for the SEC's push towards improving disclosure readability. While the SEC describes more readable disclosures as though they are unambiguously positive [SEC 1998], my results indicate that more readable disclosures may actually lead investors to overreact to information, particularly those who are the least sophisticated. In other words, investors may be too quick to accept information that is presented in a way that is easy to process. This suggests that the benefits of more readable disclosures may be less clear-cut than has been argued by the SEC [SEC 1998] and in prior accounting literature (e.g., Li [2008], You and Zhang [2009], Miller [2010]).

My findings add to the growing body of literature investigating the style of disclosures as opposed to their content. Content is the literal meaning of the information conveyed, or the concrete facts contained in a disclosure, whereas style captures the methods used to convey meaning to the audience. In response to Core's [2001] call for greater computational linguistic processing of qualitative disclosures, stylistic factors besides readability have been investigated in the archival literature, including certainty (Demers and Vega [2010]), numerical intensity (Henry [2008]), boilerplate vs. meaningful language (Nelson and Pritchard [2007]), and tone (Henry [2008], Davis, Piger and Sedor [2011], Davis and Tama-Sweet [2011], Feldman, Govindaraj, Livnat and Segal [2009], Li [2010a], Demers and Vega [2010], Rogers, Van Buskirk and Zechman [2011]). However, experimental investigation of stylistic characteristics is relatively scarce in the recent accounting literature. One exception is Hales et al. [2011], which investigates the effects of vivid vs. pallid language, above and beyond actual information content. Other recent work has looked at the determinants of the style characteristics of verbal communications (Hobson, Mayew and Venkatachalam [2011]), as well as reactions to verbal communications (Mayew and Venkatachalam [2012], Elliott, Hodge and Sedor [2012]).

A limitation of my study is that it is not clear whether investors' reactions differ in response to individual linguistic and formatting choices. I made this design choice to produce a strong treatment effect. My treatments also correspond with real-world disclosures, because the use of either high- or low-quality linguistic and formatting features is likely to be highly correlated in a given disclosure. However, there is substantial opportunity for future experimental research investigating how individual stylistic features of a disclosure affect investors' judgments. Experiments

are also likely to be well-suited to the investigation of managers' disclosure decisions (to investors, but also to auditors, other employees, clients, etc.). Sociolinguistics research supports the prediction that choice of linguistic characteristics may vary in line with specific goals and incentives.¹⁹ Based on these findings, future work might investigate whether managers intentionally or unintentionally use personal pronouns and other linguistic features to influence others' reactions as firm performance and incentives vary (Chatterjee and Hambrick [2007], Li [2010b]).

My study also adds to the literature on processing fluency, and suggests that it may be an important mechanism for both understanding results in prior studies and motivating predictions in future research. While processing fluency is mentioned only rarely in the existing accounting literature (e.g., Koonce and Lipe [2010]), a number of prior findings are consistent with outcomes that would be predicted by research on processing fluency. In their review, Alter and Oppenheimer [2009] point out that all tasks range from effortless to demanding, and that there are a number of sources of processing fluency in a given task. As a result, processing fluency effects are likely to be pervasive across accounting settings. The availability of information, or ease of retrieval, is just one manifestation of processing fluency (Alter and Oppenheimer [2009]). Thus, accounting studies showing that professional and nonprofessional investors do not spontaneously recall or consider unfavorable information about a firm shed light on how processing fluency might affect investor behavior (e.g., Kadous et al. [2006], Krische [2005]). Similarly, conceptual processing fluency relates to the ease with which individuals access relevant (and sometimes irrelevant) knowledge

¹⁹ For example, Gunsch, Brownlow, Haynes and Mabe [2000] investigate how personal pronouns and present/future tense verbs appear to be strategically used in political ads to focus voters on a positive future relationship with candidates. Negative ads tend to emphasize negative past behaviors of the opponent and frame him or her as distant.

structures to complete a task (Schwarz [2004]). Findings in Hopkins [1996] can perhaps then be thought of as evidence that financial statement classification affects conceptual processing fluency by making category-relevant information more or less accessible.

The effects of processing fluency also represent a potentially fruitful opportunity for future research. For instance, the use of “jargon” is likely to feel more fluent to those with more experience in financial reporting settings, suggesting important experience effects of processing fluency. Furthermore, processing fluency research might also suggest potential remedies for biased decision-making in accounting settings. Alter, Oppenheimer, Epley and Eyre [2007] show that in some circumstances disfluent presentation can lead to more systematic processing, presumably because disfluency can act as a signal that greater effort should be expended on the task. This suggests that processing fluency could be manipulated in a variety of ways to induce relatively more cautious or aggressive behavior among managers, investors, auditors, etc.

Finally, my study uses a source of participants that may prove useful for future research. The AMT platform provides access to more participants than would typically be available in a laboratory setting, and at a substantially lower cost. Furthermore, AMT has controls in place to help alleviate some of the typical concerns raised with online experiments (as discussed in Bloomfield and Rennekamp [2009]). On average, participants in my study have some experience with investing, have taken some accounting and finance courses, have more than a decade of full-time work experience, and take the task seriously enough to correctly answer questions that check for their attention. Future work should consider utilizing AMT (or other online

crowd-sourcing platforms) for accounting research, particularly with respect to the judgments of less-sophisticated investors.

CHAPTER 2

THE INFLUENCE OF PERFORMANCE AND REPORTING GOALS ON MANAGERS' CHOICE OF REPORTING COMPLEXITY IN DISCLOSURES

2.1 Introduction

Theory predicts that the market reaction to information will be weaker as it becomes more difficult or costly to extract (Bloomfield [2002]; Hirshleifer and Teoh [2003]). Consistent with this prediction, recent studies in the accounting literature demonstrate that reporting complexity in qualitative disclosures is associated with weaker reactions to news (You and Zhang [2009]; Rennekamp [2012]) and lower trading volume among small investors (Miller [2010]).²⁰

Some prior studies have argued that managers use this to their advantage and intentionally increase the reporting complexity of annual reports when firm performance deteriorates (e.g., Subramanian, Insley and Blackwell [1993]; Courtis [1998]; Li [2008]). For example, Li's [2008] "management obfuscation hypothesis" proposes that managers strategically increase the processing cost of adverse information in order to reduce or delay its incorporation into stock price. Using techniques from computational linguistics, Li [2008] finds the first large-sample evidence that loss firms prepare annual reports that are both longer and less-readable than firms with positive earnings. Among firms with positive earnings, firms with more transitory earnings prepare longer and less-readable annual reports.

Bloomfield [2008] acknowledges that managers may intentionally obfuscate

²⁰ For the purposes of this chapter and experiment, I define a text as higher in reporting complexity if it is (1) longer and/or (2) less readable, consistent with Miller [2010]. You and Zhang [2009] focus on reporting length, while Rennekamp [2012] focuses on readability. Miller [2010] investigates both components of reporting complexity.

bad performance to mitigate investors' reactions. However, he also suggests that the reporting complexity of disclosures might increase when firm performance is poor simply because the language used to describe poor performance may be inherently more complex. Specifically, Bloomfield [2008] suggests that reporting complexity might increase if negative events are more unique or unusual, and if unusual events require more complex language to explain.²¹ Bloomfield's [2008] explanation presents an important alternative to the idea that managers intentionally obfuscate information about poor performance.

I use an experiment to investigate these two potential explanations for reporting complexity during periods of poor performance. In my experiment, participants are asked to assume that they are the Vice President of a division in a hypothetical firm, and are told that they have been asked to draft a report to headquarters describing the division's performance in the most recent quarter.²² I manipulate between-subjects whether the division performed well or poorly in the most recent quarter, and whether budgeting decisions for the next year have already been made. Participants are then told to prepare a draft report to headquarters that conveys performance honestly. In the condition where budgeting decisions have already been made for the coming year, participants' only explicit incentive is to provide an accurate report. In the condition where budgeting decisions have not yet

²¹ Bloomfield [2008] also suggests that disclosure complexity may increase if managers engage in biased attribution, management by exception, misdirection, or are concerned about litigation. However, while biased attribution could change the actual words that managers choose, it is less clear that biased attribution would make a disclosure longer and/or less readable. Similarly, management by exception or the use of misdirection is likely to make a disclosure longer, but not necessarily less readable. Finally, addressing managers' concerns with litigation is outside the scope of my study.

²² I use a setting of *internal* performance reporting in order to remove legal and other considerations that might be present in an external performance reporting setting to investors. I believe that this decision choice reduces noise in my responses, while allowing me to still address fundamental questions about linguistic choices in performance reporting decisions.

been made, participants are still told to provide a report that is accurate, but are also told to highlight the positive aspects of the division in order to secure more funding for the future. This design choice imparts participants with a strategic reporting goal of presenting the division as favorably as possible.

The fundamental purpose of language is to communicate, and findings from sociolinguistics research support the idea that linguistic choices vary to reflect thought processes, motives and emotions, particularly during periods of personal upheaval or shared crisis (Pennebaker and Lay [2002]; Pennebaker, Mehl and Niederhoffer [2003]). Thus, I use methods from computational linguistics to analyze the natural language choices that participants make to describe the division's performance in light of their reporting goals (Tausczik and Pennebaker [2010]).

I predict and find that when participants are asked to explain the division's performance, their reports are significantly longer and less-readable when performance is bad than when performance is good. This finding holds even when participants are told that their only goal is to accurately convey performance, and is consistent with Bloomfield's [2008] suggestion that poor performance may be inherently more difficult to explain concisely and in more readable language. However, I do not find that poor performance interacts with reporting goals in the predicted way. Based on arguments in the prior literature, my initial prediction is that reporting complexity will be greatest when performance is bad and participants are given explicit instructions to prepare a report that will help the division secure future funding (i.e., a strategic reporting goal is present). While this goal does lead participants to provide significantly longer reports, the reports are significantly *more* readable than when participants are only given a goal of reporting accurately.

There are at least three potential explanations for the discrepancy between my findings and the arguments made in prior literature that reporting complexity is used to intentionally obfuscate poor performance. One is that prior studies may have instead picked up on the inherent complexity required to explain bad performance (as opposed to strategic obfuscation), as suggested by Bloomfield [2008]. Secondly, it is also possible that the discrepancy stems from differences in my setting compared to that of a typical archival study. My experiment focuses on an internal reporting setting, in order to get at the fundamental question of how individuals describe performance as incentives to present themselves favorably vary. This choice likely reduces noise in participants' responses, but may also fail to capture some of the additional considerations that managers face when making external reporting decisions. Finally, I use a setting where participants describe hypothetical performance, rather than their performance on a real task (e.g., intelligence tasks, manufacturing tasks, etc.). As discussed in more detail in Section 2.3.3, this mitigates potential confounds between linguistic choices and intelligence. However, a consequent downside of asking participants to describe hypothetical performance is that it potentially reduces their personal involvement in the task, and may influence their linguistic choices as well.

In addition to the results discussed above, I also provide some preliminary evidence on how other linguistic choices are affected by variation in performance and reporting goals. I find that participants use more personal pronouns when performance is good than when performance is bad, particularly when they have a strategic reporting goal. This is consistent with prior work suggesting that individuals can use personal pronouns to associate themselves with a message (Hyland [2005]) and make it more persuasive (Asay, Libby and Rennekamp [2012]). I also find that, when

performance is bad, participants use more past tense language when they have an accuracy goal, and more future tense language when they have a strategic reporting goal.

My findings add to the growing body of literature investigating the linguistic style of disclosures as opposed to linguistic content. Linguistic content is the literal meaning of the information conveyed, or the concrete facts contained in a disclosure, whereas linguistic *style* captures the methods used to convey meaning to the audience. Controlling for actual firm performance, recent archival studies use methods from computational linguistics to analyze large bodies of text, as suggested by Core [2001]. However, these studies focus almost exclusively on investigating whether various linguistic style factors affect trading behavior and convey incremental information to the market.²³ My study, on the other hand, uses a controlled experiment to provide complementary evidence on *managers'* disclosure choices while holding constant other factors that are not of primary interest. I find that firm performance and explicit reporting goals affect choice of reporting complexity. Future experimental research might investigate additional linguistic choices that have been shown to reflect individuals' motives and circumstances (e.g. linguistic tone, causal explanations, etc.). Future research might also investigate both written and spoken settings of managerial communication (e.g. annual reports, interviews, conference calls, etc.).

The remainder of the chapter proceeds as follows. Section 2.2 provides background information and develops my hypotheses. Sections 2.3 and 2.4 discuss my

²³ Some factors investigated include Readability (Li [2008]; Miller [2010]; Loughran and McDonald [2011]), Certainty (Demers and Vega [2010]), Numerical Intensity (Henry [2008]), Boilerplate vs. Meaningful Language (Nelson and Pritchard [2007]), and Tone (Nelson and Pritchard [2007]; Henry [2008]; Davis, Piger and Sedor [2011]; Demers and Vega [2010]; and Feldman, Govindaraj, Livnat and Segal [2009]), to name a few.

experiment and results, respectively. Section 2.5 concludes.

2.2 Background and Development of Hypotheses

At its core, language is a social construct. Words (whether written or spoken) are the primary tool that individuals use to communicate with others on a day-to-day basis. It is not surprising then that findings from research on sociolinguistics suggest that the linguistic choices of individuals reveal their social-psychological processes and vary with personal and environmental circumstances (Gee [1999]; Pennebaker et al. [2003]). In a financial reporting context, prior findings suggest that the linguistic characteristics of qualitative disclosures are likely to vary with firm performance (for reviews, see Jones and Shoemaker [1994]; Merkl-Davies and Brennan [2007]; and Li [2011]). For the purposes of this study, I focus on variation in reporting complexity, where a disclosure is higher in reporting complexity if it is longer and less readable (Li [2008]; Miller [2010]). Variation in reporting complexity may be due to intentional (conscious) choices or may be an unintentional consequence of environmental circumstances.

2.2.1 Strategic Obfuscation by Management

Much of the accounting literature assumes that choices of reporting complexity are conscious, and that managers attempt to obfuscate poor performance but highlight good performance (for reviews see Jones and Shoemaker [1994]; and Li [2008]). Using a sample of 60 U.S. firms, Subramanian, Insley and Blackwell [1993] look at how readability of the Chairman's Letter varies with firm performance. Over a two-year period, they divide their sample into firms where net profit either increases or decreases in the second year of the period, compared to the first. They find that readability is significantly higher for the group of firms where performance improves

rather than deteriorates.²⁴ Courtis [1998] finds that firms with more media attention have Chairman's Letters that are less readable. He claims that this is consistent with an obfuscation hypothesis. That is, firms with bad news may be more likely to receive media coverage, and may also be more likely to try and obfuscate the bad news by reducing the readability of the Chairman's Letter. However, Rutherford [2003] argues that previous studies of the obfuscation hypothesis (pre-2003) should be interpreted cautiously, because they rely on "small samples of tests, testing of limited amounts of text, and the application of weak tests of association".²⁵ Li [2008] provides some of the first large-sample evidence on the issue. Using the FOG index and disclosure length to measure reporting complexity, Li [2008] finds that firms with higher earnings have annual reports that are shorter and more readable. Furthermore, when performance is positive, firms with more readable annual reports have more persistent positive earnings. Li [2008] concludes that firms intentionally provide less-readable annual reports when performance is poor or when positive performance is unsustainable.

In his response to Li [2008], Bloomfield [2008] concedes that management obfuscation is a plausible hypothesis, and follows up with a case study for a single firm's MD&A over three years. The chosen firm has positive earnings in the first year, record earnings in the second year, and negative earnings in the third year. Consistent with Li's [2008] findings, Bloomfield [2008] finds that the length of the annual report is longer in the loss year and transitory positive earnings year than in the persistent

²⁴ However, Subramanian et al. [1993] conduct no analyses that would rule out reverse-causality. That is, it is possible that firms with better management (who would be more likely to fall in the group with improving performance) were already providing more readable disclosures.

²⁵ And, in fact, many of the pre-2003 studies do *not* find evidence that readability varies with performance (e.g., Courtis [1986], Jones [1988], and Clatworthy and Jones [2001]). Again, Rutherford [2003] points out that it is difficult to interpret these early studies.

positive earnings year. Bloomfield [2008] also finds that much of the increased length in the loss year is attributable to inclusion of additional information on fairly standard accounting issues, suggesting that they were added to the MD&A to bog down readers and obfuscate the firm's performance. Finally, Bloomfield [2008] also points out that firms' disclosures may be more detailed if they provide biased attribution or attempt to misdirect readers' attention, both of which are consistent with managers trying to obfuscate information and mitigate reactions.

Recent research in the accounting literature supports the idea that intentional obfuscation may be a useful strategy for managers, and finds evidence consistent with the prediction that disclosure complexity dampens investors' reactions to good and bad news. You and Zhang [2009] use word-count as their proxy for reporting complexity and find that market underreaction is more severe as the complexity of the 10-K increases. Using techniques from computational linguistics, Miller [2010] develops a proprietary measure of disclosure readability based on the SEC's plain English guidelines to show that improved 10-K readability is associated with increased trading by small investors (i.e. those who are likely to have the most difficulty with complex disclosures).²⁶ Likewise, Rennekamp [2012] uses an experimental setting to show that more readable disclosures lead to stronger reactions from investors to both good and bad news. She shows that this holds even for relatively short disclosures, where there are no significant differences in the actual amount of information that investors are able to acquire.

Prior research also suggests that managers *understand* that less-transparent disclosures reduce the likelihood that investors and analysts will be able to extract

²⁶ Miller [2010] finds similar results when he instead measures disclosure length.

information. For example, managers are more likely to engage in earnings management activities when their decisions are less-transparent, suggesting that they believe transparency makes earnings management a more value-enhancing activity (for a review, see Libby and Seybert [2009]).

2.2.2 The Inherent Complexity of Explaining Poor Performance

In addition to providing evidence supporting a management obfuscation story, Bloomfield [2008] outlines an alternative, and less sinister, explanation for variation in annual report readability. He argues that the actual language associated with poor performance may be inherently more complex, especially given that poor performance is often due to unique or unusual circumstances.²⁷

Consistent with this line of thinking, Daft and Macintosh [1981] argue that non-routine situations require greater information processing than routine situations. As a result, managers are more likely to convey non-routine information through richer mediums, such as face-to-face meetings rather than written reports (Daft, Lengel and Trevino [1987]). This matching of situational complexity to communication medium is more pronounced in high-performing managers (Daft et al. [1987]). Daft and Macintosh [1981] also suggest that routine situations are easier to analyze and summarize with quantitative information, whereas non-routine situations are more likely to be analyzed and summarized with qualitative information. This latter point is consistent with the idea that poor (and therefore non-routine) performance requires more explanation via words instead of summary numbers. More

²⁷ Other studies have looked at how reporting complexity varies in response to catering to certain types of investors. Kalay [2011] finds that sophisticated investors concentrate their trading in firms with more opaque information environments, presumably because they can take advantage of the greater information asymmetry in these types of firms. Similarly, Lehavy, Li and Merkley [2011] show that analysts also appear to be drawn to less readable disclosures, where their analysis is likely to be most needed.

explanation via words could both increase disclosure length and reduce disclosure readability, the two components of reporting complexity that I investigate in this study.

Rutherford [2003] examines the Management's Discussion and Analysis (MD&A) of firms' annual reports, and regresses reporting complexity on a variety of factors that could support either an obfuscation hypothesis or a "firm characteristics" story. While he finds some evidence that relatively risky and heavily regulated firms have less readable disclosures, he finds the greatest support for length being positively correlated with organizational size and complexity. While Rutherford's [2003] emphasis is on how organizational complexity drives reporting complexity, his findings suggest that environmental complexity (as in the case of poor performance, which is relatively more unique than good performance) could also be positively associated with reporting complexity.

Combined, findings outlined in the discussion above support the idea that bad performance requires more reporting complexity to explain. As a result, my first set of hypotheses is:

H1a: Managers provide longer disclosures when a firm is performing poorly than when a firm is performing well.

H1b: Managers provide less readable disclosures when a firm is performing poorly than when a firm is performing well.

My second set of hypotheses recognizes that the alternative argument for strategic obfuscation of poor performance by management is also plausible, and that the two possibilities need not be mutually exclusive. Given that managers appear to

understand the benefits of less-transparent disclosures, and that linguistic characteristics have been shown to vary with individuals' incentives, my second hypothesis is that:

H2a: When a firm is performing poorly, managers provide longer disclosures when they have a goal to elicit a favorable reaction than when they have a goal to convey information accurately.

H2b: When a firm is performing poorly, managers provide less readable disclosures when they have a goal to elicit a favorable reaction than when they have a goal to convey information accurately.

Figure 7 depicts H1a and H2a, and Figure 8 depicts H1b and H2b graphically. Combined, my four hypotheses predict that the reporting complexity of managers' disclosures will be influenced by the inherent difficulty of providing an explanation for bad performance, but also by intentional obfuscation of information when goals of eliciting a favorable reaction are especially salient.

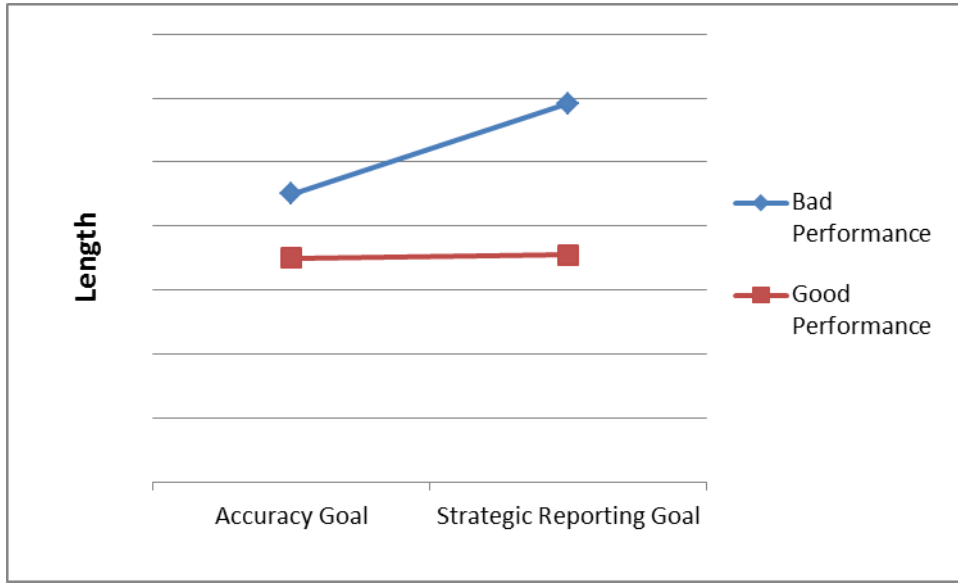


Figure 7. Expected Pattern of Results for the Length Measure of Reporting Complexity

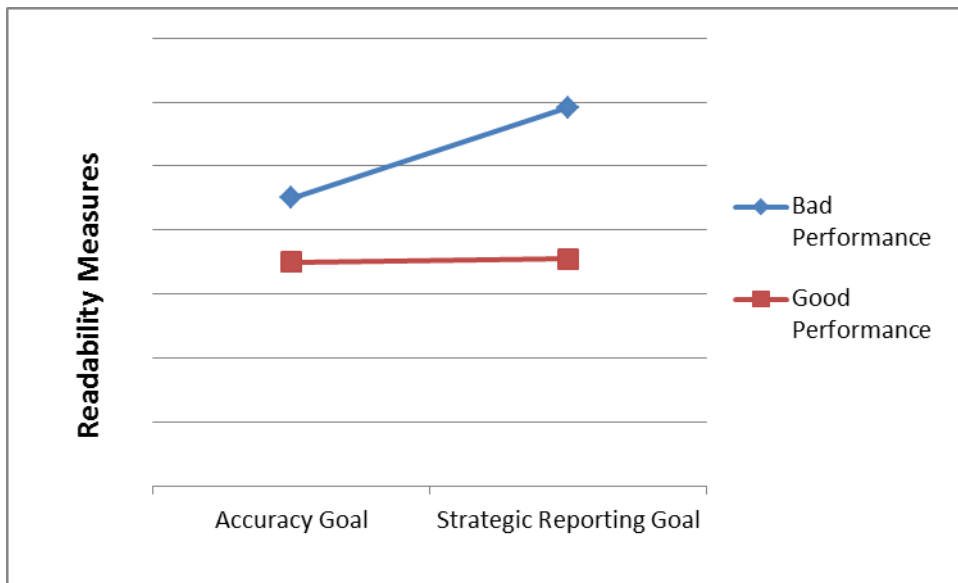


Figure 8. Expected Pattern of Results for the Readability Measures of Reporting Complexity²⁸

²⁸ As discussed in Section 2.3.4, all four readability measures (FOG, Flesch-Kincaid, BOG and READscore) are measured such that *higher* values indicate *lower* readability.

2.3 Experiment

2.3.1 Participants

Participants are 189 individuals recruited from Amazon's Mechanical Turk platform in exchange for a \$1.00 payment.²⁹ Amazon's Mechanical Turk (AMT) is an internet labor market that allows "Requesters" to pay individuals to complete "Human Intelligence Tasks" (HITs), and is an increasingly popular source of experimental data for social scientists (Paolacci, Chandler and Ipeirotis [2010]). Furthermore, studies run on AMT have been shown to reliably replicate a wide range of prior JDM findings (Paolacci et al. [2010]; Horton, Rand and Zeckhauser [2011]). The average participant is 33.77 years old, and has 12.17 years of full time work experience. Participants have completed an average of 1.95 accounting and 1.99 finance courses. On average, participants take 13 minutes to complete the task. Because my study deals with linguistic choices, I specifically recruit participants who (1) live in the United States and (2) consider English to be their native language. This particular sample of participants should have sufficient knowledge to act as managers, read a few facts, and prepare a simple report on the performance of a fictitious division within a firm.

2.3.2 Design and Manipulations

Participants are told to assume that they are the Vice President of the Beverages and Snacks division of Dexico, a hypothetical firm. My experiment uses a 2x2 between-subjects design, manipulating whether the division (1) has performed

²⁹ Initially, 200 participants were recruited. However, six participants were excluded after self-identifying as non-native English speakers despite my recruiting instructions disallowing this. Two additional participants were excluded for not providing a draft report according to the experimental instructions. Finally, three additional participants were excluded for having IP addresses identical to those of participants that had already completed the task. This last exclusion was meant to rule out responses of potential repeat participants. All inferences discussed in Section 2.4 are unchanged if responses from these eleven participants are included.

well or poorly in the most recent quarter, and (2) has or has not yet had its budget set by headquarters for the upcoming year. Participants are also told that the president of their division has asked them to draft a report to headquarters explaining the division's performance.

2.3.3 Task and Procedure

After reading a brief introduction to the task, participants are presented with a short paragraph containing my experimental manipulations. In the good (bad) performance condition, participants are told that the Beverages and Snacks division of Dexico delivered a 10% increase (decrease) in sales this quarter compared to the same quarter last year. They are also told that this performance is better (worse) than all other divisions within the company, and better (worse) than most other firms in the same industry. In the accuracy (strategic reporting) goal condition, participants are told that budgeting decisions have (have not) yet been made by headquarters for the upcoming year. In the accuracy goal condition, participants are told that their only goal is to provide an accurate report, since budgeting decisions have already been made for next year. In the strategic reporting goal condition, participants are told that since budgeting decisions have not yet been made for next year, they should provide an accurate report, but also highlight the positive aspects of the division to secure more funding. In other words, participants in the strategic reporting condition should be more attuned to presenting their division's performance in as favorable a light as possible, particularly when performance is bad.

Upon learning about the division's recent performance and budgeting status, participants are taken to a page explaining that they are about to receive four facts about their division. They are told that they may use as many or as few of the facts as

they would like to prepare their report. In order to encourage participants to write reasonably lengthy and detailed reports, they are (in all conditions) encouraged to embellish details as necessary in order to provide a more coherent explanation to headquarters for the division's performance. To facilitate the embellishment, participants are given an example before continuing on to the main task. Participants consider the hypothetical fact that "The division moved its administrative offices into a new building this quarter." They are told that, if the division performed well, they might describe the fact as "The move to a new office revitalized employees. They were more excited about coming to work, the transition to the new facilities went smoothly, and most employees expressed that the new location made their commute easier." In contrast, participants are told that if the division performed poorly, they might describe the fact as "The move to a new office demoralized employees. They were less excited about coming to work, the transition to the new facilities did not go smoothly, and most employees expressed that the new location made their commute more difficult." Reporting complexity is held as constant as possible for the good and bad performance examples.

After looking over the example, participants proceed to a page containing the actual four facts available for use in their report, and are instructed that the facts are presented in no particular order. To choose the four facts to be included, I started by presenting a larger list of facts to 60 individuals (also recruited from AMT and verified to be different participants than those in the current study). For each fact presented to them, these individuals were asked to rate on a 101-point scale whether the fact was "unambiguously negative" (0) or "unambiguously positive" (100). For the current study, I retained the four facts that did not significantly differ from the

midpoint of 50 on the scale. This was designed to ensure that participants in my study could plausibly use any of the four facts (held constant across all conditions) to explain either good or bad performance in the division.

Below the four facts is a box in which participants can type their draft report (see Figure 9). They are told that they must remain on the page for at least three minutes, in order to give them time to draft the report. After three minutes have passed, a box pops up allowing them to move forward in the study whenever they are ready.

After considering the major factors that might have influenced the performance of the Beverages and Snacks division of Dexico in the most recent quarter, you have come up with the following list of factors that you may wish to discuss in your report. They are listed in no particular order:

- 1) During the quarter the head of Market Research for the division retired, and was replaced by promoting another member of the staff.
- 2) Employee salaries for the division hover right around the average for the industry.
- 3) Two products were discontinued just before the quarter began, and two new products were introduced in their place.
- 4) Transportation costs fluctuated more than usual during the quarter.

In the space below, please write your report describing why your division performed as it did, using one or more of the facts above to support your conclusions. Again, feel free to add embellishments as you see fit in order to create a more coherent story for headquarters. Your response must be at least 300 letters long (i.e., about 75 words) to proceed.

Figure 9. Screen Shot Showing the Four Ambiguous Facts Provided to Participants for the Preparation of their Report

Finally, participants answer manipulation check questions, questions about their report, and demographic questions.

Instead of asking participants to describe the performance of a hypothetical firm, an alternative possibility would have been to ask participants to complete a real performance task and describe the actual outcome. For example, prior accounting studies have asked participants to create rebus puzzles (Kachelmeier, Reichert and Williamson [2008]; Kachelmeier and Williamson [2010]) complete a sandwich-making task (Farrell, Kadous and Towry [2008]), or complete a trivia task (Libby and Rennekamp [2012]). However, performance in these types of tasks is likely to be confounded with intelligence. As a result, if I had used a real performance task in my study, any observed differences in language choices might be driven by the intelligence underlying performance rather than good or bad performance on its own.

2.3.4 Primary Dependent Variables

Prior studies describe disclosures as higher in reporting complexity if they are (1) longer and (2) less readable (e.g., Rutherford [2003], Miller [2010]). My primary dependent variables capture these two constructs.

Length. Length is measured as the number of words that participants provide in their report, and is counted using Stylewriter software.³⁰

Measuring the length of participants' reports is simple, but measuring readability is quite a bit more complex. The increased complexity stems, at least in part, from disagreement in the literature over how best to measure "readability" (Dubay [2004]). The four measures I use, along with their benefits and costs, are described in more detail below.

FOG. The FOG index was developed in order to easily analyze the readability of written language. It is calculated as $((\text{words/sentence}) + 100 * (\text{words with three or$

³⁰ See <http://www.stylewriter-usa.com/> for more information on this software.

more syllables/words))*0.40. The output of the formula is meant to approximate the years of formal education that an individual would need to understand a given text upon a single reading. While the FOG index is easy to calculate, and its use is widespread, some have argued that it is too simple to provide a meaningful measure of readability (Dubay [2004]), particularly in business communications (Jones and Shoemaker [1994]; Miller [2010]; Loughran and McDonald [2011]). For example, multi-syllabic words such as “telecommunication” or “depreciation” would increase measures of the FOG index, although they are unlikely to be difficult to read for most investors.

Flesch-Kincaid. Like the FOG index, the Flesch-Kincaid Grade Level Formula (Flesch-Kincaid) incorporates word length and sentence length to come up with an estimate of the grade level necessary to comprehend a given text. Its formula is calculated as $0.39 * (\text{words/sentence}) + 11.8 * (\text{total syllables/words}) - 15.59$. Like the FOG index, Flesch-Kincaid is praised for its simplicity, but questioned for its ability to provide a meaningful measure of readability (Dubay [2004]). I use both the FOG and Flesch-Kincaid measures in my study because of their widespread use in other domains, with the caveat that they are unlikely to be robust measures of readability.

BOG. My last two measures of readability are more complex, but should better capture actual levels of readability than simple formulas based only on word and sentence length. The BOG index is a proprietary measure of readability developed by creators of Stylewriter software, and includes three components. The first component is “Sentence Bog”, which measures the appropriateness of sentence length for a particular type of document. Stylewriter allows users to specify what type of document is being analyzed (e.g., advertisement, technical report, student essay, etc.). Sentence

Bog is calculated as [Average Sentence Length]² divided by a “Long Sentence Limit” for the type of document specified. The second component of BOG is “Word Bog”, which evaluates the proportion of words that represent common writing style problems (e.g., passive voice, too many abbreviations and acronyms, etc.). The third component of BOG is “Pep”, which measures the use of features that make the reader’s job easier and more enjoyable (e.g., greater sentence variety, personal pronouns, short sentences, etc.). The final BOG index measure is calculated as Sentence Bog + Word Bog – Pep, with all three components adjusted for the linguistic features that are appropriate for a particular type of communication. In my setting, I select Stylewriter’s “Report” option as the type of communication to be analyzed.

READscore. In addition to providing a comprehensive measure of readability with its “BOG index”, Stylewriter allows users to measure certain individual linguistic features of a document. Miller [2010] uses this feature in his analysis of firms’ annual reports. Similarly, I develop my own measure of readability, which includes most of the features of readability that are manipulated in Rennekamp [2012]. My measure of readability is similar to Miller’s [2010] approach, and is calculated as instances of [(Passive Voice + Hidden Verbs + Superfluous words + Negations + Complex Synonyms – Personal Pronouns)*10]/Total words.^{31,32}

While my four measures of readability vary in their components and precision, they are all significantly correlated at $p < 0.10$, with the exception of the correlation between the READscore measure and the Flesch-Kincaid measure ($\rho = 0.08$, $p = 0.302$).

³¹ The number of Negations and Personal Pronouns are calculated using LIWC 2007 software rather than Stylewriter, because Stylewriter does not separately report these two linguistic style features.

³² The two manipulations included in Rennekamp [2012] but not measured in this study are (1) sentence length, and (2) whether the subject, verb and object are kept in order and together. Neither of these two features can be easily measured and incorporated into the READscore calculation.

2.4 Results

On average, participants prepare a report that is approximately 104 words long, with an average of five sentences. The number of facts used by participants in their reports does not vary by condition ($p=0.653$, two-tailed), and there are no significant differences across conditions in participants' ratings of how difficult it was to prepare their report ($p=0.299$, two-tailed). This suggests that my experiment was successful in providing participants with neutral facts upon which to base their reports (i.e., facts which could be interpreted in a positive or negative way).

For my manipulation check questions, 97.88% of participants correctly report whether the division's performance was good or bad in the most recent quarter, while 83.60% of participants correctly report whether their primary reporting goal was to report accurately vs. report in a way that presented their division as favorably as possible.³³

2.4.1 The Main Effect of Performance on Reporting Complexity

H1a and H1b predict that reporting complexity will be greater when participants explain bad performance than when they explain good performance. The components of reporting complexity I investigate are length (H1a) and readability (H1b).

Panel A of Table 4 shows descriptive statistics for my length measure, by condition. Panel B of Table 1 shows that, consistent with H1a, participants provide longer reports when performance is bad than when it is good ($p=0.053$, one-tailed).³⁴

³³ Inferences on the results described in Section 2.4.1 and 2.4.2 are unchanged if I exclude participants who answered manipulation check questions incorrectly.

³⁴ Inferences are unchanged if I instead use a nonparametric Wilcoxon Rank Sum test for the main effect of bad performance on median report length.

Table 4. Descriptive Statistics and Analysis of Variance – Report Length**Panel A. Descriptive Statistics - Mean, (Standard Deviation), and [Median]**

Performance	Primary Goal		
	Accuracy	Reporting	Overall
Good	95.347	102.220	98.478
	(40.052)	(34.711)	(37.664)
	[86]	[92]	[87.5]
	n=49	n=41	n=90
Bad	101.712	118.234	109.556
	(32.721)	(70.751)	(54.549)
	[92.5]	[91]	[92]
	n=52	n=47	n=99
Overall	98.624	110.772	
	(36.418)	(57.143)	
	[90]	[91.5]	
	n=101	n=88	

Panel B. Results of Analysis of Variance (ANOVA)

Source	S.S.	d.f.	M.S.	F-statistic	p-value
Performance	5870.945	1	5870.945	2.649	0.053*
Goal	6416.092	1	6416.092	2.895	0.091
Performance x Goal	1091.616	1	1091.616	0.493	0.484

Panel C. Results of Contrast-Coded Analysis of Variance

Source	S.S.	d.f.	M.S.	F-statistic	p-value
Performance x Goal	9630.946	1	9630.946	4.345	0.019*
Residual	3947.967	2	1973.983	0.891	0.408
Error	410069.225	185	2216.590		

Notes to Table 4. Panel A presents descriptive statistics on the length measure in my experiment, in which 189 participants draft a report explaining the performance of a division in a hypothetical firm. As discussed in Section 2.3.4, the readability measure is calculated such that *higher* values indicate *lower* readability. Panel B presents ANOVA results investigating whether firm performance affects report length. Panel C presents results from a contrast-coded ANOVA. The cells of the experiment receive

contrast weights as follows: Bad Performance/Strategic Goal = +3, Bad Performance/Accuracy Goal = +1, Good Performance/Strategic Goal = -2, and Good Performance/Accuracy Goal = -2. P-values indicated with asterisks (*) are one-tailed, given my directional predictions.

Panel A of Tables 5, 6, 7 and 8 shows descriptive statistics for my four measures of readability: FOG, Flesch-Kincaid, BOG and READscore, respectively. Panel B of Table 5 shows that H1b is not supported with the FOG measure of readability ($p=0.495$, one-tailed). Panel B of Table 6 shows that H1b is also not supported with the Flesch-Kincaid measure of readability ($p=0.451$, one-tailed).

However, as noted in Section 2.3.4, both the FOG and Flesch-Kincaid measures may not be very precise measures of readability, despite their widespread use. This is because of their relatively simple calculations, based on word length and sentence length. Figures 10 and 11 show examples of reports provided by participants that were in the Bottom 10% (“most readable”) and Top 10% (“least readable”) on both the FOG and Flesch-Kincaid measures. Both examples are around 90 words long. While the “less readable” example shown in Figure 11 does use some reasonably large words compared to the “more readable” example shown in Figure 10 (e.g., innovating, groundbreaking, synergy, etc.), they are hardly words that would be considered less readable to the average reader of a business communication.

Table 5. Descriptive Statistics and Analysis of Variance – FOG Measure of Readability

Panel A. Descriptive Statistics - Mean, (Standard Deviation), and [Median]

Performance	Primary Goal		
	Accuracy	Strategic Reporting	Overall
Good	13.311 (2.472) [13.380] n=49	13.067 (2.562) [12.480] n=41	13.186 (2.503) [13.190] n=90
Bad	13.198 (2.668) [12.865] n=52	13.160 (2.877) [12.5] n=47	13.180 (2.755) [12.860] n=99
Overall	13.252 (2.563) [12.940] n=101	13.103 (2.720) [12.490] n=88	

Panel B. Results of Analysis of Variance (ANOVA)

Source	S.S.	d.f.	M.S.	F-statistic	p-value
Performance	0.001	1	0.001	0.000	0.495*
Goal	1.137	1	1.137	0.162	0.688
Performance x Goal	0.653	1	0.653	0.093	0.761

Notes to Table 5. Panel A presents descriptive statistics on the FOG readability measure in my experiment, in which 189 participants draft a report explaining the performance of a division in a hypothetical firm. As discussed in Section 2.3.4, the readability measure is calculated such that *higher* values indicate *lower* readability. Panel B presents ANOVA results investigating whether firm performance affects readability of the report. The reported p-value for the effect of performance (indicated with an asterisk) is one-tailed, given my directional prediction.

Table 6. Descriptive Statistics and Analysis of Variance – Flesch-Kincaid Measure of Readability

Panel A. Descriptive Statistics - Mean, (Standard Deviation), and [Median]

Performance	Primary Goal		
	Accuracy	Strategic Reporting	Overall
Good	11.771 (2.301) [11.910] n=49	11.179 (2.326) [11.090] n=41	11.501 (2.319) [11.350] n=90
Bad	11.477 (2.298) [11.205] n=52	11.557 (2.414) [11.230] n=47	11.515 (2.342) [11.220] n=99
Overall	11.620 (2.293) [11.360] n=101	11.381 (2.368) [11.150] n=88	

Panel B. Results of Analysis of Variance (ANOVA)

Source	S.S.	d.f.	M.S.	F-statistic	p-value
Performance	0.083	1	0.083	0.015	0.451*
Goal	3.087	1	3.087	0.567	0.453
Performance x Goal	5.296	1	5.296	0.972	0.326

Notes to Table 6. Panel A presents descriptive statistics on the Flesch-Kincaid readability measure in my experiment, in which 189 participants draft a report explaining the performance of a division in a hypothetical firm. As discussed in Section 2.3.4, the readability measure is calculated such that *higher* values indicate *lower* readability. Panel B presents ANOVA results investigating whether firm performance affects readability of the report. The reported p-value for the effect of performance (indicated with an asterisk) is one-tailed, given my directional prediction.

“The division performed very well this quarter for a few reasons. First of all, there was a shift in the product lines. Two of our old products were cut and replaced by two new ones. This probably contributed to the division’s performance. Also, we had an important change in leadership. Our old head of Market Research retired and, although he was a good worker, he was replaced by a more energetic member of the staff. He did a great job in accepting his new responsibilities and performed in a stellar manner.”

Figure 10. Sample Report from the Bottom 10% (Most Readable) of both the FOG and Flesch-Kincaid Measures.

“Despite fluctuating transportation costs and the loss of long-time leader Jacob Estleman, the Beverage and Snack team was able to post a 10% sales increase. This was a result of not only innovating with new products, but also increasing staff synergy by replacing Jacob with our own Sally Washington, whose years of experience and close contacts within our team enabled great cooperation and team camaraderie. In the coming year, we plan to continue our groundbreaking work with further research into trending opportunities and forging an industry-leading team.”

Figure 11. Sample Report from the Top 10% (Least Readable) of both the FOG and Flesch-Kincaid Measures.

Panel B of Tables 7 and 8 shows results for the BOG and READscore measures, respectively. Again, these are the two readability scores that I argue are likely to be a better gauge of actual disclosure readability, because they capture more detailed stylistic choices that have been argued to facilitate or impede readers’ processing. Panel B of Table 7 shows that, consistent with H1b, bad performance is

associated with a significantly higher BOG score ($p < 0.001$, one-tailed). The simple main effects of bad performance are also positive and significant in both the Accuracy condition ($p < 0.001$, one-tailed, untabulated) and the Strategic Reporting condition ($p = 0.048$, one-tailed, untabulated). Recall that a higher BOG score indicates writing that is less readable.³⁵

Panel B of Table 8 shows that, also consistent with H1b, bad performance is associated with a significantly higher READscore value ($p < 0.001$, one-tailed). The simple main effects of bad performance are also positive and significant in both the Accuracy condition ($p = 0.006$, one-tailed, untabulated) and the Strategic Reporting condition ($p = 0.002$, one-tailed, untabulated). As with the other measures of readability, a higher READscore value indicates writing that is less readable.

³⁵ Inferences are identical if I instead use a nonparametric Wilcoxon Rank Sum test for the main effect of performance on the BOG measure ($p < 0.001$, one-tailed, untabulated). Inferences are also identical (at $p < 0.001$, one-tailed, untabulated) if I exclude one participant in the Bad Performance/Accuracy condition that has a BOG score of 262, which is considerably higher than any other participants. For reference, the next closest BOG score is 103, and the overall mean [median] for all participants is 57.042 [56].

Table 7. Descriptive Statistics and Analysis of Variance – BOG Measure of Readability

Panel A. Descriptive Statistics - Mean, (Standard Deviation), and [Median]

Performance	Primary Goal		
	Accuracy	Strategic Reporting	Overall
Good	54.245 (18.291) [54.000] n=49	46.610 (22.149) [44.000] n=41	50.767 (20.385) [50.500] n=90
Bad	69.558 (33.177) [66.500] n=52	55.213 (17.943) [55.000] n=47	62.747 (27.853) [63.000] n=99
Overall	62.129 (27.948) [60.000] n=101	51.205 (20.357) [51.500] n=88	

Panel B. Results of Analysis of Variance (ANOVA)

Source	S.S.	d.f.	M.S.	F-statistic	p-value
Performance	6704.889	1	6704.889	11.633	<0.001 *
Goal	5663.432	1	5663.432	9.826	0.002
Performance x Goal	527.763	1	527.763	0.916	0.340

Notes to Table 7. Panel A presents descriptive statistics on the BOG readability measure in my experiment, in which 189 participants draft a report explaining the performance of a division in a hypothetical firm. As discussed in Section 2.3.4, the readability measure is calculated such that *higher* values indicate *lower* readability. Panel B presents ANOVA results investigating whether firm performance affects readability of the report. The reported p-value for the effect of performance (indicated with an asterisk) is one-tailed, given my directional prediction.

Table 8. Descriptive Statistics and Analysis of Variance – READscore Measure of Readability

Panel A. Descriptive Statistics - Mean, (Standard Deviation), and [Median]

Performance	Primary Goal		
	Accuracy	Strategic Reporting	Overall
Good	0.026 (0.409) [0.000] n=49	-0.235 (0.536) [-0.259] n=41	-0.093 (0.486) [-0.078] n=90
Bad	0.277 (0.523) [0.272] n=52	0.072 (0.536) [0.000] n=47	0.180 (0.537) [0.175] n=99
Overall	0.155 (0.485) [0.141] n=101	-0.071 (0.555) [-0.073] n=88	

Panel B. Results of Analysis of Variance (ANOVA)

Source	S.S.	d.f.	M.S.	F-statistic	p-value
Performance	3.653	1	3.653	14.475	<0.001 *
Goal	2.543	1	2.543	10.076	0.002
Performance x Goal	0.036	1	0.036	0.141	0.708

Notes to Table 8. Panel A presents descriptive statistics on the READscore readability measure in my experiment, in which 189 participants draft a report explaining the performance of a division in a hypothetical firm. As discussed in Section 2.3.4, the readability measure is calculated such that *higher* values indicate *lower* readability. Panel B presents ANOVA results investigating whether firm performance affects readability of the report. The reported p-value for the effect of performance (indicated with an asterisk) is one-tailed, given my directional prediction.

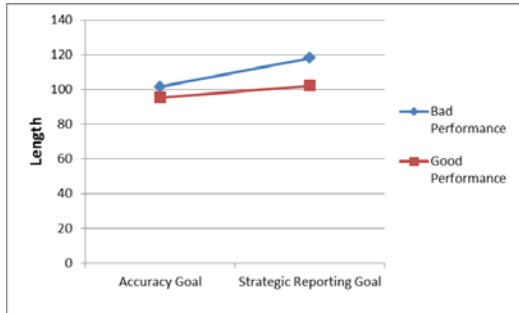
2.4.2 The Interaction between Performance and Goals on Reporting Complexity

While H1a and H1b predict a positive main effect of bad performance on reporting complexity, H2a and H2b predict that reporting complexity will be even higher in the bad performance condition when participants have a strategic reporting goal (i.e., securing more funding for the division), rather than an accuracy goal. In other words, predictions for H2a and H2b are driven by the common argument that managers *intentionally* make bad news more linguistically complex in order to elicit a more favorable, or less negative, reaction to the bad news. Figure 7 depicts the expected pattern of results for the Length measure and the four Readability measures that capture reporting complexity.

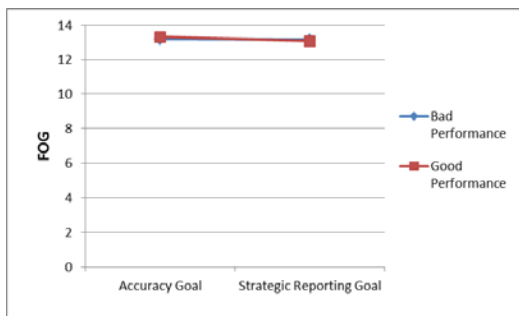
Figure 12 shows that the actual pattern of results looks similar to the expectation for H2a (depicted in Figure 7) for the length measure of reporting complexity. To test the predicted interaction in H2a, I derive contrast weights using the procedure outlined in Buckless and Ravenscroft [1990]. The specific contrast weights are as follows: -2 in the Good Performance/Accuracy Condition, -2 in the Good Performance/Strategic Reporting Condition, +3 in the Bad Performance/Strategic Reporting Condition, and +1 in the Bad Performance/Accuracy Condition. Results of this planned contrast are presented in Panel C of Table 4. As expected, I find support for the predicted interaction ($p=0.019$, one-tailed). I also find that the simple main effect of goal is significant for the bad performance condition ($p=0.041$, one-tailed, untabulated), but not for the good performance condition ($p=0.246$, one-tailed, untabulated). Thus, H2a is supported.

Figure 12. Actual Pattern of Results for the Primary Dependent Variables

Length Measure



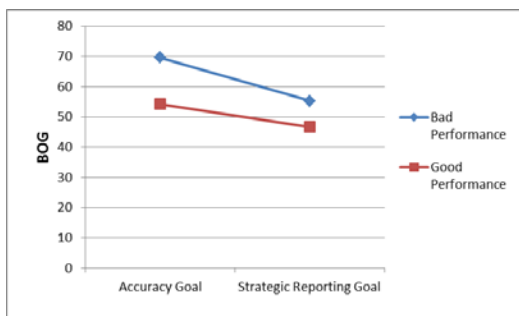
FOG Measure of Readability



F-K Measure of Readability



BOG Measure of Readability



READscore Measure of Readability

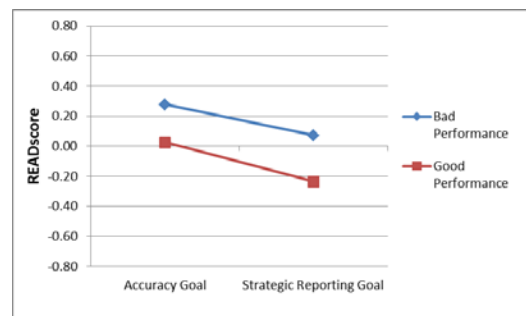


Figure 12 also shows that the actual pattern of results for each of the four measures of readability does not look similar to the expectation for H2b depicted in Figure 7. Likewise, there is no significant interaction between performance and goals

for any of the four measures. Panel B of Tables 5, 6, 7 and 8 shows that the Performance x Goal interaction has a p-value greater than 0.10 for the FOG, Flesch-Kincaid, BOG and READscore measures, respectively. Thus, H2b is not supported for any of the four measures of readability. However, I do find an unexpected main effect of the goal manipulation on both the BOG and READscore measures of readability. The strategic reporting goal leads participants to provide significantly *more* readable reports than does the accuracy goal, as measured using both the BOG ($p=0.001$, one-tailed) and READscore ($p=0.001$, one-tailed) measures. For the BOG measure, this holds for both the good performance condition ($p=0.067$, one-tailed), and the bad performance condition ($p=0.002$, one-tailed). Likewise, for the READscore measure, this holds for both the good performance condition ($p=0.008$, two-tailed) and the bad performance condition ($p=0.022$, two-tailed).

Combined, the results do not appear to support arguments regarding intentional obfuscation of bad performance. Rather, when participants are asked to explain bad performance and have a strategic reporting goal to elicit a favorable reaction, they use significantly *more* readable language. While H2a is supported, and bad performance is accompanied by more lengthy descriptions when there is a goal of eliciting a favorable reaction, the lack of results for H2b suggest an alternative interpretation of H2a. Rather than providing more lengthy disclosures to obfuscate bad performance, it is possible that participants provide more lengthy disclosures in order to provide a more comprehensive explanation for poor performance, particularly given that participants increase the readability of these disclosures. This suggests that future researchers may want to reconsider using lengthiness of disclosures as an indicator of reporting complexity.

2.4.3 Supplemental Analyses

In this section I report some additional evidence on how performance and goals affect certain linguistic choices in my experimental setting. One component of the READscore measure is the use of personal pronouns (e.g., “I” or “we”). Style guides often suggest that personal pronouns increase the readability of a written communication (see, for example, the SEC’s *Plain English Guidelines*).³⁶ In addition, Hyland [2005] argues that greater use of personal pronouns is an indication that an individual is associating themselves more with a given message. Looking only at personal pronouns, I find that participants use significantly fewer personal pronouns in their reports when performance is bad than when it is good ($p=0.013$, one-tailed, untabulated), regardless of their reporting goals. This suggests that participants may want to distance themselves from bad performance news, consistent with Hyland’s arguments [2005]. I also find that participants use significantly *more* personal pronouns when they have a goal of presenting themselves as favorably as possible, but only when performance is good ($p=0.050$, one-tailed, untabulated) rather than bad ($p=0.164$, one-tailed-untabulated). Hyland [2001] argues that personal pronouns indicate greater commitment to a given message. Thus the greater use of personal pronouns when individuals have a strategic reporting goal and performance is good might indicate that participants use personal pronouns to display commitment to the message and to make their argument more convincing. Consistent with this, Asay, Libby and Rennekamp [2012] demonstrate that the use of personal pronouns can increase the persuasiveness of management communications, even for relatively short disclosures.

³⁶ Prior accounting studies have also treated the use of personal pronouns as indicators of attribution bias (Li [2010b]) and narcissism (Chatterjee and Hambrick [2007]).

Finally, I find that the accuracy goal leads to greater use of the past tense in participants' reports ($p=0.004$, one-tailed, untabulated), regardless of whether performance is good ($p=0.016$, one-tailed, untabulated) or bad ($p=0.048$, one-tailed, untabulated). Alternatively, the strategic reporting goal to secure more funding leads to greater use of the future tense in participants' reports ($p=0.062$, one-tailed, untabulated), but only when performance is bad ($p=0.046$, one-tailed, untabulated) rather than good ($p=0.302$, one-tailed, untabulated). This latter finding also suggests that, when performance is bad, participants shift away from using the past tense and towards using the future tense as their goal shifts from providing an accurate report to providing a report that serves a strategic reporting objective. Future research might further investigate how managers use linguistic characteristics to either (1) distance from or associate themselves with a message, or (2) shift investors' attention away from less favorable information.

2.5 Conclusion

This study serves as a starting point for understanding the fundamental question of how performance and explicit goals affect managers' choice of reporting complexity in qualitative disclosures. Using an experiment, I investigate linguistic choices in a controlled setting. I predict and find that participants provide reports that are significantly longer and less-readable when performance is bad than when performance is good. This finding holds even when participants are told that their only goal is to accurately convey performance, and is consistent with Bloomfield's [2008] suggestion that poor performance may be inherently more difficult to explain concisely and in more readable language. However, I do not find that poor performance interacts with goals in the predicted manner. Arguments in prior

accounting literature (particularly Li [2008]) suggest that reporting complexity will be greatest when performance is bad and participants are given explicit instructions to prepare a report that will help the division secure future funding (i.e., a strategic reporting goal). While this goal does lead participants to provide significantly longer reports, the reports are significantly *more* readable than when participants are only given a goal of reporting accurately. Combined, the results do not appear to support arguments regarding intentional obfuscation of bad performance.

In addition to the results discussed above, I also provide some preliminary evidence on how other linguistic choices are affected by variation in performance and goals. I find that participants use more personal pronouns when performance is good than when performance is bad, particularly when they have a strategic reporting goal. This is consistent with prior work suggesting that individuals can use personal pronouns to associate themselves with a message (Hyland [2005]) and make it more persuasive (Asay, Libby and Rennekamp [2012]). I also find that, when performance is bad, participants use more past tense language when they have an accuracy goal, and more future tense language when they have a goal of creating a favorable impression. Future experimental research might investigate additional linguistic choices that have been shown to reflect individuals' motives and circumstances (e.g. linguistic tone, causal explanations, etc.). Future research might also investigate both written and spoken settings of managerial communication (e.g. annual reports, interviews, conference calls, etc.).

This study is not without some important caveats. First of all, my participants are individuals who lack real-world experience preparing financial disclosures. However, differences in reporting complexity and abstractness that emerge in response

to firm performance are expected to be driven by pervasive psychological phenomena, so the use of non-professionals as participants is appropriate (Libby, Bloomfield and Nelson [2002]). Furthermore, I argue that my participants have sufficient knowledge to understand the manipulation of managers' incentives in my experimental task. A second concern is that managers' actual disclosures in the real world may be additionally influenced through consultation with legal departments, public-relations professionals, etc. Thus, this study serves as a starting point for understanding the psychological and incentive-based drivers of qualitative disclosure decisions. It does not, however, strictly capture the final disclosures that might actually be provided to investors.

APPENDIX A – EXAMPLES OF *PLAIN ENGLISH HANDBOOK* SUGGESTIONS

PANEL A. LINGUISTIC FEATURES INCLUDED IN THE EXPERIMENTAL MANIPULATION OF READABILITY		
Features that increase readability	Example of a Less Readable Disclosure	Example of a More Readable Disclosure
Short sentences	Stephen Miller, Chief Executive Officer, stated, “Sales were above plan, earnings were higher than expectations, the Company engaged in the execution of several initiatives during the third quarter that have it better positioned for the future, and the Company’s business strategy is to increase sales by expanding distribution of its brands in new and existing markets, which is intended to raise consumer awareness and trial of its products, thus leading to increased relevance and purchase intent.”	Chief Executive Officer, Stephen Miller, stated, “Our sales were above plan, our earnings were higher than we expected, and we executed several initiatives during the third quarter that have us better positioned for the future. Our business strategy is to increase sales by expanding distribution of our brands in new and existing markets, raising consumer awareness and trial of our products, and increasing relevance and purchase intent.”
Active Voice	... which is intended to raise consumer awareness and trial of our products, thus leading to increased relevance and purchase intent.	... raising consumer awareness and trial of our products, and increasing relevance and purchase intent.
No Hidden Verbs	While sales were below plan and earnings were lower than expectations , the Company engaged in the execution of several initiatives during the third quarter that have it better positioned for the future.	While sales were below plan and earnings were lower than expected , we executed several initiatives during the third quarter that have us better positioned for the future.
No Superfluous Words	The decrease in revenue was primarily because of the fact that there was a decrease in total case sales of 34.5 percent to 2.9 million cases.	The decrease in revenue was primarily due to a decrease in total case sales of 34.5% to 2.9 million cases.
Write in the Positive	The Company is not satisfied with what has been accomplished so far in fiscal 2009.	We are dissatisfied with what has been accomplished so far in fiscal 2009.
Simple Synonyms	The Company is focused on escalating distribution of its products, building and maintaining good relationships with key distributors; and creating pioneering brands of beverages and products.	The Company is focused on expanding distribution of its products, building and maintaining good relationships with key distributors; and creating innovative brands of beverages and products.
Personal Pronouns	The recent strength of the overall economy and financial markets has positively impacted the Company’s two primary markets: the U.S. and Canada. This has not decreased consumer confidence in the economy and the Company believes has positively affected consumers’ willingness to purchase its products as they augment discretionary spending.	The recent strength of the overall economy and financial markets has positively impacted our two primary markets: the U.S. and Canada. This has increased consumer confidence in the economy and we believe has positively affected consumers’ willingness to purchase our products as they grow their discretionary spending.
Keep Subject-Verb- and Object in order and together.	The table below summarizes revenue and earnings performance for the quarter ended September 30, 2009, as compared to September 30, 2008, for Cooper Soda Co (NASDAQ: CSOD), a leader in the premium soda category and known for its unique branding and innovating marketing, which today announced results for the quarter ended September 30, 2009.	Cooper Soda Co. (NASDAQ: CSOD), a leader in the premium soda category and known for its unique branding and innovative marketing, today announced results for the quarter ended September 30, 2009. The table below summarizes revenue and earnings performance for the quarter ended September 30, 2009, as compared to September 30, 2008.

PANEL B. FORMATTING FEATURES INCLUDED IN MY EXPERIMENTAL MANIPULATION OF READABILITY														
Features that increase readability	Example of a Less Readable Disclosure	Example of a More Readable Disclosure												
Clear Headings/ Hierarchy	COOPER SODA CO. REPORTS... Portland, OR – November 5, 2009 ... Outlook	COOPER SODA CO. REPORTS <i>Portland, OR – November 5, 2009</i> ... <i>Outlook</i>												
Appropriate Layout	Cooper Soda Co. (NASDAQ: CSOD), a leader in the premium soda category and known for its unique branding and innovative marketing, today announced results for the quarter ended September 30, 2009.	Cooper Soda Co. (NASDAQ: CSOD), a leader in the premium soda category and known for its unique branding and innovative marketing, today announced results for the quarter ended September 30, 2009.												
Tables¹	The Company acquired Division A from Company X in 1986, Division B from Company Y in 1991, and Division C from Company Z in 1998.	The table below highlights the Company's Acquisitions: <table border="1"> <thead> <tr> <th>Division:</th> <th>Acquired From:</th> <th>Year of Acq.:</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>X</td> <td>1986</td> </tr> <tr> <td>B</td> <td>Y</td> <td>1991</td> </tr> <tr> <td>C</td> <td>Z</td> <td>1998</td> </tr> </tbody> </table>	Division:	Acquired From:	Year of Acq.:	A	X	1986	B	Y	1991	C	Z	1998
Division:	Acquired From:	Year of Acq.:												
A	X	1986												
B	Y	1991												
C	Z	1998												
Bullet Points	[The Company] is focused on escalating distribution of its products, building and maintaining good relationships with key distributors; and creating pioneering brands of beverages and products.	We are focused on: <ul style="list-style-type: none"> • Expanding distribution of our products; • Building and maintaining good relationships with our key distributors; and • Creating innovative beverage brands and products. 												

¹This example is not from my experiment. I use a smaller table as an example in this Appendix because of space limitations. See Appendices B and C for my actual manipulation of the use of tables.

APPENDIX B – PRESS RELEASES FROM MORE READABLE CONDITIONS

Panel A. Good News/More Readable Condition

COOPER SODA CO. REPORTS STRONG THIRD QUARTER 2010 RESULTS

Portland, OR – August 5, 2010 – Cooper Soda Co. (NASDAQ: CSOD), a leader in the premium soda category and known for its unique branding and innovative marketing, today announced results for the quarter ended June 30, 2010.

The table below summarizes revenue and earnings performance for the quarter ended June 30, 2010, as compared to June 30, 2009.

	Quarter ended June 30, 2009	Quarter ended June 30, 2010	Change
Revenues	\$18.2 million	\$24.1 million	+ \$5.9 million
Net Profit	\$3.2 million	\$4.7 million	+ \$1.5 million

The increase in revenue was primarily due to an increase in total case sales.

Outlook

Chief Executive Officer Stephen Miller stated, "Our sales were above plan, our earnings were higher than we expected, and we executed several initiatives during the third quarter that have us better positioned for the future. Our business strategy is to increase sales by expanding distribution of our brands in new and existing markets, raising consumer awareness and trial of our products, and increasing relevance and purchase intent." We are focused on:

- *Expanding distribution of our products;*
- *Building and maintaining good relationships with our key distributors; and*
- *Creating innovative beverage brands and products.*

The beverage industry, and particularly those companies selling premium beverages like ours, can be affected by macroeconomic factors including changes in economic conditions and consumer spending patterns. The recent improvement in the overall economy has positively impacted our two primary markets: the U.S. and Canada. This has increased consumer confidence and we believe has positively affected consumers' willingness to purchase our products as they grow their discretionary spending.

According to Miller, "We are satisfied with our accomplishments so far in fiscal 2010, and expect further improvement in operations for fourth quarter and annual results".

Panel B. Bad News/More Readable Condition

COOPER SODA CO. REPORTS WEAK THIRD QUARTER 2010 RESULTS

Portland, OR – August 5, 2010 – Cooper Soda Co. (NASDAQ: CSOD), a leader in the premium soda category and known for its unique branding and innovative marketing, today announced results for the quarter ended June 30, 2010.

The table below summarizes revenue and earnings performance for the quarter ended June 30, 2010, as compared to June 30, 2009.

	Quarter ended June 30, 2009	Quarter ended June 30, 2010	Change
Revenues	\$30.0 million	\$24.1 million	- \$5.9 million
Net Profit	\$6.2 million	\$4.7 million	- \$1.5 million

The decrease in revenue was primarily due to a decrease in total case sales.

Outlook

Chief Executive Officer Stephen Miller stated, "While our sales were below plan and our earnings were lower than we expected, we executed several initiatives during the third quarter that have us better positioned for the future. Our business strategy is to increase sales by expanding distribution of our brands in new and existing markets, raising consumer awareness and trial of our products, and increasing relevance and purchase intent." We are focused on:

- *Expanding distribution of our products;*
- *Building and maintaining good relationships with our key distributors; and*
- *Creating innovative beverage brands and products.*

The beverage industry, and particularly those companies selling premium beverages like ours, can be affected by macroeconomic factors including changes in economic conditions and consumer spending patterns. The recent weakness in the overall economy has negatively impacted our two primary markets: the U.S. and Canada. This has decreased consumer confidence and we believe has negatively affected consumers' willingness to purchase our products as they slow their discretionary spending.

According to Miller, "We are dissatisfied with our accomplishments so far in fiscal 2010, and expect further deterioration in operations for fourth quarter and annual results".

APPENDIX C – PRESS RELEASES FROM LESS READABLE CONDITIONS

Panel A. Good News/Less Readable Condition

COOPER SODA CO. REPORTS STRONG Q3 2010 RESULTS

Portland, OR – August 5, 2010 – The information below summarizes revenue and earnings performance for the quarter ended June 30, 2010, as compared to June 30, 2009, for Cooper Soda Co. (NASDAQ: CSOD), a leader in the premium soda category and known for its unique branding and innovative marketing, which today announced results for the quarter ended June 30, 2010.

For the quarter ended June 30, 2010, revenue was \$24.1 million, an increase of \$5.9 million from \$18.2 million in revenue for the quarter ended June 30, 2009. Net profit for the quarter ended June 30, 2010 was \$4.7 million, an increase of \$1.5 million from a net profit of \$3.2 million for the quarter ended June 30, 2009. The increase in revenue was primarily because of the fact that there was an increase in total case sales.

Outlook

Stephen Miller, Chief Executive Officer, stated, “Sales were above plan, earnings were higher than expectations, the Company engaged in the execution of several initiatives during the third quarter that have it better positioned for the future, and the Company’s business strategy is to increase sales by expanding distribution of its brands in new and existing markets, which is intended to raise consumer awareness and trial of its products, thus leading to increased relevance and purchase intent.” The Company is focused on escalating distribution of its products, building and maintaining good relationships with key distributors; and creating pioneering brands of beverages and products.

The beverage industry, and particularly those companies selling first-rate beverages like Cooper, can be affected by macroeconomic factors including changes in economic conditions and consumer spending patterns. The recent improvement in the overall economy has positively impacted the Company’s two primary markets: the U.S. and Canada. This has not decreased consumer confidence and the Company believes has positively affected consumers’ willingness to purchase its products as they augment discretionary spending. According to Miller, “The Company is not dissatisfied with what has been accomplished so far in fiscal 2010, and expects further improvement in operations for Q4 and annual results”.

Panel B. Bad News/Less Readable Condition

COOPER SODA CO. REPORTS WEAK Q3 2010 RESULTS

Portland, OR— August 5, 2010— The information below summarizes revenue and earnings performance for the quarter ended June 30, 2010, as compared to June 30, 2009, for Cooper Soda Co. (NASDAQ: CSOD), a leader in the premium soda category and known for its unique branding and innovative marketing, which today announced results for the quarter ended June 30, 2010.

For the quarter ended June 30, 2010, revenue was \$24.1 million, a decrease of \$5.9 million from \$30.0 million in revenue for the quarter ended June 30, 2009. Net profit for the quarter ended June 30, 2010 was \$4.7 million, a decrease of \$1.5 million from a net profit of \$6.2 million for the quarter ended June 30, 2009. The decrease in revenue was primarily because of the fact that there was a decrease in total case sales.

Outlook

Stephen Miller, Chief Executive Officer, stated, "While sales were below plan and earnings were lower than expectations, the Company engaged in the execution of several initiatives during the third quarter that have it better positioned for the future, and the Company's business strategy is to increase sales by expanding distribution of its brands in new and existing markets, which is intended to raise consumer awareness and trial of its products, thus leading to increased relevance and purchase intent." The Company is focused on escalating distribution of its products; building and maintaining good relationships with key distributors; and creating pioneering brands of beverages and products.

The beverage industry, and particularly those companies selling first-rate beverages like Cooper, can be affected by macroeconomic factors including changes in economic conditions and consumer spending patterns. The recent weakness in the overall economy has negatively impacted the Company's two primary markets: the U.S. and Canada. This has not increased consumer confidence and the Company believes has negatively affected consumers' willingness to purchase its products as they diminish discretionary spending. According to Miller, "The Company is not satisfied with what has been accomplished so far in fiscal 2010, and expects further deterioration in operations for Q4 and annual results".

REFERENCES

- ALTER, A. L., and OPPENHEIMER, D. M. 'Predicting Short-Term Stock Fluctuations by Using Processing Fluency.' *Proceedings of the National Academy of Sciences* 103 (2006): 9369-9372.
- ALTER, A., D. OPPENHEIMER, N. EPLEY, and R. EYRE. 'Overcoming Intuition: Metacognitive Difficulty Activates Analytic Reasoning.' *Journal of Experimental Psychology* 136 (2007): 569-576.
- ALTER, A., and D. OPPENHEIMER. 'Uniting the Tribes of Fluency to Form a Metacognitive Nation.' *Personality and Social Psychology Review* 13 (2009): 219-235.
- ASAY, S., R. LIBBY and K. RENNEKAMP. 'How Subtle Language Choices Affect the Interpretation of Qualitative Accounting Disclosures.' Unpublished paper, 2012.
- BARON, R. M., and D. A. KENNY. 'The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations.' *Journal of Personality and Social Psychology* 51 (1986): 1173-1182.
- BARTON, J., and M. MERCER. 'To Blame or Not to Blame: Analysts' Reactions to External Explanations for Poor Financial Performance.' *Journal of Accounting and Economics* 39 (2005): 509-533.
- BIDDLE, G.C., G. HILARY, and R.S. VERDI. 'How does Financial Reporting Quality Relate to Investment Efficiency?' *Journal of Accounting and Economics* 48 (2009): 112-131.
- BLOOMFIELD, R. 'The 'Incomplete Revelation Hypothesis' and Financial Reporting.' *Accounting Horizons* 16 (2002): 233-243.
- BLOOMFIELD, R. 'Discussion of 'Annual Report Readability, Current Earnings and Earnings Persistence'.' *Journal of Accounting and Economics* 45 (2008): 248-252.
- BLOOMFIELD, R., and K. RENNEKAMP. 'Experimental Research on Financial Reporting: From the Laboratory to the Virtual World.' *Foundations and Trends in Accounting* 3 (2009): 135-221.
- BRENNAN, N., GUILLAMON-SAORIN, E. and A. PIERCE. 'Impression Management: Developing and Illustrating a Scheme of Analysis for Narrative Disclosures - A Methodological Note.' *Accounting, Auditing & Accountability Journal* 22 (2009): 789-832.

- BRIÑOL, P., R. PETTY, and Z. TORMALA. 'The Malleable Meaning of Subjective Ease.' *Psychological Science* 17 (2006): 200-206.
- BUCKLESS, F.A., and S.P. RAVENSCROFT. 'Contrast Coding: A Refinement of ANOVA in Behavioral Analysis.' *The Accounting Review* 65 (1990): 933-945.
- CHATTERJEE, A., and D. HAMBRICK. 'It's All About Me: Narcissistic Chief Executive Officers and their Effects on Company Strategy and Performance.' *Administrative Science Quarterly* 52 (2007): 351-386.
- CLATWORTHY, M., and M. JONES. 'The Effect of Thematic Structure on the Variability of Annual Report Readability.' *Accounting, Auditing and Accountability Journal* 14 (2001): 311-326.
- CLOR-PROELL, S.M. 'The Effects of Expected and Actual Accounting Choices on Judgments and Decisions.' *The Accounting Review* 84 (2009): 1465-1493.
- CORE, J. 'A Review of the Empirical Disclosure Literature: A Discussion.' *Journal of Accounting and Economics* 31 (2001): 441-456.
- COURTIS, J.K. 'An Investigation into Annual Report Readability and Corporate Risk Return Relationships.' *Accounting and Business Research* 16 (1986): 285-294.
- COURTIS, J.K. 'Annual Report Readability: Tests of the Obfuscation Hypothesis.' *Accounting, Auditing and Accountability Journal* 11 (1998): 459-471.
- DAFT, R.L., and N.B. MACINTOSH. 'A Tentative Exploration into the Amount and Equivocality of Information Processing in Organizational Work Units.' *Administrative Science Quarterly* 26 (1981): 207-224.
- DAFT, R.L., R.H. LENGEL, and L.K. TREVINO. 'The Relationship between Message Equivocality, Media Selection, and Manager Performance: Implications for Information Support Systems.' *Management Information Science Quarterly* 11 (1987): 355-356.
- DAVIS, A., J. PIGER, and L. SEDOR. 'Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language.' Unpublished paper, 2011. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=875399
- DAVIS, A., and I. TAMA-SWEET. 'Managers' Use of Language across Alternative Disclosure Outlets: Earnings Press Releases versus MD&A.' Unpublished paper, 2011. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1866369
- DEMERS, E., and C. VEGA. 'Soft Information in Earnings Announcements: News or Noise?' Unpublished paper, 2010. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1153450

- DUBAY, W. *The Principles of Readability*. Costa Mesa, California: Impact Information, 2004.
- ELLIOTT, W.B., F. HODGE, J. KENNEDY, and M. PRONK. 'Are M.B.A. Students a Good Proxy for Nonprofessional Investors?' *The Accounting Review* 82 (2007): 139–168.
- ELLIOTT, W.B., F. HODGE, and L. SEDOR. 'Using Online Video to Announce a Restatement: Influences on Investment Decisions and the Mediating Role of Trust.' *The Accounting Review* 87 (2012): 513-535.
- FARRELL, A., K. KADOUS, and K. TOWRY. 'Contracting on Contemporaneous vs. Forward-Looking Measures: An Experimental Investigation.' *Contemporary Accounting Research* 25 (2008): 773-802.
- FELDMAN, R., S. GOVINDARAJ, J. LIVNAT, and B. SEGAL. 'Management's Tone Change, Post Earnings Announcement Drift and Accruals.' *Review of Accounting Studies* 15 (2009): 915-953.
- GEE, J.P. *An Introduction to Discourse Analysis: Theory and Method*. London, England: Routledge, 1999.
- GROSSMAN, S., and J. STIGLITZ. 'On the Impossibility of Informationally Efficient Markets.' *American Economic Review* 70 (1980): 393-407.
- GUNSCH, M. A., S. BROWNLOW, S.E. HAYNES, and Z. MABE. 'Differential Linguistic Content of Various Forms of Political Advertising.' *Journal of Broadcasting & Electronic Media* 44 (2000): 27-42.
- HAFNER, M., and D. STAPEL. 'Information to Go: Fluency Enhances the Usability of Primed Information.' *Journal of Experimental Social Psychology* 46 (2010): 73-84.
- HALES, J., X. KUANG and S. VENKATARAMAN. 'Who Believes the Hype? An Experimental Investigation of How Language Affects Investor Judgments.' *Journal of Accounting Research* 49 (2011): 223-255.
- HAYES, A. 'Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium.' *Communication Monographs* 76 (2009): 408-420.
- HEIMAN, V. B. 'Auditors' Assessments of the Likelihood of Error Explanations in Analytical Review.' *The Accounting Review* 65 (1990): 875-890.
- HENRY, E. 'Are Investors Influenced by How Earnings Press Releases are Written?' *Journal of Business Communication* 45 (2008): 363-407.

- HIRSHLEIFER, D. and H.S. TEOH. 'Limited Attention, Information Disclosure and Financial Reporting.' *Journal of Accounting and Economics* 36 (2003): 337-386.
- HOBSON, J., W. MAYEW, and M. VENKATACHALAM. 'Analyzing Speech to Detect Financial Misreporting.' *Forthcoming in the Journal of Accounting Research, 2011.*
- HOPKINS, P. 'The Effect of Financial Statement Classification of Hybrid Financial Instruments on Financial Analysts' Stock Price Judgments.' *Journal of Accounting Research* 34 (1996): 33-50.
- HORTON, J., and L.B. CHILTON. 'The Labor Economics of Paid Crowdsourcing.' *Proceedings of the 11th ACM Conference on Electronic Commerce.* Cambridge, Massachusetts, 2010.
- HORTON, J., D. RAND, and R. ZECKHAUSER. 'The Online Laboratory: Conducting Experiments in a Real Labor Market.' *Experimental Economics* 14 (2011): 1-27.
- HYLAND, K. 'Humble Servants of the Discipline? Self-Mention in Research Articles.' *English for Special Purposes* 20 (2001): 207-226.
- HYLAND, K. *Metadiscourse: Exploring Interaction in Writing.* Oxford, England: Continuum, 2005.
- JONES, M.J. 'A Longitudinal Study of the Readability of the Chairman's Narratives in the Corporate Reports of a UK Company.' *Accounting and Business Research* 18 (1988): 297-305.
- JONES, M.J., and P.A. SHOEMAKER. 'Accounting Narratives: A Review of Empirical Studies of Content and Readability.' *Journal of Accounting Literature* 13 (1994): 142-184.
- KACHELMEIER, S., B. REICHERT, and M. WILLIAMSON. 'Measuring and Motivating Quantity, Creativity, or Both.' *Journal of Accounting Research* 46 (2008): 341-373.
- KACHELMEIER, S., and M. WILLIAMSON. 'Attracting Creativity: The Initial and Aggregate Effects of Contract Selection on Creativity-Weighted Productivity.' *The Accounting Review* 85 (2010): 1669-1691.
- KADOUS, K., S. KRISCHE, and L. SEDOR. 'Using Counter-Explanation to Limit Analysts' Forecast Optimism.' *The Accounting Review* 81 (2006): 377-397.
- KAHNEMAN, D., and A. TVERSKY. 'On the Reality of Cognitive Illusions.' *Psychological Review* 103 (1996): 582-591.

- KALAY, A. 'Investor Sophistication, Disclosure, and the Information Environment of the Firm.' Unpublished paper, 2011. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1762809
- KENNEDY, J. 'Debiasing the Curse of Knowledge in Audit Judgment.' *The Accounting Review* 70 (1995): 249-273.
- KENNY, D. A., D. A. KASHY, and N. BOLGER. 'Data Analysis in Social Psychology' in *The Handbook of Social Psychology Vol. 1*, edited by D. Gilbert, S. Fiske, and G. Lindzey, 233-265. Boston, MA. McGraw-Hill, 1998.
- KOONCE, L. 'Explanation and Counterexplanation during Audit Analytical Review.' *The Accounting Review* 67 (1992): 59-76.
- KOONCE, L., and M. LIPE. 'Earnings Trend and Performance Relative to Benchmarks: How Consistency Influences Their Joint Use.' *Journal of Accounting Research* 48 (2010): 859-884.
- KRISCHE, S.D. 'Investors' Evaluations of Strategic Prior Period Benchmark Disclosures in Earnings Announcements.' *The Accounting Review* 80 (2005): 243-268.
- LABROO, A., and S. KIM. 'The 'Instrumentality Heuristic': Why Metacognitive Difficulty is Desirable During Goal Pursuit.' *Psychological Science* 20 (2009): 127-134.
- LEHAVY, R., F. LI, and K. MERKLEY. 'The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts.' *The Accounting Review* 86 (2011): 1087-1115.
- LI, F. 'Annual Report Readability, Current Earnings, and Earnings Persistence.' *Journal of Accounting and Economics* 45 (2008): 221-247.
- LI, F. 'The Information Content of Forward-Looking Statements in Corporate Filings – A Naïve Bayesian Machine Learning Approach.' *Journal of Accounting Research* 48 (2010a): 1049-1102.
- LI, F. 'Managers' Self-Serving Attribution Bias and Corporate Financial Policies.' Unpublished paper, The University of Michigan, 2010b. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1639005
- LI, F. 'Textual Analysis of Corporate Disclosures: A Survey of the Literature.' *Forthcoming in the Journal of Accounting Literature, 2011.*
- LIBBY, R., R. BLOOMFIELD, and M. NELSON. 'Experimental Research in Financial Accounting.' *Accounting, Organizations and Society* 27 (2002): 775-810.

- LIBBY, R., and N. SEYBERT. 'Behavioural Studies of the Effects of Regulation on Earnings Management and Accounting Choice.' In C. Chapman, D. Cooper and P. Miller (Eds.), *Accounting Organizations and Society: Essays in Honour of Anthony Hopwood* (290-314). Oxford, England: Oxford University Press, 2009.
- LIBBY, R., and K. RENNEKAMP. 'Self-Serving Attribution Bias, Overconfidence, and the Issuance of Voluntary Disclosures.' *Journal of Accounting Research* 50 (2012): 197-231.
- LOUGHRAN, T., and B. MCDONALD. 'Measuring Readability in Financial Disclosures.' Unpublished paper, 2011. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1920411
- MACKINNON, D., J. KRULL, and C. LOCKWOOD. 'Equivalence of the Mediation, Confounding and Suppression Effect.' *Prevention Science* 1 (2000): 173-181.
- MAYEW, W., and M. VENKATACHALAM. 'The Power of Voice. Managerial Affective States and Future Firm Performance.' *Journal of Finance* 67 (2012): 1-44.
- MCGLONE, M. S., and J. TOFIGHBAKHS. 'Birds of a Feather Flock Conjointly (?): Rhyme as Reason in Aphorisms.' *Psychological Science* 11 (2000): 424-428.
- MERCER, M. 'How Do Investors Assess the Credibility of Management Disclosures?' *Accounting Horizons* 18 (2004): 185-196.
- MERCER, M. 'The Fleeting Effects of Disclosure Forthcomingness on Management's Reporting Credibility.' *The Accounting Review* 80 (2005): 723-744.
- MERKL-DAVIES, D.M., and N. BRENNAN. 'Discretionary Disclosure Strategies in Corporate Narratives: Incremental Information or Impression Management?' *Journal of Accounting Literature* 26 (2007): 116-194.
- MILLER, B. 'The Effects of Reporting Complexity on Small and Large Investor Trading.' *The Accounting Review* 85 (2010): 2107-2143.
- MOFFITT, K., and M. BURNS. 'What Does that Mean? Investigating Obfuscation and Readability Cues as Indicators of Deception in Fraudulent Reports.' San Francisco, California: Proceedings of the Fifteenth Americas Conference on Information Systems, 2009.
- NELSON, K., and A. PRITCHARD. 'Litigation Risk and Voluntary Disclosure: The Use of Meaningful Cautionary Language.' Unpublished paper, 2007. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=998590

- NOVEMSKY, N., R. DHAR, N. SCHWARZ, and I. SIMONSON. 'Preference Fluency in Choice.' *Journal of Marketing Research* 44 (2007): 347-356.
- NUNNALLY, J. *Psychometric Theory*. 2nd edition. New York, NY: McGraw-Hill, 1978.
- OPPENHEIMER, D. 'Consequences of Erudite Vernacular Utilized Irrespective of Necessity: Problems with Using Long Words Needlessly.' *Applied Cognitive Psychology* 20 (2006):139-156.
- PAOLACCI, G., J. CHANDLER and P. IPEIROTIS. 'Running Experiments on Amazon Mechanical Turk.' *Judgment and Decision-Making* 5 (2010): 411-419.
- PENNEBAKER, J.W., and T.C. LAY. 'Language Use and Personality during Crisis: Analyses of Mayor Rudolph Giuliani's Press Conferences.' *Journal of Personality and Social Psychology* 77 (2002): 1296-1312.
- PENNEBAKER, J.W., M.R. MEHL, and K.G. NIEDERHOFFER. 'Psychological Aspects of Natural Language Use: Our Words, Our Selves.' *Annual Review of Psychology* 54 (2003): 547-577.
- RENNEKAMP, K. 'Processing Fluency and Investors' Reactions to Disclosure Readability.' *Forthcoming in the Journal of Accounting Research*, 2012.
- ROGERS, J., A. VAN BUSKIRK, and S. ZECHMAN. 'Disclosure Tone and Shareholder Litigation.' *The Accounting Review* 86 (2011): 2155-2183.
- RUPAR, K. 'Significance of Forecast Precision: The Importance of Ex Ante Expectations.' Unpublished paper, 2011. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1752217
- RUTHERFORD, B.A. 'Obfuscation, Textual Complexity, and the Role of Regulated Narrative Accounting Disclosure in Corporate Governance.' *Journal of Management and Governance* 7 (2003): 187-210.
- SCHEFFÉ, H. *The Analysis of Variance*. New York: John Wiley & Sons, Inc., 1959.
- SCHWARZ, N., H. BLESS, F. STRACK, G. KLUMPP, H. RITTENAUER-SCHATKA, and A. SIMONS. 'Ease of Retrieval as Information: Another Look at the Availability Heuristic.' *Journal of Personality and Social Psychology* 61 (1991): 195-202.
- SCHWARZ, N. 'Metacognitive Experiences in Consumer Judgment and Decision Making.' *Journal of Consumer Psychology* 14 (2004): 332-348.

- Securities and Exchange Commission. (1998). 'A Plain English Handbook: How to Create Clear SEC Disclosure.' SEC Office of Investor Education and Assistance. Available at: <http://www.sec.gov/pdf/handbook.pdf>.
- SEDOR, L. 'An Explanation for Unintentional Optimism in Analysts' Earnings Forecasts.' *The Accounting Review* 77 (2002): 731-753.
- SHAH, A., and D. OPPENHEIMER. 'Easy Does it: The Role of Fluency in Cue Weighting.' *Judgment and Decision Making* 2 (2007): 371-379.
- SHROUT, P., and N. BOLGER. 'Mediation in Experimental and Non-Experimental Studies: New Procedures and Recommendations.' *Psychological Methods* 7 (2002): 422-445.
- SUBRAMANIAN, R., R. INSLEY and R. BLACKWELL. 'Performance and Readability: A Comparison of Annual Reports of Profitable and Unprofitable Corporations.' *The Journal of Business Communication* 30 (1993): 49-61.
- TAN, H-T., R. LIBBY and J. HUNTON. 'Analysts' Reactions to Earnings Preannouncement Strategies.' *Journal of Accounting Research* 40 (2002): 223-246.
- TAN, H-T., E. WANG and B. ZHOU. 'How Does Readability Affect Financial Disclosure Credibility? Message Consistency Matters.' Unpublished paper, 2011. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1942879
- TAUSCZIK, Y. and J.W. PENNEBAKER. 'The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods.' *Journal of Language and Social Psychology* 29 (2010): 24-54.
- VENKATARAMAN, S. 'The Joint Impact of Commitment to Disclosure and Prior Forecast Accuracy on Managers' Reporting Credibility.' Unpublished paper, The Georgia Institute of Technology, 2008. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1267510
- YANG, H. 'Capital Market Consequences to Managers' Voluntary Disclosure Styles.' *Journal of Accounting and Economics* 53 (2012): 167-184.
- YOU, H., and X. ZHANG. 'Financial Reporting Complexity and Investor Underreaction to 10-K Information.' *Review of Accounting Studies* 14 (2009): 559-586.
- ZHAO, X., J.G. LYNCH JR., and Q. CHEN. 'Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis.' *Journal of Consumer Research*, 37(2010): 197-206.