The Pennsylvania State University

The Graduate School

The Mary Jean and Frank P. Smeal College of Business

DATA OVERLOAD AND INVESTOR TRADING

A Dissertation in

Business Administration

by

Brian P. Miller

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2008

The dissertation of Brian P. Miller was reviewed and approved* by the following:

Paul E. Fischer
Professor of Accounting
Department Head of Accounting
Dissertation Co-Adviser / Co-Chair of Committee

Karl Muller
Associate Professor of Accounting
Chair of Graduate Program in Accounting
Smeal Research Fellow in Accounting
Dissertation Co-Adviser / Co-Chair of Committee

Andrew Leone
Associate Professor of Accounting

Arun Upneja
Associate Professor of Hospitality Management and PIC International Programs

*Signatures are on file in the Graduate School

ABSTRACT:

This paper examines the effects of annual report length and readability on small and large investors' trading behavior. Motivated by critics concerns over the effects of longer and more opaque annual reports, I use computational linguistic software to measure the length (number of words) and readability (*Fog Index* and *Plain English Index*) of 10-K filings. After controlling for the information content of the report, I find evidence that small investors reduce their trading when filings increase in length. Additional analysis indicates that longer reports are also associated with reduced consensus and short-term trading profits for small investors. I fail to find similar evidence for large investors, which is consistent with their superior processing skills. Further, despite the SEC's efforts to encourage plain English disclosure, I find little support that readability has an effect on investors' trading behavior.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank the members of my dissertation committee for their guidance and support: Paul Fischer (co-chair), Karl Muller (co-chair), Andrew Leone, and Arun Upneja. Useful comments have also been provided by Walid Al-Issa, Orie Barron, Neil Bhattacharya, Christine Cheng, Dan Givoly, Karen Hennes, Steven Huddart, Ben Lansford, Feng Li, Jim Vincent, Hal White and seminar participants at Carnegie Mellon University, Duke University, Indiana University, Michigan State University, Ohio State University, Penn State University, Purdue University, Southern Methodist University, University of Michigan, University of North Carolina-Chapel Hill, University of Notre Dame, University of Rochester, University of Toronto, and Washington University – St. Louis. I also express gratitude to Justin Patterson and the StyleWriter software engineers for programming assistance. I acknowledge financial support from the Smeal Dissertation Research Award.

I also want to express my gratitude for the mentorship I received from the faculty at Penn State. Specifically, I would like to thank Charlie Smith for his consistent moral support and his longstanding support of the Ph.D. program. My committee members have been great role models. Paul Fischer has not only consistently challenged me to think deeply, but has always been willing to sacrifice his time to chat about my research and even join me on the occasional road trip. Karl Muller has been a great advocate and has helped me cut through numerous obstacles, so that I could focus on my research and teaching. He has consistently modeled every aspect of scholarship, for which I am truly grateful. I have also benefited greatly from Andy Leone, who has been a great

encouragement to me through his positive attitude, tenacity for research, and generosity with his time.

Further, I would like to thank the many Penn State Ph.D. students who have blazed the trail before me, as well as those that have gone through the program with me. I would specifically like to thank Alan Jagolinzer for his encouragement to come to Penn State and his advice throughout the process. I would also like to thank Ben Lansford for his friendship and lending a listening ear when things got difficult. I am especially grateful to Hal White who has been the great friend and confidant throughout the entire program. I thank Karen Hennes, Maya Thevenot, and Jeremy Morris for their friendship and keeping things fun during the countless late nights and weekends during coursework. Craig Crossland also deserves credit for keeping me sane through courses in philosophy and professional development. Finally, I thank Walid Al-Issa for his friendship and countless hours discussing the deeper things of life.

In conclusion, I would like to thank my entire family who has sacrificed immensely for me to complete the program. Specifically, I thank my parents for their generosity and constant moral support. To Caleb, Joshua, and Melea for sacrificing some of your time with dad so he could get work done. Finally, I would like to thank my wife and best friend, Renee, for the encouragement she has provided and standing with me through a lot more than she ever bargained for. I look forward to the next phase of our lives together.

1. Introduction and Motivation

The securities industry has historically been driven by the fundamental concept that "all investors, whether large institutions or private individuals, should have access to certain basic facts about an investment prior to buying it, and so long as they hold it" (SEC 2007). However, simply providing access to infinite amounts of data does not necessarily aid investors, unless those investors can efficiently extract useful information from that data (Bloomfield 2002). Despite efforts by regulators to encourage companies to make financial disclosures more understandable (Firtel 1999), annual reports have instead become longer and less readable (Li 2008).[1] Accordingly, both legal scholars and regulators continue to raise concerns about investors' ability to process these longer and more opaque filings (Paredes 2003).

Recent empirical research provides evidence that longer and less readable reports are associated with lower earnings (Li 2008) and that firms subject to greater shareholder litigation tend to use less boilerplate and more readable language in their disclosures (Nelson and Pritchard 2007). Both papers are motivated by the assumption that management believes longer and/or less readable reports affect investor behavior, but neither study documents that investor behavior is actually affected by these report attributes. This study contributes to the existing literature by examining whether annual report length and readability affect investor trading behavior (holding information content constant).

[1] The Securities and Exchange Commission (SEC) has made several attempts to encourage disclosure clarity. Most recently, they adopted plain English regulations that encourage firms to provide more readable disclosures (SEC 1998).

I use computational linguistics to measure the effects of length (number of words) and readability (Fog Index and Plain English Index) on various metrics of investor trading behavior.[2] Specifically, I provide evidence on the effects of these report attributes on 1) trading volume; 2) trading consensus; and 3) short-term trading profitability. I investigate the impact of these measures on overall trading volume for all investors, as well as the specific trading activity of large and small investor groups. However, because the effects of length and readability are likely to be most pronounced amongst smaller investors (due to their limitations in processing capabilities, expertise, and scale), the primary focus of this study is on small trade behavior.[3]

To investigate the effects of report length and readability on investors' ability to extract useful information, I examine abnormal trading activity (volume and transactions) around the filing. *A priori,* it is unclear whether the attributes of length and readability will lead to either increased or decreased trading around the 10-K filing. For instance, faced with longer and/or less readable reports, some investors may elect not to process the report because doing so is too costly. As a consequence, these investors will not initiate trades in response to the report (Grossman and Stiglitz 1980; Bloomfield 2002). Thus, after controlling for information content, longer and/or less readable disclosures may lead to lower trading volume. On the other hand, longer and/or less readable reports may increase disagreement among investors that do elect to process the report.

[2] Consistent with prior research (Li 2008), I measure readability using the Fog Index (syllables per word and words per sentence). In addition to this measure, I use proprietary software to create a multidimensional plain English readability index that is based on the writing factors outlined by the SEC plain English guidelines (SEC 1998).

[3] Following prior research (Cready 1988; Lee 1992; Lee and Radhakrishna 2000), I infer whether the investor is small or large from the size of the trade.

Accordingly, disagreement may increase with longer or less readable reports because these attributes are likely associated with dispersion in the inferences drawn by the investors. Such disagreement, in turn, would lead to higher trading volume (Karpoff 1986; Bamber, Barron, and Stober 1997; 1999).

In addition to trading activity, I also examine the effects of report length and readability on the trading consensus among small and large investor trade groups. Prior experimental evidence suggests that nonprofessional investors faced with increased disclosure generate (or infer) private information whereas sophisticated investors are largely unaffected by the increased disclosure (Barron, Byard, and Enis 2004). Based on this evidence, I hypothesize that longer and/or less readable reports will lead to lower consensus among small investors.

Finally, I examine the short-term trading profitability of both small and large investors around the filing. This examination is motivated by a concern that lengthy and unreadable reports may influence investors to make suboptimal investment decisions. If small investors are unable to completely process longer and/or less readable reports, they will be less likely to buy (sell) when the short-window return is positive (negative).[4] I predict that when faced with longer and/or less readable reports the processing constraints of small investors will lead them to make more suboptimal trades.

I examine approximately 13,000 10-K's filed between 1995 and 2006 and find evidence consistent with longer reports affecting small investor trade behavior around the

[4] Since actual trading profits of small and large investors are unobservable, I modify an approach developed by Asthana et al. (2004) to analyze whether the direction of each trade is more likely to be a buy (sell) when the return in the subsequent five-day window after the trade is positive (negative).

filing. Specifically, I find that longer reports are associated with lower levels of small investor trade activity. For perspective, an increase of 15,000 words (roughly equivalent to the 64% (i.e., 55 page) increase in report length from 1995 to 2006) results in a reduction in small abnormal volume around the filing of approximately 20%. Additional tests indicate that longer reports are associated with a slight decrease in consensus among small investors and provide some evidence that small investors are more likely to make suboptimal trades when trading around longer annual reports. I fail to document similar evidence for large traders, which is consistent with their superior processing skills.

Demonstrating the effects of lengthy disclosures on small investors provides evidence to regulators assessing disclosure regulation and considering whether to adopt "sunset" provisions for disclosure requirements.[5] Further, in contrast to the SEC's 1998 plain English initiative that emphasized the importance of clear writing, I find little significant evidence that report readability is associated with either small or large trading activity. However, the failure to document a significant relationship between readability and trading volume should be interpreted with caution, as the lack of association could be due to the inability of the measures used in this study to completely capture financial disclosure readability.[6]

This paper continues with Section II, which provides additional background and hypothesis development. Section III outlines the research design. Sample selection and

[5] Radin (2007) recommends a shift toward "sunset" provisions on future disclosure regulations, where the mandatory disclosure would cease to be required after a specific date unless regulators take action to extend it.

[6] For instance, the readability measures employed in this paper do not address the SEC's concern over financial statement formatting.

descriptive data are provided in Section IV. Empirical results are summarized in Section V, Section VI summarizes various robustness checks and provides discussion regarding certain caveats, and Section VII concludes.

2. Background and Hypothesis Development

Efforts by the SEC to make financial data more readable and understandable date back to the 1933 Securities Act (Firtel 1999).[7] For instance, in 1969 the SEC released the *Wheat Report* indicating that the length and complexity of prospectuses prevented the average investor from readily understanding them. The report went on to recommend against unnecessarily long, complex and/or verbose writing. These concerns over lengthy reports were reiterated by Arthur Levitt in his 1997 remarks to the Securities Regulation Institute when he asserted that "[i]n many cases, the problem is not a lack of information; quite the opposite. Too much information can be as much a problem as too little. More disclosure does not always mean better disclosure" (Levitt 1997).

Under Levitt's leadership, the SEC adopted the 1998 plain English regulation, which required issuers to use plain English principles in the organization, language, and design of prospectuses.[8] In addition to the specific rules regarding prospectuses, the SEC also released *A Plain English Handbook: How to Create Clear SEC Disclosure Documents* as a way to encourage the use of plain English attributes in all mandatory disclosures. Along with reiterating the pitfalls of excessively lengthy documents and paragraphs, the SEC discusses several common problems that they have encountered in public filings. These

[7] Firtel (1999) provides a thorough review of the SEC's regulatory progression regarding financial statement readability, the history of plain English in other settings, and the debate on the intended audience of disclosure.

[8] SEC Rule 421(d) became effective October 1, 1998.

concerns include the use of lengthy sentences, passive voice, superfluous words, legal and financial jargon, unnecessary details, and unreadable design and layout. Most recently, the SEC has focused its attention on the readability of mutual fund (Glassman 2005) and executive compensation disclosures (Cox 2007).

In addition to concerns over the readability and complexity of disclosures, some contend that financial reports have simply become too long for investors to effectively process (Paredes 2003). Consistent with these concerns, prior evidence documents a substantial increase in report length over the past decade (Li 2008).[9] However, whether the additional data is beneficial to investors is unclear. In fact, some accounting practitioners argue that the useful information disclosed in annual report filings is now hidden amongst a plethora of boilerplate, redundant, immaterial, or even irrelevant data making disclosures increasingly difficult to process (Radin 2007).

2.1. Literature

Despite the importance that practitioners and regulators have placed on disclosure clarity and length, there is little large scale empirical evidence on the effects of these attributes on financial statement accessibility. Most of the early work in this area investigates the readability of financial statements and footnotes in small sample sizes. Jones and Shoemaker (1994) summarize this literature and conclude that financial

[9] Radin (2007) points out that recent increases in report length can be traced to the amount of new disclosure devoted to risk, compensation, pension accounting, and stock options since the collapse of Enron. However, it is also plausible that firms have voluntarily increased disclosure over time. The focus of this paper is how length affects different investors' trade behavior (after controlling for information content) and not necessarily the causes of increased disclosure.

disclosures are very difficult to read and are "inaccessible" to a large proportion of unsophisticated investors.

As previously discussed, recent readability research focuses primarily on managerial incentives to disclose more or less readable reports under different circumstances. For instance, Li (2008) provides large sample evidence that financial statements are longer and harder to read when firms have lower earnings and suggests that managers may opportunistically alter the readability of annual reports to hide adverse information from investors. However, Li fails to find a significant correlation between report readability and future stock returns, suggesting that any implications of disclosure readability are impounded into stock prices. Nelson and Pritchard (2007) find that firms subject to greater shareholder litigation use more readable language in their disclosures and avoid boilerplate warnings that are unlikely to garner legal protection under the Private Securities Litigation Reform Act of 1995. In general, recent research on length and readability focuses on managerial incentives, but does not address whether these attributes affect investor trade behavior.[10]

This study also contributes to a broad literature on price and trading volume reactions to earnings releases (Beaver 1968; Morse 1981), 10-K filings and 10-Q filings (Cready and Mynatt 1991; Asthana and Balsam 2001; Griffin 2003; Li and Ramesh 2007).[11] In

[10] Recent linguistics research on financial disclosures also investigates the market response to tonality (i.e., optimistic and pessimistic tone) in the media (Tetlock 2007 and Tetlock, Saar-Tsechansky, and Macskassy 2007) and earnings press releases (Davis, Piger, and Sedor 2007). This study differs from this work in at least three ways: 1) the linguistic measure is *readability* (not *tone*), 2) the investor response measure is trading behavior by different investor types (not overall market reaction), and 3) the setting is 10-K filings (not the media or earnings reports).

[11] Li and Ramesh (2007) document that after controlling for the concurrent release of earnings information the significant stock market and volume reaction for quarterly and small business filings disappear. Thus, in

addition to overall price and volume reactions, several studies provide evidence of differential trading behavior of small and large investors to information events across several settings: earnings announcements (Bhattacharya 2001; Shanthikumar 2004); first time EDGAR filings (Asthana et al. 2004); pro-forma earnings releases (Allee et al. 2007; Bhattacharya et al. 2007); and security analyst stock recommendations (Mikhail et al. 2007). Most relevant to this study, Asthana et al. (2004) document that the SEC efforts to make 10-K filings available on EDGAR resulted in increased abnormal trading volume and short-term profits for small investors but had no significant impact on large investors. As such, I limit my sample to electronic filings available on EDGAR and document evidence that simply making data electronically available may not aid small investors unless they are capable of processing that data into meaningful information. Hence, I contribute to this literature by providing evidence on whether the way data is presented has any effect on investors' ability to extract useful information (holding electronic access constant).

2.2. Hypothesis Development

The primary objective of this paper is to examine whether longer and less readable reports affect trading investor behavior, but the impact of these disclosure attributes is not straightforward. First, given advances in technology, it is unclear that report attributes such as length or readability will have an affect on investors' trading behavior. Second, the existence of other information sources (e.g., media and analyst reports) may mitigate the effects of these attributes on investor trade behavior. In summary, although evidence

---

this study I exclude all small business filers, restrict my investigation to annual filings, and control for the number of days between the earnings announcement and the 10-K filing.

exists that annual reports have become longer and less readable (Li 2008), it is unclear whether these attributes will affect investor trading behavior.

2.2.1. Trading Volume

To assess the impact of report length and readability on investor trade behavior, I examine abnormal trading activity around the filing to investigate the effects of report length and readability on investor access to information.[12] The impact of longer and/or less readable reports on investor trade behavior around the filing is ambiguous. On the one hand, faced with longer and less readable reports, some investors could elect not to process the report because doing so is too costly. As a consequence, these investors will not initiate trades in response to the report (Grossman and Stiglitz 1980; Bloomfield 2002). Thus, after controlling for the information content in the report, longer or less readable disclosures may lead to lower trading volume.

On the other hand, to the extent that longer or less readable reports lead to more disagreement among those investors who do elect to process the report, these attributes may also lead to increased trading activity. Although most analytical models suggest that disagreement of some form drives trading around disclosure releases (Karpoff 1986; Kim and Verrecchia 1991, 1994, 1997; Kandel and Pearson 1995), Karpoff (1986) raises the possibility that this disagreement could stem from information interpreted differently by market agents. Subsequent empirical work by Barron (1995) and Bamber, Barron, and

[12] I focus on trading volume as opposed to the overall market reaction for several reasons: First, Cready and Hurtt (2002) provide evidence that trading volume provides the most visible indicator of investors' response to public disclosures. Second, trading volume allows me to differentiate between small and large investor trade, a characteristic which is unobservable in market reaction studies. Finally, from a practical standpoint, trading volume has the advantage of allowing me to use simultaneous market returns to control for information content of the filing.

Stober (1997; 1999) provides evidence that the trading around earnings announcements is at least partially attributable to newfound disagreement. As such, I contend that longer and less readable reports are harder to interpret and could lead investors to draw different inferences from the same report (Kandel and Pearson 1995). To the extent that length and readability lead to greater disagreement, I expect to find more abnormal trading.[13]

In summary, whether the attributes of length and readability lead to either increased or decreased trading around the 10-K filing is ambiguous. Formally, the first set of hypotheses (in alternative form) is as follows:

H1a: Abnormal investor trade increases or decreases when 10-K reports are longer.

H1b: Abnormal investor trade increases or decreases when 10-K reports are less readable.

As previously discussed, because smaller investors have limited processing capabilities, industry expertise, and scale, they are most likely to be affected by report length and readability. Hence, in addition to examining the effects of reporting attributes on total abnormal volume, I also investigate the effects on small and large investor trading volume. I predict that the effects of length and readability will be most pronounced in the small investor group.

[13] A finding of increased abnormal trade volume would also be consistent with recent experimental evidence. Smith (2005), for example, documents that increasing the quantity and consistency of information leads to greater confidence and trading aggressiveness.

2.2.2. Trading Consensus

In addition to trading volume, I also investigate whether longer and/or less readable reports lead to more dispersion in the inferences drawn by small and large investors that do elect to trade. Barron et al. (2004) provide experimental evidence that increases in disclosure lead nonprofessional investors to generate (or infer) private information while sophisticated professional investors are unaffected by the increased disclosure. This evidence is consistent with professional investors (smaller investors) being more homogeneous (diverse) with respect to their training and occupational selection and therefore more (less) likely to have common interpretations of specific disclosures.

As discussed in greater detail later, I measure small and large investors' trading consensus as the absolute value of daily net buys (buys minus sales) of shares traded deflated by total volume of shares traded (buys plus sales) on the same day. If length or lack of readability results in a high level of disagreement amongst an investor group, their buys and sells would off-set and trading consensus would approach zero. Alternatively, if investors agree about the information content of the report, they are more likely to trade in the same direction and trading consensus would approach one. Conditional on trade, I hypothesize that longer and less readable reports will lead to more disagreement within an investor class, and these effects will be greatest for the small investor class. More formally, my second set of hypotheses is as follows (stated in alternative form):

H2a: Trade consensus within a class of investors decreases when 10-K reports are longer.

H2b: Trade consensus within a class of investors decreases when 10-K reports are less readable.

2.2.3. Trading Profits

Regulators may ultimately be most interested in trading profits. Hence, I also investigate whether the direction of trades is associated with the information disclosed in the 10-K, as reflected in the share price change subsequent to trade. More specifically, I investigate whether investors are more likely to buy (sell) when the short-window return is positive (negative). If small investors are unable to process the longer and less readable reports as effectively as their larger more sophisticated counterparts, I expect to find small (large) investors to have less (more) profitable short-term trades.[14] Thus, my third set of hypotheses is as follows (stated in alternative form):

H3a: Small (Large) investor short-term profits decrease (increase) when 10-K reports are longer.

H3b: Small (Large) investor short-term profits decrease (increase) when 10-K reports are less readable.

3. Research Design

The research design consists of three stages. The first stage examines trading response to longer and less readable annual reports using abnormal volume and transactions. The second stage examines the effects of length and readability on small and large investor trading consensus. Whereas, the third stage examines the effects of these report attributes on small and large investor trading profitability.

[14] This prediction is consistent with prior experimental research which provides evidence that, at some point, individuals become overloaded with information and make worse decisions than if less information were available (e.g., Chewning and Harrell 1990, Stocks and Harrell 1995, Stocks and Tuttle 1998).

3.1. Abnormal Trading

To examine the effects of length and readability on the combined trading activity of all investor groups (i.e., small, medium, and large), I estimate the following regression model:

$$\text{ABN\_TRADE} = \sum_{i=1}^{n_i} \alpha_i d_i + \begin{bmatrix} \alpha_1 \text{LENGTH}_{it} + \alpha_2 \text{READ\_SCR}_{it} + \alpha_3 \text{AFTEAD}_{it} + \alpha_4 \text{AFTEXD}_{it} \\ + \alpha_5 \text{ABS\_RET}_{it} + \alpha_6 \text{MV}_{it} + \alpha_7 \text{MTB}_{it} + \alpha_8 \text{EARNINGS}_{it} \\ + \alpha_9 \text{NBSEG}_{it} + \alpha_{10} \text{NGSEG}_{it} + \alpha_{11} \text{NA\_FOLL}_{it} \end{bmatrix} + \varepsilon_{it} \quad (1a),$$

where the dependent variable, ABN_TRADE is the relevant measure of abnormal trading volume or transaction frequency.[15] I use two measures of abnormal trading volume (transaction frequency) to estimate the effects of length and readability on investors trading behavior.[16] First, I measure excess volume as in Barron et al. (2005), where EXVOL (EXNUM) is defined as the natural log of the cumulative trading volume (transactions) over the five-day event period (-1, 3) minus the natural log of the median volume (transactions) for contiguous five-day periods during the non-filing period (-49, -5).[17] Consistent with Asthana et al. (2004), I define abnormal trading volume AVOL (ANUM) as the mean daily trading volume (transactions) during the event period (-1, 3) minus the mean daily trading volume (transactions) during the non-filing period (-49, -5),

[15] As Asthana et al. (2004) point out, the number of individuals owning computers and gaining access to the internet has increased dramatically over the sample period, while the cost of on-line trading has gradually decreased. The use of abnormal volume (versus raw volume) should mitigate the effect of these market-wide economic and technological changes over the event period.

[16] Cready and Ramanan (1995) contend that the number of trades (or transactions) provides a useful metric in assessing trading around information events. Hence, I examine both volume and transaction frequency.

[17] I modify the measure to lengthen the window around the event period in order to provide ample time for investors to process the report. I also use a shortened non-filing period window to improve computational efficiency with TAQ by reducing the number of trading days analyzed and to match the period used in Asthana et al. (2004).

deflated by the standard deviation of daily trading volume (transactions) during the non-filing period (-49, -5).

LENGTH is the relevant metric of financial statement length, measured as either WORDS or TABLES+WORDS.[18] WORDS is equal to the logarithm of the number of words in the document. I follow Li (2008) and use the logarithm rather than the raw number of words because of the skewness in the number of words across firms and the few extreme values. Although this measure provides a useful metric of the amount of data in text format, it fails to incorporate the data included in tables. Further, it is unclear whether the quantity of data in tables acts as a substitute or complement to the data in the text. Hence, I create a variable TABLES+WORDS, defined as the natural log of the total words plus table cells contained in the document. However, as described in Appendix 1, since the ability to count the number of table cells requires firms to comply with a specific HTML standard, this measure is only available for a subset of firms after 1999.

READ_SCR is the relevant metric of readability, where readability is either measured as the Fog Index (READ_FOG) or the plain English Index (READ_PE). The Fog Index, originally developed by Gunning (1952), provides a simple, well known, and widely accepted formula for measuring readability. The measure captures text complexity as a function of syllables per word and words per sentence, and provides an approximation of the number of years of formal education a reader of average intelligence would need to be able to read the text once and understand that piece of writing. Generally, a Fog score

[18] Since useful information is likely contained in all parts of the 10-K filing, users must process all the data in the document to get the complete set of available information. Hence, I examine the entire 10-K document including MD&A, footnotes, contracts, and other exhibits.

greater than 18 means the text is unreadable. Consistent with prior literature, I measure READ_FOG as the ((*words per sentence + percent of complex words)*.4)*, where complex words are defined as words with three syllables or more. I follow Li (2008), and use the Lingua-EN-Fathom package of Perl to analyze the raw 10-K files and calculate both *READ_FOG* and *LENGTH*. A more detailed explanation of the process used to edit (compute) the document (scores) is provided in Appendix 1 (2).[19]

The Fog Index was not developed to specifically measure the attributes encouraged by the SEC's *Plain English Handbook*, and thus may measure the SEC's notion of financial statement readability with error. Accordingly, I use a proprietary computational software program, StyleWriter, to develop a unique multidimensional measure of financial readability based on the writing factors specifically outlined by the SEC plain English guidelines. As described in Appendix 2, I first identify the specific plain English problems identified by the SEC and match them to the specific variables identified by StyleWriter as shown in Figure 1.

FIGURE 1 – Plain English Writing Factors

SEC Plain English Problems	StyleWriter Plain English Measure
Passive voice	Passive Verbs
Weak / Hidden verbs	Hidden Verbs
Superfluous words	Overwriting (overuse of qualifying words)
Legal and financial jargon	Legal Words and Jargon /Abstract Words
Numerous defined terms	N/A
Abstract words	Jargon / Abstract Words
Unnecessary details	Tautologies / Overwriting
Long sentences	Number of Words / Avg. Sentence Length
Unreadable design and layout	N/A

[19] In addition to Li (2008), this Perl module has been used in various fields including information science and business communication. Examples include Collins-Thompson and Callan (2005) and Muresan et al. (2006).

Based on these guidelines identified by the SEC, I then create the plain English measure, *READ_PE,* defined as the number of errors identified by Stylewriter (i.e., Passive Verb + Hidden Verb + Overwriting + Legal Words and Jargon + Tautologies)*10 scaled by the approximate number of sentences in the document (number of words / average sentence length). The scalar enables me to parse out the effects of the document length, but perhaps more importantly allows me to address the SEC's concern about sentence length (longer sentences decrease the denominator and thus increase *READ_PE).*[20]

Consistent with the H1a and H1b, my expectation is that the coefficients on LENGTH (READ_SCR) will be negative if increases in report length (readability) result in increased processing costs leading to lower abnormal trade and positive if increases in report length (readability) lead investors to draw different conclusions about the information contained in the report and generate additional trade.

As I am interested in investigating investor's response to report attributes, it is important to control for other items that are likely to affect investor trading. For instance, I control for information content of the report by including the variable ABS_RET, which I define as the absolute value of the abnormal (market model) return on the firm's common stock, calculated over the 5-day event window (-1, 3).[21] My expectation is that more information content will lead to higher abnormal trading volume (Bamber and

[20] As discussed in detail in Appendix 2, I use *StyleWriter Plain English* software to obtain my plain English readability measure. It is important to note that although the *READ_PE* measure is directly based on specific components that the SEC claims are tied to poor financial writing, this specific measure has not been directly validated for financial statement readability.

[21] The market model abnormal return is estimated using CRSP-population data from a 255-day estimation period ending 46 days before the filing date.

Cheon 1995; Bamber et al. 1997). Additionally, Li (2008) documents that firms with lower earnings tend to have longer and less readable reports. To control for the possibility that LENGTH and READ_FOG are merely proxying for earnings information, I follow Li (2008) and include EARNINGS defined as operating earnings scaled by total assets (Compustat data178 / data6).

Additionally, Asthana et al. (2004) point out the importance of controlling for the effect of the timing of the 10-K filing on trading activity. Accordingly, to control for the effects of earnings information disclosed prior to the 10-K release, I include the variable AFTEAD, which is defined as the number of days after the preliminary earnings announcement date that the 10-K is filed. I expect that the longer after the earnings announcement the 10-K is filed, the lower the abnormal trading around the filing.[22] Prior research also finds evidence that the timing of information disclosure is related to the quality of that information. For example, firms delay releasing bad news (Kross and Schroeder 1984) and, consequently, the timeliness of that release affects the market reaction (Chambers and Penman 1984). Based on these findings, I also include a variable AFTEXD, which is defined as the number of days the current 10-K is filed after the date of the previous year's 10-K filing. I expect the longer the report is delayed after the expected filing date (i.e., previous year's filing date) the more abnormal volume.

[22] As with any event study, it is always possible that other information released during the event window could affect the results. For example, some analyst may process the information in the filings and make this data available to both small and large investors in the 3 day window after the filing. However, related information releases are likely to act as a filtering mechanism and would therefore bias against finding an effect on information processing. Other significant information unrelated to the 10-K filing is released during the window would likely only add noise.

$\sum_{i=1}^{n_i} \alpha_i d_i$ represents firm fixed effects, which are included in all regressions to mitigate firm-specific attributes influencing the results. As the firms may change over time, I also add controls for specific firm characteristics. To control for potential deviations in firm size, I include the independent variable MV, which I measure as the logarithm of the market value of equity at the end of the fiscal period. Since larger firms are likely to have better information environments (Atiase 1985; Bamber 1987), I expect that increases in firm size will be associated with decreases in abnormal trading around the filing.

I also expect that there will be less new information in the 10-K filings when the firm has greater analyst coverage. As such, I include NA_FOLL, which I define as the log of one plus the number of analysts as reported on I/B/E/S detail and excluded files. My expectation is that, all else equal, greater analyst following is associated with lower abnormal trade volume. Further, to control for changes in the effect of the firm's future growth opportunities on volume, I include the variable MTB, which is defined as the market value to book value of equity at the close of the current fiscal year. Finally, to control for deviations in firm complexity, I include the logarithm of one plus the number of business segments (NBSEG) and geographic segments (NGSEG).[23]

To minimize the effects of outliers, all variables are winsorized by year at the top and bottom 1%. Table 1 summarizes all variable definitions. Finally, all regressions are performed with clustered robust standard errors (Huber 1967; White 1980) to control for within year correlation.

[23] In addition to the controls reported in this analysis, I also add a control for whether the firm had a merger during the year as this would also increase firm complexity. I find (untabulated) that adding the control does not affect the significance of the results reported in this study.

3.1.1. Small and Large Investors Abnormal Trading

In addition to examining the total trade activity of all investors, I also examine the specific trade activity of small and large investors estimating the following equation:

$$\begin{aligned} ABN\_TRADE = & \sum_{i=1}^{n_i} \alpha_i d_{SML,i} + SML^* \begin{bmatrix} \alpha_1 LENGTH_{it} + \alpha_2 READ\_SCR_{it} + \alpha_3 AFTEAD_{it} + \alpha_4 AFTEXD_{it} \\ + \alpha_5 ABS\_RET_{it} + \alpha_6 MV_{it} + \alpha_7 MTB_{it} + \alpha_8 EARNINGS_{it} \\ + \alpha_9 NBSEG_{it} + \alpha_{10} NGSEG_{it} + \alpha_{11} NA\_FOLL_{it} \end{bmatrix} \\ & \sum_{i=1}^{n_i} \beta_i d_{LRG,i} + LRG^* \begin{bmatrix} \beta_1 LENGTH_{it} + \beta_2 READ\_SCR_{it} + \beta_3 AFTEAD_{it} + \beta_4 AFTEXD_{it} \\ + \beta_5 ABS\_RET_{it} + \beta_6 MV_{it} + \beta_7 MTB_{it} + \beta_8 EARNINGS_{it} \\ + \beta_9 NBSEG_{it} + \beta_{10} NGSEG_{it} + \beta_{11} NA\_FOLL_{it} \end{bmatrix} + \varepsilon_{it} \end{aligned} \quad (1b),$$

I define small and large traders as in Bhattacharya (2001), where trades are assumed to be made by small investors if the dollar amount of the trade is less than or equal to $5,000; when the dollar amount of the trade is greater than or equal to $50,000, the trade is assumed to be made by a large investor.[24] Consistent with prior literature, I ignore medium-sized trades to create a 'buffer-zone' between the trading activities of small and large investors.[25] I follow Asthana et al. (2004), I further modify this formulation so that for firms with a share price greater than $50, I define small trades as those less than or equal to 100 times the share price.[26] However, in order to keep a sufficient buffer

[24] Lee and Radhakrishna (2000) find that these trade size bins perform well in identifying trades initiated by individual investors and institutions. Recent literature (Barber et al. 2006) argues that the widespread introduction of decimalization in 2001 combined with computerized trading algorithms to break up institutional trades is likely to add noise to these measures. However, there is no reason to believe that this measurement error biases the results.

[25] Eliminating medium-sized trades increases the power of the test, since large investors may try to break up their trades to disguise their identity (Kyle 1985, Meulbroek 1992, Barclay and Warner 1993) but, for a variety of reasons are unlikely to make small trades (Bhattacharya et al. 2007).

[26] Consistent with prior literature, I also exclude the opening trade because it is often the sum of multiple orders and including it could add noise to the measures (Lee and Ready 1991; Lee 1992; Bhattacharya et al. 2007). Further, I only include trades with a 'regular sales' condition code. Bhattacharya et al. (2007 p. 587) points out that these "trades result from continuous two-sided auctions involving market orders, limit orders, and buys and sells against the specialists' inventories. This is not the case when the condition code indicates something other than a 'regular sale' (e.g., large block trades or stopped orders)."

between small and large trades, I exclude firms with a share price greater than $100 (i.e., maximum small trade is $10,000).

Equation (1b) represents the stacking of two regressions that allows statistical tests of differences in coefficient estimates across the small and large investor groups.[27] Hence, SML (LRG) is an indicator variable equal to 1 when the dependent variable represents the abnormal trading behavior for small (large) investors and 0 otherwise. Accordingly, the α (β) coefficients measure associations between small (large) investor trade behavior and the dependent variables (e.g., length and readability).

$\sum_{i=1}^{n_i} \alpha_i d_{SML,i}$ and $\sum_{i=1}^{n_i} \alpha_i d_{LRG,i}$ represent the fixed effects for the small and large investor trade observations, respectively. All other variables have been previously defined. My expectation for both small and large trade groups is that the coefficients on LENGTH (READ_SCR) will be negative if increases in report length or readability result in increased processing costs leading to lower abnormal trade and will be positive if increases in report length or readability lead investors to draw different conclusions about the information in the report and generate additional trade. I also examine whether small investors are more likely to be affected by longer reports than larger investors.

[27] See Maddala (2001) for discussion of stacked regressions. Since the technique captures any (potential) correlations across the error terms, stacking allows statistical tests of coefficients across equations. The main assumption underlying this technique is that the error terms from each regression have the same distribution. Under this assumption, stacking will result in consistent coefficient estimates. The current design was selected because it allows me to present coefficient estimates for both large and small investors. However, I could also have grouped the small and large investor observations and included an indicator variable for large trade observations, which when interacted with all the explanatory variables would capture the incremental effects of large traders over small traders.

Accordingly, I provide a test of coefficients across SML*LENGTH and LRG*LENGTH and expect the effects of length to be greater for small investors.

3.2. Trading Consensus

I also examine whether increases in report length and/or readability lead investors within the same class (e.g., small investors) to draw different inferences from the same reported information. Specifically, I use the following regression to examine whether length and readability decreases trading consensus around the filing:

$$\begin{aligned}\text{TRADE\_CONS} = &\sum_{i=1}^{n_i} \alpha_i d_{SML,\,i} + SML* \begin{bmatrix} \alpha_1 LENGTH_{it} + \alpha_2 READ\_SCR_{it} + \alpha_3 AFTEAD_{it} + \alpha_4 AFTEXD_{it} \\ + \alpha_5 ABS\_RET_{it} + \alpha_6 MV_{it} + \alpha_7 MTB_{it} + \alpha_8 EARNINGS_{it} \\ + \alpha_9 NBSEG_{it} + \alpha_{10} NGSEG_{it} + \alpha_{11} NA\_FOLL_{it} \end{bmatrix} \\ &\sum_{i=1}^{n_i} \beta_i d_{LRG,\,i} + LRG* \begin{bmatrix} \beta_1 LENGTH_{it} + \beta_2 READ\_SCR_{it} + \beta_3 AFTEAD_{it} + \beta_4 AFTEXD_{it} \\ + \beta_5 ABS\_RET_{it} + \beta_6 MV_{it} + \beta_7 MTB_{it} + \beta_8 EARNINGS_{it} \\ + \beta_9 NBSEG_{it} + \beta_{10} NGSEG_{it} + \beta_{11} NA\_FOLL_{it} \end{bmatrix} + \varepsilon_{it}\end{aligned} \quad (2)$$

To gauge the amount of trade consensus by investor groups around an annual report, I first classify all buys and sells using the standard Lee and Ready (1991) algorithm.[28] If the trade is above (below) the bid/ask midpoint, then the trade is deemed a buy (sell). Trades occurring at the midpoint are classified using the tick test.[29] Next, I create a variable TRADE_CONS, which is defined as the absolute value of daily net buys of shares traded (buys minus sales) deflated by total volume of shares traded (absolute value

[28] This algorithm was later tested by Lee and Radhakrishna (2000) using TORQ data, and it appears to correctly classify the vast majority of trades.

[29] The tick test compares the trade price to adjacent trades. Generally, a trade is classified as buy (sell) when the price is higher (lower) than the price of the previous trade. In cases when the price is the same as the previous trade (a zero tick), the classification of buy (sell) defaults to the last trade where there was a price change.

of buys plus sales) on the same day for each investor group (e.g., small). All other variables have been previously defined.

If small or large investors process the 10-K and agree (disagree) on the information's effect on the firm's terminal value, then they are likely to trade in a similar (different) direction to other investors in that same group leading to higher (lower) TRADE_CONS. Consistent with H2, I expect longer and less readable reports to increase disagreement for both small and large investor groups. Hence, I predict negative coefficients on both LENGTH and READ_SCR. Further, I also examine whether longer reports affect small investor disagreement more than larger investors. Accordingly, I provide a test of coefficients across SML*LENGTH and LRG*LENGTH and expect that longer reports will have a greater affect on the consensus of small investors.

3.3. Trading Profitability

I use the following regression to examine the trading profitability of small and large investors:

$$\begin{aligned}\text{TRADE\_RET} = &\sum_{i=1}^{n_i} \alpha_i d_{SML,\,i} + SML^* \begin{bmatrix} \alpha_1 LENGTH_{it} + \alpha_2 READ\_SCR_{it} + \alpha_3 AFTEAD_{it} + \alpha_4 AFTEXD_{it} \\ + \alpha_5 ABS\_RET_{it} + \alpha_6 MV_{it} + \alpha_7 MTB_{it} + \alpha_8 EARNINGS_{it} \\ + \alpha_9 NBSEG_{it} + \alpha_{10} NGSEG_{it} + \alpha_{11} NA\_FOLL_{it} \end{bmatrix} \\ &+ \sum_{i=1}^{n_i} \beta_i d_{LRG,\,i} + LRG^* \begin{bmatrix} \beta_1 LENGTH_{it} + \beta_2 READ\_SCR_{it} + \beta_3 AFTEAD_{it} + \beta_4 AFTEXD_{it} \\ + \beta_5 ABS\_RET_{it} + \beta_6 MV_{it} + \beta_7 MTB_{it} + \beta_8 EARNINGS_{it} \\ + \beta_9 NBSEG_{it} + \beta_{10} NGSEG_{it} + \beta_{11} NA\_FOLL_{it} \end{bmatrix} + \varepsilon_{it}\end{aligned} \quad (3)$$

I measure trading profitability using a modified version of a metric initially developed by Asthana et al. (2004). The general idea behind this measure is to examine whether the direction of small or large trades is associated with the information disclosed

in the 10-K, as reflected in the share price change subsequent to trade. If investors are properly processing the information in the reports, then they should trade in the correct short-term direction. For instance, one would expect to find investors buying (selling) when subsequent returns are positive (negative). [30]

I modify the approach in Asthana et al. (2004) and calculate the approximate profitability of each individual trade by comparing the trade's execution price to the closing price of the stock five days later.[31] More specifically, I first classify all buys and sells using the standard Lee and Ready (1991) algorithm described above. I then calculate an absolute return measure for each individual trade by subtracting the price at which the trade executed from the subsequent 5 day closing price, scaled by the price at which the trade executed. I then multiply the return from the previous step by 1 (-1) if the transaction was a buy (sale) to create the return for each trade. Finally, I create a variable TRADE_RET, which is defined as the group's (e.g., small) average signed return for all trades that occurred during the five-day window surrounding the 10-K filing (-1, 3).[32]

I expect to find that both the trading profitability of small investors will be negatively affected by longer and less readable reports, while large investors will be able to use their

[30] As Asthana et al. (2004) point out, without actual investor data it is impossible to observe the actual gains realized. Hence, the metric estimates the gains that could have been achieved if the investors closed out the position after 5 days.

[31] Asthana et al. (2004) measure of trading profitability calculated as the daily net buys (buys minus sales) multiplied by change in price for the subsequent five day period and deflated by total shares traded (buys plus sales) on the same day. By definition the measure assumes that all trades during the day took place at the closing price for that day, which is likely to add noise to the measure. I attempt to provide a more precise measure of trading profits.

[32] I follow Asthana et al. (2004) and calculate the trading profits in a short window (i.e., five days) as opposed to a longer window based on the assumption that the information should be impounded within five days and a longer window would in most cases contain additional information events which would likely add noise to the measure.

superior processing skills to generate additional trading gains when reports are longer and less readable. Accordingly, I predict negative coefficients for SML*LENGTH and LRG*READ_SCR and positive coefficients for LRG*LENGTH and LRG*READ_SCR. Finally, to examine whether longer reports lead to lower trading profits for small investors than large investors, I provide a test of coefficients across SML*LENGTH and LRG*LENGTH, and expect that longer reports will reduce the trading profits of small investors more than large investors.

4. Sample and Descriptive Data

4.1. Sample Collection

I gather my sample by downloading all 10-K (and 10-K405) filings on EDGAR from 1994 to 2006.[33] I then require a firm-year match on PERMNO (from CRSP), GVKEY (from COMPUSTAT), CIK (from EDGAR), and TAQ data for small and large volume around the 10-K filing.[34] Consistent with prior research, I eliminate observations where the ticker identifier on TAQ changed during the year, the stock price was less than one dollar, or the firm split their stock or issued a stock dividend during the year. I further eliminate all observations where there was insufficient text remaining after I eliminated the header, table, and other information as described in Appendix 1. In order to calculate the AFTEXD variable, I require the prior year filing date to be available on EDGAR, which eliminates all 1994 observations and a significant number of other observations.

[33]Forms 10-K and 10-K405 are identical in substance, except that form 10-K405 indicates that an officer or director of a public company failed to file a Form 4 (or related Form 3 or Form 5) within the required time period. I eliminate all other forms of 10-K filings (e.g., 10-KSB, 10-K/A, etc.) because they are likely to contain characteristics that may affect trading behavior.

[34] For comparison purposes between large and small investors, I require non-zero mean trade for each investor type during the control and the 5-day window surrounding the 10-K filing.

Finally, I eliminate observations where there is missing Compustat data (e.g., MV, ASSETS, NBSEG, NGSEG, etc.) or where analyst following is unavailable on I/B/E/S. The final sample consists of 3,949 unique firms with 13,252 firm year observations. This number reduces to 4,687 observations with TABLES+WORDS available. Table 2 provides a detailed description of the sample selection procedure.

4.2. Sample Characteristics and Correlations

Table 3 provides sample characteristics. Panel A provides means, standard deviations, and medians for several variables of interest. The mean (median) EXVOL_TOT is .13 (.05), and the means (medians) for EXVOL_SML and EXVOL_LRG are .12 (.06) and .59 (.083), respectively. The evidence of abnormal trading around the filing is consistent with survey evidence that small investors utilize the information in the annual report as a primary source for corporate financial information (Mezick 2001).[35] The mean WORDS is 10.39, which translates into an average of 32,533 words per 10-K. The mean Fog Index is 19.95, which classifies the vast majority of financial statements as unreadable according to most interpretations of the index.[36] The mean ABS_RET is 4.8%, indicating that there is an abnormal market reaction during the 10-K filing period. The remaining statistics in Panel A reveal that the firms in this sample

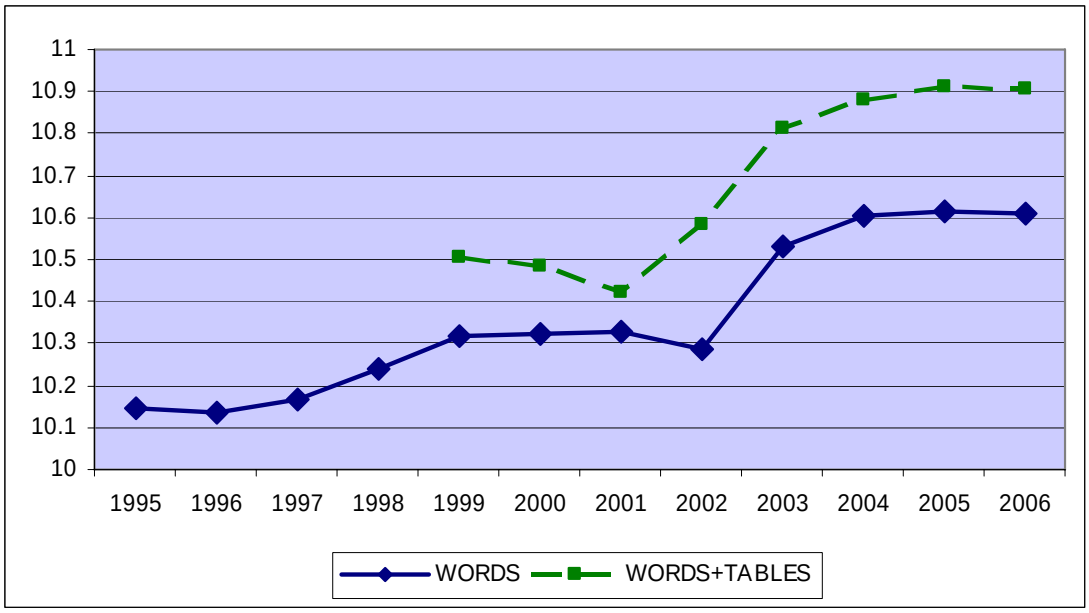
[35] Mezick (2001) surveys individual investors and finds that 59% of the investors in her survey use the annual report as their primary source for corporate financial information. The survey also finds that nearly one-third of investors use EDGAR to access financial data.

[36] The length and Fog scores are consistent with Li (2008) who reports mean (median) length of 10.08 (10.05) and mean (median) Fog Index of 19.4 and 19.2 for his sample of more than 50,000 firm years from 1994-2004. The slightly higher results observed in this study are at least partially attributable to two factors. First, I include 10-K's filed in 2005 and 2006, which have more words and higher readability scores (Table 3 - Panels B and C). Second, because of more restrictive data requirements (e.g., TAQ), the firms in this study are larger than Li (2008).

are relatively large and profitable. Hence, one caveat is that the inferences drawn from this study may not relate to smaller or less profitable firms.

Panel B (C) of Table 3 provides information on length (readability) across time. Consistent with Li (2008), both WORDS and TABLES_WORDS appear to be increasing over time. Figure 2 provides a chart of reporting the increase in document length as measured in WORDS and WORDS+TABLES during the sample period.[37] The pattern in both measures appears consistent with an increase in the amount of data reported in text and tables in 10-K filings over the past decade.

FIGURE 2 – Length of 10-K Document over Time



Panel C of Table 3 also provides evidence that READ_FOG appears to show an increasing trend over the sample period. Although READ_PE appears to follow a similar pattern to READ_FOG across time, there is no apparent increase in READ_PE during the

[37]As previously noted the calculation of table cells requires firms to comply with a specific HTML standard, hence, this measure is only available for a subset of firms after 1999.

sample period. Figure 3 reports the trend in readability during the sample period for both readability indices.

FIGURE 3 – Readability of 10-K Documents over Time

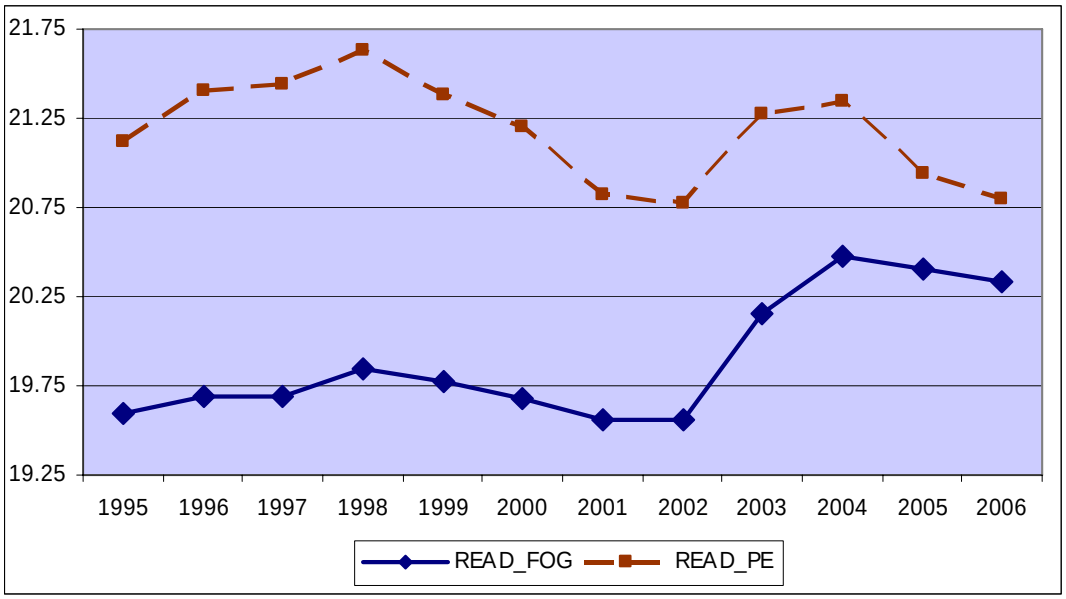


Table 4 provides Pearson and Spearman correlations of variables for both small (Panel A) and large (Panel B) investor trade groups, respectively. Consistent with Li (2008), I find significant (<1% two-tailed) negative correlations between EARNINGS and both measures of length (WORDS and WORDS+TABLES) and readability (READ_FOG and READ_PE). Further, I find that the measures of readability (READ_FOG and READ_PE) are significantly positively associated with each other (<1% two-tailed). However, the rather modest correlations (62% Pearson and 57% Spearman) suggest that the plain English measure is picking up somewhat different attributes of readability from the Fog index. Finally, I find a high correlation between WORDS and TABLES+WORDS (85% Pearson and 96% Spearman), which suggests that the quantity of data in tables is proportional to the length of text. Hence, measuring

word count should be a reasonable proxy for the data contained in both text and tabular formats.

5. Sample and Descriptive Data

5.1. Trading Activity

Table 5 reports the multivariate regression results for the effects of length and readability on excess (EXVOL) and abnormal (AVOL) trading volume, respectively. Overall, the results are consistent with small (but not large) investors trading less when reports are longer. Specifically, in column [4] of Panel A, the coefficient on SML*WORDS is -.024 (t-statistic -3.15), the coefficient on SML*WORDS+TABLES in column [5] is -.056 (t-statistic -2.86), and the coefficient on SML*LENGTH in column [6] is -.0217 (t-statistic -2.60).

I find only limited evidence of an effect of report length on total investor trade behavior. More specifically, although the coefficient on WORDS in column [1] of Panel A is significant (t-statistic -2.86), the coefficients on length in columns [2] and [3] are insignificant. Further, there is no significant evidence that large investors are affected by longer reports. However, when I compare the coefficients across SML*LENGTH and LRG*LENGTH in columns [4-6] of Panel A, I find only limited evidence that small investors are more affected by longer reports than large investors. More specifically, I find that the coefficient on SML*LENGTH is significantly more negative than the coefficient on LRG*LENGTH only in column [6] (F-statistics of 2.49).

Across all columns in Panel A, there is no significant evidence that less readable reports have an effect on excess volume, with the exception of a decrease in abnormal

trade among large investors in column [6] where the coefficient on READ_PE is -.012 (t-statistic -2.61). This evidence is consistent with an increase in large investor processing cost when reports are less readable. Finally, the majority of control variables are significant in the expected direction.

Panel B of Table 5 reports the results for abnormal volume (AVOL). The primary results are consistent with the significant evidence reported in Panel A. For instance, in column [4] of Panel B, the coefficient on SML*WORDS is -.039 (t-statistic -2.13), while in column [5] the coefficient on SML*WORDS+TABLES is -.145 (t-statistic -3.00). The coefficient on SML*LENGTH is negative but insignificant in column [6], while the coefficient on LRG*LENGTH becomes significant (t-statistic 2.32). Columns [4], [5], and [6] of Panel B reveal that the coefficients on SML*LENGTH are significantly more negative than the coefficients on LRG*LENGTH (F-statistics of 3.10, 2.19, 5.71 respectively). In summary, this evidence is consistent with small investors' abnormal trade being lower than large investors when reports are longer.

Most other findings in Table 5 Panel B are consistent with Panel A, with a few notable exceptions. First, the effect of length on total trade volume in column [1] remains negative, but is insignificant in Panel B. Second, in the sub-sample of firms reported in column [5], I find some significant evidence (t-statistic 2.00) that less readable reports (i.e., higher READ_FOG) appear to be associated with higher levels of small investor trading. This finding is consistent with small investors having increased disagreement when reports are less readable, but the evidence is limited to the sub-sample of firms where TABLES+WORDS is available.

Panels A and B of Table 6 build on the evidence reported in Table 5 by reporting results for excess transactions (EXNUM) and abnormal transactions (ANUM), respectively. Overall, the results are consistent with the previous table and provide additional support for the association between longer reports and lower amounts of small investor trade. For instance, the coefficient on SML*WORDS is -.026 (t-statistic -3.41) in column [4] of Panel A, while in column [5] the coefficient on SML*WORDS+TABLES is -.062 (t-statistic -3.57), and in column [6] the coefficient on SML*WORDS is -.0228 (t-statistic -3.06). Further, in Panel B the coefficients on SML*WORDS and SML*WORDS_TABLES remain significant (t-statistics -2.43 and -4.14, respectively) in columns [4] and [5]. I continue to find some evidence that longer reports are more likely to lead to less small investor abnormal trade versus the abnormal trade of large investors. Specifically, I find significant evidence that the coefficients on SML*LENGTH are less than the coefficients on LRG*LENGTH in columns [5] and [6] of Panel A (F-statistics 2.85 and 4.83, respectively) and columns [4], [5], and [6] of Panel B (F-statistics 3.89, 4.34, and 4.06, respectively). Finally, similar to Table 5, I find only minimal significant evidence of an effect of readability on small investor trade behavior.

Overall, the results in Tables 5 and 6 provide evidence consistent with small investors trading less when reports are longer, but little evidence that length has much of an effect on large investors. To provide perspective, an increase of 15,000 words (roughly equivalent to the 64% (55 page) increase in report length from 1995 to 2006) results in

approximately a 20% reduction in small abnormal volume around the filing.[38] Specifically, an increase of this magnitude reduces excess trade around the filing from the overall sample median of 6.0% to 4.8%.[39] In summary, after controlling for information content and other firm factors, longer filings are associated with less abnormal trade for small but not for large investors.

5.2. Trading Consensus

The negative association between report length and small investor trade volume in the previous section is consistent with longer reports increasing processing costs. However, the findings could also be consistent with longer reports providing more clarity and therefore less disagreement amongst small investors. In this section, I examine the effects of length on small investor trade consensus to provide evidence on whether the decrease volume is due to increased processing costs or increased consensus.

Table 7 provides regression results for small and large investors' consensus trade behavior (TRADE_CONS). Overall, the results are consistent with longer reports leading to more disagreement amongst small investors that elect to trade. For example, the coefficient on SML*WORDS in column [1] is significant -.0071 (t-statistic -1.49). Although the evidence in column [1] is only weakly significant, the coefficients on SML*TABLES+WORDS and SML*WORDS in columns [2] and [3] provide stronger evidence of an effect (t-statistics -3.68 and -3.99, respectively). Further, when I test

[38] The number of words increased from a median of approximately 24,000 words in 1995 to a median of approximately 39,000 words in 2006. The page estimate provided assumes 270 words per page, which is the approximate number of words contained on a standard 8½ by 11 page that is double-spaced with 12 point font.

[39] This analysis is based on the coefficients as reported in column [4], Panel A of Table 5. Similar results hold for Panel A of Table 6.

across coefficients in columns [2] and [3], I find significant evidence that longer reports result in greater disagreement among small investors than among large investors (F-statistics 6.26 and 3.99, respectively). Although the association between length and trade consensus is statistically significant, the economic effects appear to be minimal. For instance, an increase of 15,000 words results in a decrease in consensus from a median value of 19.9% down to 19.5%. Finally, I fail to find evidence that READ_FOG has any significant effect on trade consensus. I do find some evidence in column [3] that READ_PE is associated with lower levels of large investor trade consensus.[40] In summary, longer reports appear to be associated with a slight decrease in consensus amongst small investors. These findings provide support that the evidence in the previous section is likely due to increased processing costs and not increased consensus.

5.3. Trading Direction

Table 8 provides regression results for the analysis of small and large investors' short-term trading profitability (5-day window). The evidence is consistent with small investors making more sub-optimal trades when reports are longer. For instance, the coefficient on SML*WORDS in columns [1] and [3] are -.0003 and -.0005 (t-statistic -1.80 and 3.08, respectively). The coefficients on SML*TABLES+WORDS column [2] is negative but insignificant. I fail to find significant evidence that large investors are profiting when reports are longer. However, columns [1] and [3] provide evidence that small investors are more likely than large investors to make sub-optimal trades when

[40] The positive association between READ_PE and TRADE_CONSENSUS is inconsistent with the prediction and calls for further investigation. This unexpected association may be attributable to the difficulty in measuring financial statement readability, which calls for further validation of the readability measures among financial statement users.

reports are longer (F-statistic 3.19, 18.71). For perspective, the effect of an increase of 15,000 words results in an approximately 67 basis point annualized reduction in trading profits (before transaction costs).[41] In summary, Table 8 provides some evidence consistent with longer reports leading to less profitable trades among small investors.[42]

6. Robustness Checks and Caveats

6.1. Robustness Checks

In addition to the primary tests reported in this paper, I run certain robustness tests to examine the sensitivity of the results found in this paper. As previously discussed, I include a control variable in the multivariate regressions, AFTEAD, to control for the effects of earnings information disclosed prior to the 10-K release. However, there is some concern that the trading around 10-K filings can be affected by earnings announcements that are released in within a short window prior to the 10-K filing. To ensure that these potentially confounding events do not effect the results, I eliminate all observations where AFTEAD is less than or equal to 5 days and find (untabulated) that the results of this paper are unaffected by eliminating these observations.

There is also some concern that some of the 10-K filings may be shorter in length because they incorporate financial statement information by reference to other filings/documents. In order to address this potential concern, I use a PERL script to

[41] In addition to the equally weighted returns reported in Table 8, I also perform the analysis after value weighting the returns by trade size. I find (untabulated) that the significant differences between the trading returns of small and large investors are unaffected by this alternative measure. However, the coefficient on WORDS (WORDS+TABLES) in columns 1 (2) becomes insignificant.

[42] As in Table 7, the strong positive association between READ_PE and TRADE_RET is inconsistent with the prediction and calls for further investigation.

identify these firms that include their financial statements by reference, and find that the results are unaffected by eliminating these observations.[43]

6.2. Caveats

Certain caveats are in order. First, to the extent that firm specific risk or business complexity are correlated with length and that small and large investors react differently to these attributes, it is possible that these underlying factors contribute to the results. In an attempt to rule out these alternative explanations, I include firm fixed effects and variables to control for both variations in business complexity (e.g., firm size, number of business and geographic segments) and material events that are potentially correlated with changes in firm risk (e.g., mergers). However, to the extent that the proxies employed for risk and business complexity do not adequately control for these factors, it is possible that the findings could be attributable to these alternative explanations.

Further, it is difficult to completely control for the information content of the report. Prior literature has shown that longer disclosures have been associated with more information content in certain settings. For example, Francis et al. (2002) document that longer earnings announcements are positively associated to the absolute market reactions of the press releases. In contrast to Francis et al. (2002), the focus of this study is on the effects of 10-K filing length and readability on investor trading behavior after holding information content constant. A potential concern in this study is that the absolute value of the market reaction may be a noisy proxy for the information content of the report. To the extent that longer reports are associated with more information and the market

[43] Specific regular expressions are available upon request from the author.

response fails to control for the report content, the expectation would be that longer reports would be associated with increased trading for both small and large investors. However, the evidence in this study is consistent with longer reports being associated with decreased small investor trade volume. In summary, although the control for information content may contain measurement error, it is unclear that this error would bias toward the findings in this study.

7. Conclusions

There is much discussion among accounting practitioners and legal scholars regarding the effects of longer and more opaque annual reports on the average investor (Paredes 2003, Radin 2007). Despite these concerns and regulators efforts to encourage clearer financial disclosures, minimal large sample empirical evidence exists on whether these attributes have an impact on investors' trading behavior. I use computational linguistic software to measure the length (number of words) and readability (Fog Index and Plain English Index) of 10-K filings. After controlling for both information content and firm effects, I find evidence consistent with longer reports affecting small investor trading activity around the filing date but not for large investors.

Specifically, I find that longer reports are associated with lower levels of small investor abnormal trade volume and transaction frequency. However, I find minimal evidence of an effect on large investor trade activity. I also find evidence that longer reports are associated with a slight decrease in consensus amongst small investors' trade behavior but again find no effect on large investor consensus. Further, I provide evidence consistent with small investors being more likely to make suboptimal trades when trading

around longer annual reports. Finally, I find little evidence of readability affecting the metrics of investor trading behavior investigated in this paper.

The evidence provided in this study should prove useful to academics and regulators. For instance, prior research provides evidence that making 10-K filings available on EDGAR had an effect on the abnormal trading volume and pattern of small investors but not large investors (Asthana et al. 2004). I investigate firms whose reports were filed on EDGAR and document evidence that despite having electronic access to data, small investors are affected by longer reports. Hence, simply making data publicly available may not aid small investors, unless they can efficiently process that data into meaningful information.

REFERENCES

Allee, K., N. Bhattacharya, E. Black, and T. Christensen. 2007. Pro forma disclosure and investor sophistication: external validation of experimental evidence using archival data. Accounting, Organizations and Society 32 (April): 201-222.

Asthana, S. and S. Balsam. 2001. The effect of EDGAR on the market reaction to 10-K Filings. *Journal of Accounting and Public Policy* 20 (December): 349-372.

Asthana, S., S. Balsam, S. Sankaraguruswamy. 2004. Differential response of small versus large investors to 10-K filings on EDGAR. *The Accounting Review* 79 (July): 571-589.

Atiase, R. K. 1985. Predisclosure information, firm capitalization and security price behavior around earnings announcements. *Journal of Accounting Research* 23 (Spring): 21–36.

Bamber, L. S. 1987. Unexpected earnings, firm size, and trading volume around quarterly earnings announcements. *The Accounting Review* 62 (July): 510–532.

Bamber, L. and Y. Cheon. 1995. Differential price and volume reactions to accounting earnings announcements. *The Accounting Review* 70 (July): 417-441.

Bamber, L., O Barron, and T. Stober. 1997. Trading volume and different aspects of disagreement coincident with earnings announcements. *The Accounting Review* 72 (October): 575-597.

Bamber, L., O Barron, and T. Stober. 1999. Differential interpretations and trading volume. *Journal of Financial and Quantitative Analysis* 34 (September): 369-386.

Barber, B., T. Odean, and N. Zhu. 2006. Do noise traders move markets? *EFA 2006 Zurich Meetings Paper*

Barclay, M., and J. Warner. 1993. Stealth trading and volatility: Which trades move prices? *Journal of Financial Economics* (December): 281–305.

Barron, O. 1995. Trading volume and belief revisions that differ among individual analysts. *The Accounting Review* 70 (October): 581-597.

Barron, O., D. Byard and C. Enis. 2004. Leveling the information playing field. *Review of Accounting and Finance* 3 (4) 21-46.

Barron, O., D. Harris and M. Stanford. 2005. Evidence that investor's trade on private event-period information around earnings announcements. *The Accounting Review* 80 (2) 403-421.

Beaver, W. H. 1968. The information content of annual earnings announcements. *Journal of Accounting Research* 6 (Supplement): 67-92.

Bhattacharya, N. 2001. Investors' trade size and trading responses around earnings announcements: An empirical investigation. *The Accounting Review* 76 (April): 221–244.

Bhattacharya, N., E. Black, T. Christensen and R. Mergenthaler, 2007. Who trades on pro forma earnings information? *The Accounting Review* 82 (May): 581–619.

Bloomfield, R. 2002. The "Incomplete Revelation Hypothesis" and financial reporting. *Accounting Horizons* 16 (September): 233-243.

Chambers, A. and S. Penman. 1984. Timeliness of reporting and the stock price reaction to earnings announcements. *Journal of Accounting Research* 22 (Spring): 21-47.

Chewning, E. and A. Harrell. 1990. The effects of information loan on decision makers' cue utilization levels and decision quality in a financial distress decision task. *Accounting Organizations and Society* 15 (6): 527-542.

Collins-Thompson, K. and J. Callan. 2005, Predicting reading difficulty with statistical language models, *Journal of the American Society for Information Science and Technology* 56 (November): 1448-1462.

Cox, C. 2007, Closing Remarks to the Second Annual Corporate Governance Summit, University of Southern California, Marshall School of Business. http://www.sec.gov/news/speech/2007/spch032307cc.htm.

Cready, W. M. 1988. Information value and investor wealth: The case of earnings announcements. *Journal of Accounting Research* 26 (Spring): 1-27.

Cready, W. and P. Mynatt. 1991. The information content of annual reports: A Price and Trading Response Analysis. *The Accounting Review* 66 (April): 291-312.

Cready, W. and R. Ramanan. 1995. Detecting trading response using transactions-based research designs. *Review of Quantitative Finance and Accounting* 5 (June): 203-221.

Cready, W. and D. Hurtt. 2002. Assessing investor response to information events using return and volume metrics. *The Accounting Review* 77 (October): 891-909.

Davis, A., J. Piger, and L. Sedor. (2007) Beyond the numbers: an analysis of optimistic and pessimistic language in earnings press releases. University of Oregon Working Paper (September).

Firtel, K. 1999. Plain English: A reappraisal of the intended audience of disclosure under the securities act of 1933, *Southern California Law Review* 72 (March): 851-897.

Francis, J., K. Schipper, and L. Vincent. 2002. Expanded disclosures and the increased usefulness of earnings announcements, *The Accounting Review* 77 (July): 515-546.

Glassman, C. 2005, Remarks at the plain language association international's fifth international conference, http://www.sec.gov/news/speech/spch110405cag.htm.

Griffin, P. A. 2003. Got information? Investor response to form 10-K and form 10-Q EDGAR filings. *Review of Accounting Studies* 8 (December): 433-460.

Grossman, S. and J. Stiglitz. 1980. On the impossibility of informationally efficient markets, *American Economic Review* 70 (June): 393-407.

Gunning, R. 1952. *The technique of clear writing.* McGraw-Hill International Book Co; New York, NY.

Huber, P.J. 1967. The behavior of maximum likelihood estimates under nonstandard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley, CA.

Jones, M. and P. Shoemaker. 1994, Accounting narratives: a review of empirical studies of content and readability. *Journal of Accounting Literature* 13: 142-185.

Kandel, E. and N. Pearson. 1995. Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy* 103 (August): 831-872.

Karpoff, M. 1986. A theory of trading volume. *Journal of Finance* 41 (December): 1069-1087.

Kim, O. and R. Verrecchia. 1991. Trading volume and price reactions to public announcements. *Journal of Accounting Research* 29 (Autumn): 302-321.

Kim, O. and R. Verrecchia. 1994. Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics* 17 (January): 41-67.

Kim, O. and R. Verrecchia. 1997. Pre-announcement and event-period private information. *Journal of Accounting and Economics* 24 (December): 395-419.

Kross, W. and D. Schroeder. 1984. An empirical investigation of the effect of quarterly earnings announcement timing on stock returns. *Journal of Accounting Research* 22 (Spring): 153-176.

Kyle, A. 1985. Continuous auction and insider trading. *Econometrica* (November): 1315–1335.

Lee, C. and M. J. Ready. 1991. Inferring trade direction from intraday data. *The Journal of Finance* 46 (June): 733-746.

Lee, C. 1992. Earnings news and small traders: An intra-day analysis. *Journal of Accounting and Economics* 15 (June/September): 265-302.

Lee C. and B. Radhakrishna. 2000. Inferring investor behavior: Evidence from TORQ data. *Journal of Financial Markets* 3 (2): 83-111.

Levitt, A. 1997. Corporate Finance in the Information Age, remarks delivered to the Securities Regulation Institute, San Diego, California (January 23, 1997).

Li, F. 2008 (forthcoming). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*.

Li, E. and K. Ramesh. 2007. Market reactions surrounding the filing of periodic SEC reports. Michigan State University Working Paper.

Maddala, G. 2001. *Introduction to Econometrics.* 3rd edition: John Wiley & Sons.

Mikhail, M., B. Walther, and Willis, R. (2007). When security analysts talk, who listens? *The Accounting Review*, 82: 1227-1253.

Meulbroek, L. 1992. An empirical analysis of illegal insider trading. *The Journal of Finance* (December): 1661–1699.

Mezick, E. 2001. Investing Wisely: Individual investors as information seekers. *Journal of Business & Finance Librarianship* 7 (1): 3-17

Morse, D. 1981. Price and trading volume reaction surrounding earnings announcements: A closer examination. *Journal of Accounting Research* (Autumn): 374-383.

Muresan, G., M. Cole, C. Smith, L. Liu, and N. Belkin. 2006. Does familiarity breed content? Taking account of familiarity with a topic in personalizing information retrieval. Proceedings of the Hawaii International Conference on System Sciences (HICSS-39).

Nelson, K. and A. Pritchard. 2007. Litigation risk and voluntary disclosure: the use of meaningful cautionary language. *Rice University Working Paper*.

Paredes, T. 2003. Blinded by the light: information overload and its consequences for securities regulation. *Washington University Law Quarterly* 81 (Summer): 417-485.

Radin, A. 2007. Have we created financial statement disclosure overload? *The CPA Journal Online*, (November) http://www.nysscpa.org/cpajournal/2007/1107/perspectives/p6.htm.

Rogers, W. 1993, Regression standard errors in clustered samples, *Stata Technical Bulletin* 13, College Station, TX: Stata Press.

Securities and Exchange Commission. 1998. A plain English handbook: How to create clear SEC disclosure. SEC Office of Investor Education and Assistance. http://www.sec.gov/pdf/handbook.pdf.

Securities and Exchange Commission. 2007. The investor's advocate: How the SEC protects investors, Maintains Market Integrity, and Facilitates Capital Formation *http://www.sec.gov/about/whatwedo.shtml.*

Shanthikumar, D. 2004. Small trade reactions to consecutive earnings surprises. Stanford University Dissertation: 33-89.

Smith. S. 2005. Information quantity, information consistency, and the confidence of unsophisticated investors. Cornell University Dissertation.

Stocks, M. and A. Harrell. 1995. The impact of an increase in accounting information level on the judgment of individuals and groups, *Accounting Organizations and Society* 20 (Oct-Nov): 685-700.

Stocks, M. and B. Tuttle. 1998. An examination of information presentation effects on financial distress predictions. *Advances in Accounting Information Systems* 6: 107-128.

Tetlock, P. 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*. LXII (3): 1139-1168.

Tetlock, P., M. Saar-Tsechansky, and S. Macskassy. 2007. More than words: quantifying language to measure firms' fundamentals. University of Texas at Austin Working Paper. (September).

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–83.

APPENDIX A: 10-K Filing Edit Procedures

This Appendix provides details of the method used to calculate the readability indices. I first download all 10-K reports from EDGAR and then edit the filings before calculating the readability scores. I follow the process documented in Li (2008) with some slight modifications.

I delete the heading information that is contained between <SEC-HEADER> and </SEC-HEADER> and all text that begins with <TABLE> and ends with </TABLE>.[44] I also eliminate all lines that contain <S> or <C> or special characters <...> and <&..>, along with all lines that contain strings such as <TEXT>, <DOCUMENT>, <PAGE>, <TYPE>, or /PRIVACY-ENHANCED/.[45] Further, I convert embedded HTML code to proper ASCII characters and run a Perl package to strip all remaining HTML tags (for further information see http:/www.ascii.cl/htmlcodes.htm). All ".jpg", ".pdf", and ".gif" files are eliminated. Finally, I delete all paragraphs (defined as two lines containing text with a blank line before and after) with more than 50% non-alphabetic characters to ensure that all tables, tabulated texts, and financial statements are eliminated.

Although this approach is not perfect in eliminating all numeric or HTML code, my review of a large number of these filings indicates that the code does a reasonable job at editing the text. Further, the En Fathom module with Perl attempts to distinguish words from acronyms and other characters, which should minimize any noise attributable to remaining HTML code. There is no reason to believe that any remaining HTML code biases the findings in this paper.

To count the number of cells included in the document, I locate all of the HTML tags that correspond to the syntax <TABLE or <table (prior to their deletion as documented above). The number of cells within a table is then computed by counting the number of <TD or <td tags found within a table. Finally, I sum the number of cells across all tables and add that count to the number of words in the document to create the TABLES+WORDS variable.

---

[44] I follow Li (2008) and delete tables "that begin with <TABLE> and end with </TABLE>." However, I find that certain reports contain significant amounts of text between table tags. For example, in Lowe's Companies 1994 10-K (http://www.sec.gov/Archives/edgar/data/60667/0000060667-94-000015.txt), large portions of text, including the entire MD&A section, are eliminated under this approach. Hence, I repeat the entire analysis without deleting the information between the table tags (but keeping the requirement to delete all paragraphs with more than 50% non-alphabetic characters). The results are quantitatively similar using both approaches.

[45] Li (2008) eliminates all paragraphs containing these tags. However, these strings often fail to contain end tags making it difficult to delete specific sections. To address this issue, I eliminate only the lines containing this coding.

APPENDIX B: Calculation of READ_FOG and READ_PE Indices

This Appendix explains the details of measuring length and creating readability indices after the text has been edited (as described in Appendix 1).

FOG Readability

After the file has gone through the edit process described in Appendix 1, the text is then analyzed using the En-Fathom package in Perl.[46] This package calculates a variety of textual statistics, including number of characters, number of words, percent of complex words, number of sentences, number of text lines, number of paragraphs, syllables per word, and words per sentence. Based on these statistics, the package produces the Fog readability index.

Plain English Readability

I use a text-editing program called StyleWriter to compute the plain English readability score used in this paper. This software is designed as an add-in editing program to be used in text processing software, and thus, requires the document to be opened in a 'word processor' before the document can be analyzed. This limitation results in several additional processing requirements.

First, each document must be opened in text processing software and then individually analyzed by StyleWriter.[47] I automate this process using a keyboard/mouse click macro to open and score each document, but the need to open and analyze each document individually severely limits the number of documents that can be efficiently processed. Second, the Perl output in .txt format has a '¶' symbol (i.e., carriage return) at the end of each line. This means StyleWriter sees each line as a paragraph, which invalidates several of the plain English results. Hence, prior to processing the documents, I use a Perl script to replace all '¶' with a space and keep all instances where there were two consecutive '¶' symbols (i.e., actual paragraph break). Finally, the program is designed to write out limited summary statistics to allow writers to improve their writing style. However, as originally designed, the program does not output the frequency of all plain English violations. Therefore, I worked with StyleWriter programmers to modify the program to enable me to export the relevant plain English violations for each document.

---

[46] Li (2008) provides a validation of the precision of the Perl Lingua-En-Fathom module by comparing the module's output with manual calculations of the same text and concludes that although some measurement error exists, there is no reason to believe that the measurement error is systematic or biased.

[47] Additional information about StyleWriter can be found at http://www.stylewriter-usa.com/.

APPENDIX C: Tables

Table 1
Variable Definitions

Variable	Definitions (Alphabetical Order)
ABN_TRADE	- Relevant measure of trading behavior (i.e., EXVOL, AVOL, TRADE_CONS or TRADE_RET)
ABS_RET	- Market adjusted return, defined as the absolute value of the abnormal (market model) share price return over the event period (–1, 3)
AFTEAD	- Number of days after the earnings announcement date, defined as the number of calendar days the 10-K is filed after the preliminary earnings announcement date
AFTEXD	- Number of days after the expected filing date, defined as the number of calendar days the 10-K is filed after the expected filing date (last year's 10-K filing date)
ANUM	- Abnormal number of transactions, defined as the mean daily number of transactions during the event period (–1, 3) minus the mean daily number of transactions during the non-filing period (–49, –5), deflated by the standard deviation of daily number of transactions during the non-filing period (–49, –5)
AVOL	- Abnormal trading volume, defined as the mean daily trading volume during the event period (–1, 3) minus the mean daily trading volume during the non-filing period (–49, –5), deflated by the standard deviation of daily trading volume during the non-filing period (–49, –5)
EARNINGS	- Operating earnings scaled by assets, defined as Compustat items (data178 / data6)
EXNUM	- Event-period excess number of transactions, defined as the natural log of the cumulative number of transactions over the five-day event period (-1,3) minus the natural log of the median number of transactions for contiguous five-day periods during the non-filing period (-49, -5)
EXVOL	- Event-period excess trading volume, defined as the natural log of the cumulative trading volume over the five-day event period (-1,3) minus the natural log of the median volume for contiguous five-day periods during the non-filing period (-49, -5)
LENGTH	- Relevant measure of length (i.e., WORDS or TABLES+WORDS)
LRG	- Large trades - when suffixed to a variable, implies large trades (greater than or equal to $50,000)
NA_FOLL	- Number of analysts following the firm, natural logarithm of one plus the number of unique analysts providing a forecast from the I/B/E/S Detail and Excluded Files
MV	- Market value of equity, defined as the natural logarithm of market value of common equity at the close of the current fiscal year (data25 * data199) from Compustat
MTB	- Market-to-book, defined as the market value of the firm's equity divided by its book value ((data25 * data199) / data6) from Compustat
NBSEG	- Number of business segments, defined as the natural log of one plus the number of business segments
NGSEG	- Number of business segments, defined as the natural log of one plus the number of geographic segments
READ_FOG	- Fog index, calculated as [(words per sentence + percent of complex words) *.40] using Perl's En Fathom module
READ_PE	- Plain English index, calculated using Stylewriter software as described in Appendix 2 ((# of errors attributable to: Passive Verbs + Hidden Verbs + Overwriting + Legal Words & Jargon + Tautologies) / (number of sentences))*10.
READ_SCR	- Relevant measure of readability (i.e., READ_FOG or READ_PE)
SML	- Small Trades - when suffixed to a variable, implies small trades (if share price is < $50, then dollar amount of trade is less than or equal to $5,000; if share price > 50, but less than or equal to 100, then dollar amount of trade is less than or equal to 100 * price; if share price is greater than 100, the firm is excluded from the sample)
TABLES + WORDS	- Natural logarithm of the total of number of words + the number of table cells included in each document
TRADE_ CONS	- Trade Consensus, defined as the absolute value of the daily net buys during the event period (buys minus sales), deflated by the total number of shares traded (buys plus sales) during the event period.
TRADE_ RET	- Average short term return per transaction during the event period (-1,3), where the return for each individual trade is calculated by subtracting the execution price of the trade from the subsequent 5-day closing price all scaled by the price at which the trade executed.
TOT	- Total trades - when suffixed to a variable, implies all trades (i.e., small, medium, and large)
WORDS	- Natural logarithm of the number of words included in the 10-K report

Table 2
Sample Selection

	No. firm/year observations
10-K Filings from Edgar (1994-2006) with GVKEY and PERMO Identifiers and TAQ data	21,189
less where the TAQ ticker changed during the year	-129
less where there was a stock split or stock dividend	-1,523
less where the stock price was less than $1	-122
less where the filing is missing LENGTH and FOG from Perl En Fathom	-4
less where there is missing prior year filing date (AFTEXD)	-3,285
less missing Compustat data (MV, ASSETS, NBSEG, NGSEG, etc.)	-999
less missing I/B/E/S data to calculate analyst following	-1,875
Observations available for primary analysis	**13,252**
Observations available for # of table cells in 10-K available (TABLES+WORDS) calculation	**4,867**

Note: The ability to count the number of table cells requires firms to comply with a specific HTML standard. Hence, TABLES+WORDS is only available for a subset of firms after 1999.

Table 3
Sample Characteristics

Panel A - Distribution of Variables

Variables	# OBS	Mean	Std. Dev.	25th Percentile	Median	75th Percentile
EXVOL_TOT	13252	0.127	0.598	-0.266	0.051	0.438
EXVOL_SML	13252	0.122	0.517	-0.177	0.060	0.352
EXVOL_LRG	13252	0.587	2.392	-0.446	0.083	0.733
EXNUM_TOT	13252	0.121	0.465	-0.166	0.053	0.333
EXNUM_SML	13252	0.115	0.467	-0.166	0.059	0.334
EXNUM_LRG	13252	0.126	0.724	-0.334	0.068	0.560
WORDS	13252	10.390	0.796	9.968	10.387	10.848
WORDS+TABLES	13252	10.828	0.571	10.511	10.818	11.164
READ_FOG	13252	19.946	1.784	18.843	19.769	20.877
READ_PE	13252	21.190	5.254	17.560	19.494	23.746
ABS_RET	13252	0.048	0.054	0.014	0.031	0.063
MTB	13252	1.599	1.709	0.579	1.061	1.935
MV ($ Billion)	13252	2.723	7.408	0.229	0.595	1.768
EARNINGS	13252	0.052	0.156	0.030	0.077	0.125

Panel B - Mean and Median Length Across Time

	# OBS		WORDS		WORDS+TABLES	
Year	WORDS	WORDS+TABLES	Mean	Median	Mean	Median
1995-1996	772	-	10.14	10.09	-	-
1997-1998	2500	-	10.21	10.14	-	-
1999-2000	2694	85	10.32	10.24	10.49	10.44
2001-2002	2242	690	10.31	10.34	10.54	10.58
2003-2004	2570	1895	10.57	10.53	10.85	10.82
2005-2006	2474	2197	10.61	10.60	10.91	10.89

Panel C - Mean and Median Readability Across Time

	# OBS		READ_FOG		READ_PE	
Year	READ_FOG	READ_PE	Mean	Median	Mean	Median
1995-1996	772	772	19.65	19.40	21.29	20.26
1997-1998	2500	2500	19.79	19.65	21.56	20.47
1999-2000	2694	2694	19.73	19.54	21.30	20.05
2001-2002	2242	2242	19.56	19.44	20.80	19.55
2003-2004	2570	2570	20.33	20.05	21.32	20.18
2005-2006	2474	2474	20.37	20.23	20.87	19.90

See Table 1 for variable definitions. To minimize the effects of outliers, all variables are winsorized at the top and bottom 1%.

Table 4
Pearson and Spearman Correlations

Panel A - Small Trades

	EXVOL_SML	EXNUM_SML	TRADE_CONS_SML	TRADE_RET_SML	WORDS	TABLES_WORDS	READ_FOG	READ_PE	MV	EARNINGS
EXVOL_SML		0.9584 ***	-0.0152 **	0.0058	-0.0275 ***	-0.0235 *	-0.0041	-0.0095	-0.0492 ***	-0.0091
EXNUM_SML	0.9536 ***		-0.0180 ***	0.0096	-0.0262 ***	-0.0278 *	-0.0038	-0.0063	-0.0500 ***	0.0043
TRADE_CONS_SML	-0.0380 ***	-0.0496 ***		-0.0662 ***	-0.2588 ***	-0.1721 ***	-0.1203 ***	-0.0061	-0.4838 ***	0.0138 **
TRADE_RET_SML	-0.0029	0.0013	-0.0209		0.0275 ***	-0.0054	0.0035	-0.0105	0.1224 ***	0.0373 ***
WORDS	-0.0244 ***	-0.0275 ***	-0.1978 ***	0.0089		0.9641 ***	0.4674 ***	0.6640 ***	0.2223 ***	-0.1925 ***
TABLES_WORDS	-0.0387 ***	-0.0441 ***	-0.1740 ***	-0.0182	0.8539 ***		0.3700 ***	0.5634 ***	0.2860 ***	-0.1460 ***
READ_FOG	-0.0016	-0.0023	-0.0907 ***	0.0082	0.5383 ***	0.4563 ***		0.5691 ***	0.0278 ***	-0.1276 ***
READ_PE	-0.0032	-0.0041	0.0040	0.0049	0.6349 ***	0.5922 ***	0.6232 ***		-0.0145 *	-0.0888 ***
MV	-0.0102	-0.0100	-0.2225 ***	0.0186 **	0.1222 ***	0.1486 ***	0.0201 **	-0.0359 ***		0.2649 ***
EARNINGS	-0.0336 ***	-0.0227 ***	0.0097	0.0386 ***	-0.0859 ***	-0.0596 ***	-0.0556 ***	-0.0393 ***	0.1284 ***	

Panel B - Large Trades

	EXVOL_LRG	EXNUM_LRG	TRADE_CONS_LRG	TRADE_RET_LRG	WORDS	TABLES_WORDS	READ_FOG	READ_PE	MV	EARNINGS
EXVOL_LRG		0.8734 ***	-0.0167 ***	-0.0056	-0.0440 ***	-0.0351 **	0.0030	-0.0085	-0.1829 ***	-0.0643 ***
EXNUM_LRG	0.6158 ***		-0.0835 ***	0.0021	-0.0399 ***	-0.0309 **	-0.0044	-0.0055	-0.1561 ***	-0.0351 ***
TRADE_CONS_LRG	0.1879 ***	-0.1050 ***		-0.0306 ***	-0.1577 ***	-0.1964 ***	-0.0290 ***	-0.0091	-0.6510 ***	-0.1603 ***
TRADE_RET_LRG	-0.0367 ***	-0.0089	-0.0372 ***		0.0139	-0.0018	-0.0069	-0.0126	0.0722 ***	0.0406 ***
WORDS	-0.0640 ***	-0.0294 ***	-0.1246 ***	0.0037		0.9641 ***	0.4674 ***	0.6640 ***	0.2223 ***	-0.1925 ***
TABLES_WORDS	-0.0958 ***	-0.0300 **	-0.1706 ***	0.0054	0.8539 ***		0.3700 ***	0.5634 ***	0.2860 ***	-0.1460 ***
READ_FOG	-0.0045	-0.0004 **	-0.0259 **	-0.0045	0.5383 ***	0.4563 ***		0.5691 ***	0.0278 ***	-0.1276 ***
READ_PE	-0.0183 **	-0.0076	-0.0140	-0.0055	0.6349 ***	0.5922 ***	0.6232 ***		-0.0145 *	-0.0888 ***
MV	-0.0807 ***	-0.0413 ***	-0.3478 ***	0.0321 ***	0.1222 ***	0.1486 ***	0.0201 **	-0.0359 ***		0.2649 ***
EARNINGS	-0.1331 ***	-0.0354 ***	-0.1586 ***	0.0671 ***	-0.0859 ***	-0.0596 ***	-0.0556 ***	-0.0393 ***	0.1284 ***	

See Table 1 for variable definitions. This table reports Pearson correlation coefficients in the lower-left and Spearman correlation coefficients in the upper-right. Panel A (B) reports correlations for small (large) trades. There are 13,252 for all variables in both Panels, with the exception of TABLES_WORDS which has 4,867. To minimize the effects of outliers, all variables are winsorized at the top and bottom 1%. ***, **, * indicate two-tail statistical significance at the 1%, 5%, and 10% level, respectively.

Table 5
The Effects of Length and Readability on Abnormal Volume

PANEL A - Excess Volume (EXVOL)

	Hyp Sign	EXVOL TOTAL			EXVOL		EXVOL		EXVOL	
					SMALL	LARGE	SMALL	LARGE	SMALL	LARGE
		[1]	[2]	[3]	[4]		[5]		[6]	
WORDS	+ / -	-0.0198 ***		-0.0034	-0.0244 ***	-0.0246			-0.0217 ***	0.0222
		(-2.86)		(-0.49)	(-3.15)	(-0.81)			(-2.60)	(0.84)
TABLES + WORDS	+ / -		-0.0223				-0.0562 ***	-0.0591		
			(-0.96)				(-2.86)	(-1.21)		
READ_FOG	+ / -	0.0059 *	0.0028		0.0049	-0.0056	0.0080	0.0120		
		(1.72)	(0.58)		(1.32)	(-0.71)	(1.51)	(1.10)		
READ_PE	+ / -			-0.0018					0.0008	-0.0115 ***
				(-1.47)					(0.89)	(-2.61)
AFTEAD	-	-0.0041 ***	-0.0077 ***	-0.0041 ***	-0.0018 **	-0.0134 ***	-0.0042 ***	-0.0189 ***	-0.0018 ***	-0.0131 ***
		(-4.94)	(-8.06)	(-4.89)	(-2.26)	(-7.43)	(-5.98)	(-9.31)	(-2.35)	(-7.34)
AFTEXD	+	0.0013 ***	0.0024 ***	0.0013 ***	0.0012 ***	0.0065 ***	0.0010 ***	0.0108 ***	0.0012 ***	0.0065 ***
		(2.37)	(3.07)	(2.42)	(3.12)	(2.67)	(3.06)	(2.75)	(3.19)	(2.68)
ABS_RET	+	3.3626 ***	4.6937 ***	3.3595 ***	3.1032 ***	3.3507 ***	3.6397 ***	6.1653 ***	3.0983 ***	3.3666 ***
		(8.08)	(12.55)	(8.06)	(9.42)	(4.84)	(16.69)	(8.55)	(9.39)	(4.83)
MV	-	-0.0471 ***	-0.0931 ***	-0.0483 ***	0.0065	-0.6457 ***	-0.0379 ***	-0.6570 ***	0.0068	-0.6524 ***
		(-2.46)	(-2.41)	(-2.52)	(0.30)	(-8.97)	(-3.08)	(-5.33)	(0.32)	(-9.17)
MTB	?	-0.0035	-0.0001	-0.0033	0.0141 ***	-0.0466 ***	0.0048	-0.0130	0.0139 ***	-0.0447
		(-0.57)	(-0.01)	(-0.53)	(2.73)	(-2.83)	(1.42)	(-0.44)	(2.72)	(-2.78)
EARNINGS	?	0.0520	-0.1482	0.0567	0.1141 **	-0.7182 **	0.0618	-1.0385 *	0.1154 **	-0.7080 **
		(0.73)	(-1.07)	(0.79)	(1.98)	(-2.27)	(0.76)	(-1.91)	(2.01)	(-2.26)
NBSEG	+ / -	-0.0027	0.0094	-0.0041	-0.0228	0.1036 **	-0.0219	0.0127	-0.0231	0.0994 **
		(-0.14)	(0.30)	(-0.23)	(-0.92)	(2.04)	(-1.19)	(0.15)	(-0.93)	(1.98)
NGSEG	+ / -	0.0320	0.0434	0.0320	0.0214	0.0209	0.0430 **	0.4057 **	0.0213	0.0209
		(1.50)	(1.36)	(1.50)	(0.95)	(0.19)	(2.50)	(1.87)	(0.95)	(0.20)
NA_FOLL	-	-0.0336 **	-0.0303	-0.0345 **	-0.0396 **	-0.2208 ***	-0.0541 ***	-0.1738 *	-0.0400 ***	-0.2219 ***
		(-2.12)	(-0.81)	(-2.19)	(-2.21)	(-3.57)	(-2.99)	(-1.52)	(-2.25)	(-3.57)
Firm Fixed Effects		*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
R-Square		9.69%	16.57%	9.69%		6.95%		7.74%		6.99%
Observations		13,252	4,867	13,252		26,504		9,734		26,504
Test of (SML*LENGTH - LRG*LENGTH) < 0						0.0002		0.0029		-0.0439 #

This table reports the regression results for excess volume (AVOL). Columns 1-3 report results from estimating Eq. (1a) for total excess volume (i.e., small, medium, and large investors combined). Columns 4-6 report results from Eq. (1b) using stacked regressions of small and large investor groups. All variables are defined in Table 1.The constant is not reported as fixed firm effect regressions are estimated. The regressions are performed with clustered robust standard errors (Rogers 1993) to control for within year correlation. To minimize the effects of outliers, all variables are winsorized at the top and bottom 1% by year. ***, **, * indicate two-tail statistical significance of coefficient estimates at the 1%, 5%, and 10% level when no prediction is given and one-tailed significance when predicted. # # #, # #, # indicate one-tail statistical significance of differences between small and large investor coefficients for LENGTH (i.e., WORDS or WORD+TABLES) using F-statistics.

Table 5 (Con't)
The Effects of Length and Readability on Abnormal Volume

PANEL B - Abnormal Volume (AVOL)

	Hyp Sign	AVOL TOTAL			AVOL SMALL / LARGE		AVOL SMALL / LARGE		AVOL SMALL / LARGE	
		[1]	[2]	[3]	[4]		[5]		[6]	
WORDS	+ / -	-0.0262 (-1.50)		-0.0015 (-0.08)	-0.0394 ** (-2.13)	0.0032 (0.20)			-0.0235 (-1.18)	0.0386 ** (2.32)
TABLES + WORDS	+ / -		-0.0638 (-1.26)				-0.1454 *** (-3.00)	-0.0439 (-0.90)		
READ_FOG	+ / -	0.0122 (1.49)	0.0104 (0.78)		0.0142 (1.50)	0.0074 (0.93)	0.0246 ** (2.00)	0.0096 (0.85)		
READ_PE	+ / -			-0.0017 (-0.60)					0.0007 (0.24)	-0.0054 ** (-2.01)
AFTEAD	-	-0.0080 *** (-4.13)	-0.0143 *** (-7.12)	-0.0081 *** (-4.13)	-0.0060 *** (-2.79)	-0.0069 *** (-4.50)	-0.0119 *** (-5.78)	-0.0116 *** (-6.96)	-0.0060 ** (-2.90)	-0.0067 *** (-4.43)
AFTEXD	+	0.0022 ** (1.61)	0.0049 *** (2.41)	0.0022 ** (1.63)	0.0019 ** (2.07)	0.0016 (1.17)	0.0030 *** (3.38)	0.0044 ** (1.95)	0.0019 *** (2.10)	0.0016 (1.18)
ABS_RET	+	6.5850 *** (6.79)	10.4357 *** (7.64)	6.5766 *** (6.77)	7.4668 *** (9.42)	4.4956 *** (6.47)	9.5578 *** (10.59)	7.5714 *** (8.94)	7.4543 *** (9.37)	4.4947 *** (6.45)
MV	-	0.0182 (0.59)	-0.0332 (-0.34)	0.0167 (0.55)	0.0934 *** (3.26)	-0.0378 (-1.05)	-0.0318 (-0.73)	-0.0038 (-0.04)	0.0933 *** (3.22)	-0.0412 (-1.16)
MTB	?	-0.0162 ** (-2.16)	-0.0121 (-0.43)	-0.0161 ** (-2.15)	-0.0082 (-0.77)	-0.0141 ** (-2.09)	0.0107 (0.76)	-0.0282 (-1.16)	-0.0085 (-0.79)	-0.0134 ** (-1.95)
EARNINGS	?	0.0528 (0.60)	-0.2710 (-1.09)	0.0604 (0.69)	0.0382 (0.58)	-0.0039 (-0.04)	-0.1428 (-0.82)	-0.1787 (-0.80)	0.0439 (0.67)	0.0055 (0.05)
NBSEG	+ / -	0.0123 (0.66)	0.0873 (1.56)	0.0101 (0.53)	0.0069 (0.27)	0.0231 (1.13)	0.0060 (0.14)	0.0720 (1.25)	0.0054 (0.21)	0.0200 (0.96)
NGSEG	+ / -	0.0146 (0.61)	0.0051 (0.09)	0.0146 (0.60)	-0.0071 (-0.17)	-0.0011 (-0.05)	0.0097 (0.19)	0.0132 (0.17)	-0.0071 (-0.17)	-0.0012 (-0.05)
NA_FOLL	-	-0.0702 *** (-3.63)	-0.0689 (-0.98)	-0.0716 *** (-3.75)	-0.1002 *** (-2.69)	-0.0524 *** (-2.44)	-0.1205 ** (-2.19)	-0.0348 (-0.55)	-0.1014 *** (-2.77)	-0.0539 *** (-2.47)
Firm Fixed Effects		*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
R-Square		11.24%	16.18%	11.22%	9.48%		12.29%		9.48%	
Observations		13,252	4,867	13,252	26,504		9,734		26,504	
Test of (SML*LENGTH - LRG*LENGTH) < 0					-0.0425 ##		-0.1016 #		-0.0621 ##	

This table reports the regression results for abnormal volume (AVOL). Columns 1-3 report results from estimating Eq. (1a) for total abnormal volume (i.e., small, medium, and large investors combined). Columns 4-6 report results from Eq. (1b) using stacked regressions of small and large investor groups. All variables are defined in Table 1.The constant is not reported as fixed firm effect regressions are estimated. The regressions are performed with clustered robust standard errors (Rogers 1993) to control for within year correlation. To minimize the effects of outliers, all variables are winsorized at the top and bottom 1% by year. ***, **, * indicate two-tail statistical significance of coefficient estimates at the 1%, 5%, and 10% level when no prediction is given and one-tailed significance when predicted. # # #, # #, # indicate one-tail statistical significance of differences between small and large investor coefficients for LENGTH (i.e., WORDS or WORD+TABLES) using F-statistics.

Table 6
The Effects of Table Count and Readability on Abnormal Transactions

PANEL A - Excess Transactions (EXNUM)

	Hyp Sign	EXNUM TOTAL			EXNUM SMALL	EXNUM LARGE	EXNUM SMALL	EXNUM LARGE	EXNUM SMALL	EXNUM LARGE
		[1]	[2]	[3]	[4]		[5]		[6]	
WORDS	+ / -	-0.0281 *** (-4.60)		-0.0250 *** (-3.26)	-0.0257 *** (-3.41)	-0.0150 * (-1.83)			-0.0228 *** (-3.06)	0.0028 (0.32)
TABLES + WORDS	+ / -		-0.0527 *** (-3.81)				-0.0623 *** (-3.57)	-0.0098 (-0.38)		
READ_FOG	+ / -	0.0058 * (1.76)	0.0051 (1.08)		0.0050 (1.42)	0.0056 (1.44)	0.0083 (1.59)	0.0024 (0.37)		
READ_PE	+ / -			0.0010 (0.93)					0.0008 (0.94)	-0.0022 (-1.58)
AFTEAD	-	-0.0022 *** (-3.19)	-0.0042 *** (-5.36)	-0.0023 *** (-3.32)	-0.0017 *** (-2.37)	-0.0058 *** (-4.88)	-0.0038 *** (-4.34)	-0.0101 *** (-7.56)	-0.0018 *** (-2.47)	-0.0058 *** (-4.86)
AFTEXD	+	0.0007 ** (1.75)	0.0013 *** (3.27)	0.0007 ** (1.81)	0.0008 *** (2.54)	0.0018 ** (1.76)	0.0010 *** (3.37)	0.0048 *** (3.72)	0.0009 *** (2.62)	0.0018 * (1.78)
ABS_RET	+	2.9294 *** (9.14)	3.6127 *** (12.13)	2.9236 *** (9.09)	2.8366 *** (9.14)	2.8394 *** (6.78)	3.3451 *** (13.41)	4.6569 *** (8.58)	2.8317 *** (9.09)	2.8369 *** (6.76)
MV	-	-0.0338 * (-1.52)	-0.0647 *** (-3.92)	-0.0334 * (-1.52)	-0.0097 (-0.49)	-0.1448 *** (-6.81)	-0.0524 *** (-5.67)	-0.1508 *** (-2.97)	-0.0094 (-0.48)	-0.1462 *** (-6.93)
MTB	?	0.0002 (0.04)	-0.0064 (-0.92)	0.0000 (0.00)	0.0103 ** (2.15)	-0.0013 (-0.21)	-0.0043 (-1.02)	-0.0326 (-1.57)	0.0101 *** (2.13)	-0.0010 (-0.17)
EARNINGS	?	0.1220 ** (2.14)	0.0594 (0.62)	0.1235 ** (2.20)	0.1195 ** (2.42)	0.0051 (0.05)	0.0828 (1.00)	-0.1321 (-1.41)	0.1209 ** (2.48)	0.0101 (0.11)
NBSEG	+ / -	-0.0251 (-1.07)	0.0073 (0.37)	-0.0254 (-1.08)	-0.0153 (-0.77)	-0.0106 (-0.55)	-0.0109 (-0.59)	0.0398 (1.28)	-0.0156 (-0.78)	-0.0122 (-0.63)
NGSEG	+ / -	0.0269 (1.35)	0.0453 ** (2.09)	0.0269 (1.35)	0.0150 (0.72)	0.0264 (1.08)	0.0446 ** (2.02)	0.0367 (0.63)	0.0150 (0.72)	0.0264 (1.08)
NA_FOLL	-	-0.0505 *** (-2.72)	-0.0465 ** (-1.89)	-0.0509 *** (-2.76)	-0.0551 *** (-2.82)	-0.0411 ** (-2.01)	-0.0624 *** (-3.24)	-0.0175 (-0.31)	-0.0554 *** (-2.86)	-0.0420 ** (-2.04)
Firm Fixed Effects		*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
R-Square		11.88%	17.54%	11.86%		7.70%		11.38%		7.70%
Observations		13,252	4,867	13,252		26,504		9,734		26,504
Test of (SML*LENGTH - LRG*LENGTH) < 0						-0.0107		-0.0525 #		-0.0256 #

This table reports the regression results for excess number of transactions (EXNUM). Columns 1-3 report results from estimating Eq. (1a) for total number of excess transactions (i.e., small, medium, and large investors combined). Columns 4-6 report results from Eq. (1b) using stacked regressions of small and large investor groups. All variables are defined in Table 1.The constant is not reported as fixed firm effect regressions are estimated. The regressions are performed with clustered robust standard errors (Rogers 1993) to control for within year correlation. To minimize the effects of outliers, all variables are winsorized at the top and bottom 1% by year. ***, **, * indicate two-tail statistical significance of coefficient estimates at the 1%, 5%, and 10% level when no prediction is given and one-tailed significance when predicted. # # #, # #, # indicate one-tail statistical significance of differences between small and large investor coefficients for LENGTH (i.e., WORDS or WORD+TABLES) using F-statistics.

Table 6 (Con't)
The Effects of Table Count and Readability on Abnormal Transactions

PANEL B - Abnormal Transactions (ANUM)

	Hyp Sign	ANUM TOTAL			ANUM SMALL	ANUM LARGE	ANUM SMALL	ANUM LARGE	ANUM SMALL	ANUM LARGE
		[1]	[2]	[3]	[4]		[5]		[6]	
WORDS	+ / -	-0.0505 **		-0.0427 **	-0.0457 **	-0.0022			-0.0314	0.0205
		(-2.49)		(-2.17)	(-2.43)	(-0.19)			(-1.57)	(1.27)
TABLES + WORDS	+ / -		-0.1525 ***				-0.1537 ***	-0.0378		
			(-4.72)				(-4.14)	(-0.91)		
READ_FOG	+ / -	0.0117	0.0135		0.0138	0.0072	0.0235 **	0.0034		
		(1.40)	(1.28)		(1.46)	(1.27)	(2.04)	(0.35)		
READ_PE	+ / -			0.0017					0.0009	-0.0027
				(0.66)					(0.34)	(-0.99)
AFTEAD	-	-0.0070 ***	-0.0118 ***	-0.0071 ***	-0.0060 ***	-0.0070 ***	-0.0111 ***	-0.0110 ***	-0.0060 ***	-0.0069 ***
		(-3.28)	(-5.69)	(-3.38)	(-2.69)	(-4.61)	(-4.74)	(-6.83)	(-2.80)	(-4.61)
AFTEXD	+	0.0020 **	0.0038 ***	0.0020 **	0.0020 ***	0.0010	0.0033 ***	0.0042 ***	0.0020 ***	0.0011
		(2.10)	(2.50)	(2.12)	(2.40)	(0.82)	(3.82)	(2.07)	(2.44)	(0.83)
ABS_RET	+	7.5096 ***	10.2560 ***	7.4983 ***	6.8412 ***	4.7952 ***	9.0241 ***	7.0663 ***	6.8288 ***	4.7920 ***
		(8.83)	(9.16)	(8.79)	(8.46)	(7.64)	(10.35)	(7.87)	(8.39)	(7.61)
MV	-	0.0295	-0.0750	0.0302	0.0529 **	-0.0832	-0.0609 **	-0.0714	0.0530 **	-0.0851
		(1.17)	(-1.20)	(1.19)	(1.92)	(-2.81)	(-1.75)	(-0.77)	(1.91)	(-2.91)
MTB	?	-0.0157	-0.0052	-0.0162	-0.0089	-0.0017	-0.0040	-0.0105	-0.0092	-0.0014
		(-1.47)	(-0.21)	(-1.51)	(-0.81)	(-0.24)	(-0.29)	(-0.36)	(-0.84)	(-0.19)
EARNINGS	?	0.0878	-0.0462	0.0913	0.0680	0.0711	-0.0561	-0.1265	0.0734	0.0775
		(1.13)	(-0.21)	(1.19)	(1.04)	(1.04)	(-0.31)	(-1.05)	(1.14)	(1.11)
NBSEG	+ / -	0.0321	0.0943 ***	0.0314	0.0168	0.0183	0.0261	0.0690	0.0155	0.0162
		(1.45)	(2.39)	(1.42)	(0.59)	(0.72)	(0.66)	(1.52)	(0.54)	(0.65)
NGSEG	+ / -	-0.0121	-0.0169	-0.0121	-0.0154	-0.0049	-0.0044	0.0262	-0.0154	-0.0049
		(-0.36)	(-0.30)	(-0.36)	(-0.35)	(-0.25)	(-0.07)	(0.58)	(-0.35)	(-0.24)
NA_FOLL	-	-0.1353 ***	-0.1239 **	-0.1362 ***	-0.1300 ***	-0.0663 ***	-0.1443 ***	-0.0246	-0.1312 ***	-0.0674 ***
		(-4.08)	(-1.75)	(-4.17)	(-3.15)	(-3.74)	(-2.37)	(-0.40)	(-3.22)	(-3.79)
Firm Fixed Effects		*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
R-Square		14.84%	18.17%	14.82%	10.30%		13.08%		10.29%	
Observations		13,252	4,867	13,252	26,504		9,734		26,504	
Test of (SML*LENGTH - LRG*LENGTH) < 0					-0.0436 ##		-0.1159 ##		-0.0519 ##	

This table reports the regression results for abnormal number of transactions (ANUM). Columns 1-3 report results from estimating Eq. (1a) for total number of abnormal transactions (i.e., small, medium, and large investors combined). Columns 4-6 report results from Eq. (1b) using stacked regressions of small and large investor groups. All variables are defined in Table 1.The constant is not reported as fixed firm effect regressions are estimated. The regressions are performed with clustered robust standard errors (Rogers 1993) to control for within year correlation. To minimize the effects of outliers, all variables are winsorized at the top and bottom 1% by year. ***, **, * indicate two-tail statistical significance of coefficient estimates at the 1%, 5%, and 10% level when no prediction is given and one-tailed significance when predicted. # # #, # #, # indicate one-tail statistical significance of differences between small and large investor coefficients for LENGTH (i.e., WORDS or WORD+TABLES) using F-statistics.

Table 7
The Effects of Length and Readability on Trade Consensus

	Hyp Sign	TRADE_CONS SMALL	TRADE_CONS LARGE	TRADE_CONS SMALL	TRADE_CONS LARGE	TRADE_CONS SMALL	TRADE_CONS LARGE
		[1]		[2]		[3]	
WORDS	-	-0.0071 *	-0.0008			-0.0147 **	0.0048
		(-1.49)	(-0.26)			(-1.66)	(1.16)
TABLES + WORDS	-			-0.0163 ***	0.0017		
				(-3.68)	(0.30)		
READ_FOG	-	-0.0009	-0.0003	0.0014	0.0005		
		(-1.00)	(-0.22)	(1.08)	(0.41)		
READ_PE	-					0.0014	-0.0013 **
						(1.71)	(-2.32)
AFTEAD	+	0.0018 ***	-0.0004	0.0008 ***	-0.0001	0.0018 ***	-0.0003
		(9.13)	(-1.70)	(3.63)	(-0.15)	(9.82)	(-1.53)
AFTEXD	?	-0.0002	0.0004	-0.0001	0.0008 ***	-0.0002	0.0004
		(-0.95)	(1.36)	(-0.51)	(3.15)	(-0.95)	(1.37)
ABS_RET	-	-0.1338 ***	-0.4528 ***	0.0116	-0.5971 ***	-0.1344 ***	-0.4513 ***
		(-2.46)	(-6.93)	(0.34)	(-6.10)	(-2.46)	(-6.91)
MV	-	-0.0819 ***	-0.1185 ***	-0.0418 ***	-0.1295 ***	-0.0811 ***	-0.1193 ***
		(-16.25)	(-23.77)	(-9.57)	(-22.67)	(-16.53)	(-23.96)
MTB	?	0.0149 ***	-0.0174 ***	0.0065 ***	-0.0014	0.0147 ***	-0.0172 ***
		(8.85)	(-19.12)	(5.98)	(-0.44)	(8.58)	(-19.03)
EARNINGS	?	0.0463 ***	-0.0736 ***	-0.0301 **	-0.1130 **	0.0444 ***	-0.0723 ***
		(2.94)	(-3.89)	(-2.06)	(-2.04)	(2.96)	(-3.81)
NBSEG	-	-0.0368 ***	0.0104 **	-0.0049	-0.0040	-0.0361 ***	0.0099 **
		(-6.32)	(1.80)	(-1.02)	(-0.33)	(-6.20)	(1.71)
NGSEG	-	0.0286	0.0013	-0.0033	0.0049	0.0286	0.0013
		(3.91)	(0.15)	(-0.47)	(0.28)	(3.97)	(0.15)
NA_FOLL	-	-0.0038	-0.0131 ***	-0.0099 **	0.0101	-0.0035	-0.0132 ***
		(-0.80)	(-2.90)	(-1.76)	(2.51)	(-0.74)	(-2.94)
Firm Fixed Effects		*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
R-Square			15.56%		10.65%		15.63%
Observations			26,504		9,734		26,504
Test of (SML*LENGTH - LRG*LENGTH) < 0			-0.0063		-0.0180 ##		-0.0196 ##

This table reports the regression results for trade consensus (TRADE_CONS). Columns 1-3 report results from Eq. (2) using stacked regression to estimate the trading consensus of small and large investor groups. All variables are defined in Table 1. The constant is not reported as fixed firm effect regressions are estimated. The regressions are performed with clustered robust standard errors (Rogers 1993) to control for within year correlation. To minimize the effects of outliers, all variables are winsorized at the top and bottom 1% by year. ***, **, * indicate two-tail statistical significance of coefficient estimates at the 1%, 5%, and 10% level when no prediction is given and one-tailed significance when predicted. # # #, # #, # indicate one-tail statistical significance of differences between small and large investor coefficients for LENGTH (i.e., WORDS or WORD+TABLES) using F-statistics.

Table 8
The Effects of Length and Readability on Short-Term Trading Returns

	Hyp. Sign		TRADE_RET		TRADE_RET		TRADE_RET	
	SML	LRG	SMALL	LARGE	SMALL	LARGE	SMALL	LARGE
			[1]		[2]		[3]	
WORDS	-	+	-0.0003 **	0.0003			-0.0005 ***	0.0003
			(-1.80)	(1.07)			(-3.08)	(1.59)
WORDS	-	+			-0.0003	0.0000		
					(-1.01)	(0.03)		
READ_FOG	-	+	0.0000	-0.0001	0.0000	-0.0001		
			(0.34)	(-0.58)	(0.11)	(-0.33)		
READ_PE	-	+					0.0001	0.0000
							(1.71)	(-0.73)
AFTEAD	?		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
			(-0.39)	(-0.24)	(0.13)	(0.45)	(-0.53)	(-0.18)
AFTEXD	?		0.0000 *	0.0000	0.0000	0.0000	0.0000 *	0.0000
			(1.91)	(0.22)	(0.76)	(-1.42)	(1.92)	(0.21)
ABS_RET	+		0.0162 ***	-0.0022	0.0155 ***	-0.0119	0.0161 ***	-0.0021
			(6.73)	(-0.38)	(2.74)	(-1.75)	(6.81)	(-0.37)
MV	?		0.0014 ***	0.0008 *	-0.0005	0.0007	0.0014 ***	0.0008 *
			(3.12)	(1.91)	(-1.25)	(0.70)	(3.24)	(1.89)
MTB	?		-0.0003 *	0.0000	-0.0002	0.0000	-0.0003 *	0.0000
			(-1.93)	(-0.20)	(-0.83)	(0.04)	(-1.99)	(-0.17)
EARNINGS	?		0.0048 **	0.0050 ***	0.0048 **	0.0051	0.0048 **	0.0050 **
			(2.33)	(3.82)	(2.19)	(1.19)	(2.29)	(3.83)
NBSEG	-		0.0000	0.0000	0.0007 ***	-0.0005	0.0000	0.0000
			(-0.12)	(-0.31)	(2.39)	(-0.93)	(-0.07)	(-0.34)
NGSEG	-		-0.0004	0.0006	-0.0008	-0.0005	-0.0004	0.0006
			(-0.85)	(1.13)	(-1.05)	(-0.38)	(-0.85)	(1.13)
NA_FOLL	?		-0.0003	0.0001	-0.0003	-0.0010 **	-0.0003	0.0001
			(-1.36)	(0.33)	(-0.64)	(-2.12)	(-1.34)	(0.34)
Firm Fixed Effects			*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
R-Square				0.50%		0.62%		0.51%
Observations				26,504		9,734		26,504
Test of (SML*LENGTH - LRG*LENGTH) < 0				-0.0006 ##		-0.0003		-0.0008 ###

This table reports the regression results for short-term trading profits (TRADE_RET). Columns 1-3 report results from Eq. (3) using stacked regression to estimate the trading profits of small and large investor groups. All variables are defined in Table 1. The constant is not reported as fixed firm effect regressions are estimated. The regressions are performed with clustered robust standard errors (Rogers 1993) to control for within year correlation. To minimize the effects of outliers, all variables are winsorized at the top and bottom 1% by year. ***, **, * indicate two-tail statistical significance of coefficient estimates at the 1%, 5%, and 10% level when no prediction is given and one-tailed significance when predicted. # # #, # #, # indicate one-tail statistical significance of differences between small and large investor coefficients for LENGTH (i.e., WORDS or WORD+TABLES) using F-statistics.

Brian P. Miller, CPA

EDUCATION AND PROFESSIONAL CERTIFICATION

Ph.D.	Accounting, Pennsylvania State University	(Anticipated) 2008
M.B.A.	Finance & International Business, University of Cincinnati	1998
B.A.	Finance & Accounting, Cedarville University	1994

Certified Public Accountant, Ohio (Active)

ACADEMIC EXPERIENCE

The Pennsylvania State University, University Park, PA 2003-Present

Cedarville University, Cedarville, Ohio 2001-2003

Masters College, Santa Clarita, California 2000-2001

PROFESSIONAL EXPERIENCE

Procter and Gamble, Cincinnati, Ohio 1998-2000
Financial Cost Forecast Manager – Fabric Care
Cost Analyst – New Business Development
Financial Analyst – N.A. Fabric Care
Field Study – Corporate New Ventures

B.K.D., LLP (8th largest CPA and Advisory Service Firm), Bloomington, Indiana 1994-1997
Staff Auditor

Pepsi Cola Bottling Co., Indianapolis, Indiana 1993
Sales Tax Project Coordinator – Internship

RESEARCH

Dissertation

- "Data Overload and Investor Trading" - *Dissertation Committee: Paul Fischer (Co-Chair), Andrew Leone, Karl Muller (Co-Chair)*

Research Papers

- "The Importance of Distinguishing Errors from Irregularities in Restatement Research: The Case of Restatements and CEO/CFO Turnover" (with K. Hennes and A. Leone). AAA *2007 Financial Accounting and Reporting Section (FARS) Meeting Paper.* Currently under review at The Accounting Review (Forthcoming).
- "Are Uncontested Director Elections Meaningful Polls?" (with P. Fischer, J. Gramlich, H. White) AAA *2007 Financial Accounting and Reporting Section (FARS) Meeting Paper.* Preparing for resubmission to the Journal of Accounting and Economics.

Other Publications (Course Materials)

- Pop's, Inc. – Managerial Accounting as a Strategic Decision Tool (with J. Austin and K. Schappell), *AICPA Professor Practitioner Case Development Program (2004).*

HONORS, AWARDS, AND ACTIVITIES

Pennsylvania State University: Ossian R. MacKenzie Doctoral Teaching Award nominee (2007)
American Accounting Association Doctoral Consortium Fellow (2007)
Smeal Dissertation Research Award (2006)
G. Kenneth Nelson Scholarship (2005-2008)
J. Kenneth and Nancy N. Jones Graduate Scholarship (2006-2007)
Jane O. Burns Graduate Scholarship (2005-2006)
Donald and Regina Harrison Graduate Scholarship (2004-2005)
Robert B. Graham Graduate Scholarship (2003-2005)

University of Cincinnati: University Scholarship - Scholarship for graduate studies (1997-1998)

Cedarville University: D. R. Wood Scholarship - excellence academics, leadership, athletics
Four Year Varsity Athlete in Cross Country, National Qualifier