A Plain English Measure of Financial Reporting Readability

Samuel B. Bonsall IV
The Ohio State University

Andrew J. Leone
University of Miami

Brian P. Miller
Indiana University

April 2015

ABSTRACT

This study highlights the advantages of using multifaceted measures of readability based on the plain English attributes recommended by the SEC (e.g., passive voice, hidden verbs). Over the past decade a burgeoning literature has emerged using textual analysis to examine the impact of financial reporting readability on the capital markets. Loughran and McDonald (2014) provide evidence that components of the most commonly used measure of readability, the Fog index, are poorly specified in financial applications and suggest an alternative proxy for reporting readability based on the file size of a 10-K document. We find significant measurement error in the use of file size as a proxy for financial reporting readability. Specifically, the vast majority of variation in file size over time and across firms is driven by the inclusion of content unrelated to the underlying text in the 10-K (e.g., HTML, XML, pdf and jpeg file attachments). Further, we document that only these non-textual components of file size are significant determinants of post-filing stock return volatility, which questions the validity of using file size as a proxy for readability. Given these issues, we develop a measure of readability grounded in plain English attributes and then validate our measure using a quasi-exogenous shock provided by the *1998 Plain English Mandate* requiring firms issuing prospectuses to communicate in plain English. We show that compared to the annual filings of firms not filing prospectuses, the firms filing prospectuses experienced significantly greater improvements in plain English measures of readability. In contrast, we find no evidence of decreases in file size, which suggests that total file size does not capture the attributes of readability promoted by the SEC. In summary, we provide evidence consistent with file size having substantial measurement error as a proxy for readability and suggest that measures based on plain English attributes better capture the financial reporting readability attributes highlighted by the SEC.

We are grateful for discussions and helpful comments from Kimball Chapman, Khrystyna Bochkay, Roman Chychyka, Patrick Hopkins, Michelle Hutchens, Jeff McMullin, Sundaresh Ramnath, K.R. Subramanyam, Eric Weisbrod, and workshop participants at Hong Kong Polytechnic University, The Pennsylvania State University, Singapore Management University, and the University of Buffalo. We also express gratitude to the StyleWriter software engineers for programming assistance and Bradley Carrico and Barrett Wheeler for research assistance.

1. Introduction

The Securities and Exchange Commission's (SEC) commitment to making investor communications more readable and understandable dates back to the Securities Act of 1933 (Firtel 1999). More recently the SEC has taken additional specific steps to improve readability and understandability by publishing a plain English guide which provides all financial statement preparers specific principles to improve communications with investors (SEC 1998). Despite this SEC specific guidance, most of the proxies for reporting readability implemented by accounting and finance researchers fail to incorporate these plain English principles.

For finance and accounting researchers, the Fog Index has historically been the primary measure of financial reporting readability. The Fog Index—a measure developed by Gunning (1952) and based on average sentence length and proportion of complex words—gained researcher acceptance largely because it provided a simple and well-known formula for measuring readability.[1] However, as pointed out by Loughran and McDonald (2014), business texts contain an extremely high percentage of words that the Fog Index mechanically classifies as complex because they are multisyllabic but in fact are well understood by investors and analysts (e.g., the word "company"). Loughran and McDonald (2014) show that, due to the uniqueness of business writing, the complex words component of the Fog Index is often poorly specified in financial applications and conclude that the Fog Index does not appropriately capture readability for business documents.

Given the shortcomings of the Fog Index, Loughran and McDonald (2014, p. 1644) "recommend using the file size of the 10-K as an easily calculated proxy for document readability." At first glance, the file size of a 10-K document has some appealing features; most

[1] Examples of studies using the Fog Index include Li (2008); Biddle, Hilary, and Verdi (2009); Miller (2010); Lehavy, Li, and Merkley (2011); Dougal, Engelberg, Garcia, and Parsons (2012); Lawrence (2013); and DeFranco, Hope, Vyas, and Zhou (2014).

notably it is simple to measure and amenable to replication. The measure's simplicity likely explains its increasing use among researchers as a primary measure of textual readability.[2]

Despite its simplicity, a careful examination of the sources of variation in 10-K file size suggests that there are significant theoretical and empirical shortcomings to Loughran and McDonald's (2014) readability proxy. From a theoretical perspective, there is no compelling reason to expect file size to be a sufficient proxy. At best, text length captures one plain English attribute that the SEC highlights as facilitating clear presentation of complex information (i.e., unnecessary details).

For file size to be a useful readability proxy the amount of text in the document (e.g., number of words in a document) would need to be a primary driver of readability and the amount of text in the filing would need to be the primary source of variation in file size. Although the amount of text might capture some aspects of readability related to unnecessary details, the measure does not differentiate the actual language used in the document and is also likely to be at least partially driven by differences in firm complexity (e.g., more lines of business or M&A activity).

Even if text length is an important driver of readability (after controlling for business complexity, M&A activity, etc.), text length explains only a minor portion of the variation in file size.[3] To illustrate, the average 10-K filing size on EDGAR in 2012 is 12.48 megabytes, but the average file size of the text is only 0.28 megabytes. As such, even if there was a 50 percent increase in the size of the text in a 10-K, the total file size would only increase by 1.1 percent. Therefore, it is highly unlikely that file size captures variation in the presentation of unnecessary text.

[2] At the time of this manuscript, the Loughran and McDonald (2014) study had over thirty Google Scholar citations suggesting that a large number of researchers are embracing the new measure.

[3] See the Appendix A for a detailed discussion and examples of forces of variation in file size.

The empirical shortcomings of using file size to measure the amount of unnecessary text primarily stem from the tremendous amount of growth in non-textual components that have little to do with the actual text contained in the 10-K document, the only part of the filing that could plausibly relate to the construct of readability for a financial report user.[4] The growth in these components is illustrated in Figure 1, where not-textual components such as XBRL and HTML, which did not exist in 1994, accounted for 77.3 and 13.8 percent of the total file size in 2012, respectively. During this same time period, the proportion of the 10-K related to text shrank from approximately 35 percent in 1994 to approximately 2 percent in 2012. This evidence further illustrates that there is substantial measurement error in using file size to capture variation in the presentation of unnecessary text.

Given the shortcomings of both the Fog Index and total file size as proxies for readability, we develop a more comprehensive measure of financial reporting readability based on the plain English principles outlined by the SEC. Specifically, we use a computational linguistics software program, *StyleWriter—Plain English Editor*, to develop two multidimensional measures of readability based on the factors outlined by the SEC's *Plain English Handbook* (1998). Our first measure, the *Bog Index*, is *StyleWriter*'s standard measure of readability, which incorporates negative plain English attributes, such as the use of passive verbs, style problems, and long sentences along with positive attributes that make writing more interesting. However, to better ensure that we capture all aspects of the SEC's plain English report, we also construct a summary measure of readability that captures a broad range of plain English attributes. This second measure, *Plain English Factor*, uses factor analysis that combines not only the attributes

[4] For expositional efficiency, throughout the paper we use the term 'non-textual components' of the 10-K filing as a way to describe the file size of anything not related to the text of Form 10-K itself or text in Exhibit 13. This exhibit contains sections of companies' annual reports, such as the Management Discussion and Analysis, that are often incorporated by reference in Form 10-K.

of the *Bog Index*, but also other plain English factors identified by the SEC such as hidden verbs, abstract words, jargon, and document length (i.e., number of words).

We validate our plain English measures using a quasi-exogenous shock provided by the *1998 Plain English Mandate* that required firms issuing a prospectus after October 1, 1998 to use plain English. Under an assumption of compliance with the SEC mandate, we expect prospectus-filing firms to improve the readability of their prospectuses and their related financial filings more than firms unaffected by the regulation (i.e., shelf registrants).[5] Consistent with our predictions, we show that compared to the annual filings of non-prospectus firms that raise capital, the filings of prospectus firms that raise capital experience significantly greater improvements in plain English measures of readability following the *1998 Plain English Mandate*. We also examine how the Loughran and McDonald (2014) readability measure, file size of 10-K filings, changes around the mandate for prospectus filers compared to non-prospectus filers. In contrast to the results we document using the plain English measures of readability, we find no statistically significant decrease in the file size for prospectus filers suggesting that total file size fails to capture the attributes of readability promoted by the SEC.

To further illustrate the potential shortcomings of using file size as a measure of readability, we also examine the time series variation of the textual and non-textual components of file size and find that non-textual components drive most of the time-series variation in the file size of 10-K filings. Additionally, we replicate the main tests from Loughran and McDonald (2014) and find that the non-textual components of the 10-K are significant determinants of post-filing stock return volatility, but the textual component is not. Specifically, we document that file size related to the numerical table data appear to be driving the overall relationship between file size and

[5] It is important to point out that the 1998 plain English regulation requires only the firms filing a prospectus to use plain-language in their prospectus. We predict that the required changes in a firm's prospectus will translate to more clarity in the narrative portion of the 10-K filing.

future stock return volatility. This evidence is consistent with increased volatility stemming from either 1) increased firm complexity or 2) differential interpretation of tabulated data. Although both of these alternatives are economically plausible, neither explanation can be traced to the differential readability of the text included in the financial disclosures. Accordingly, our evidence suggests that file size appears to be a poorly specified measure of financial reporting readability.

We next return to our proposed measures of readability based on the plain English attributes prescribed by the SEC. In addition to validating our plain English measures using the *1998 Plain English Mandate*, we show that both measures of plain English readability are associated with the outcome variables used by Loughran and McDonald (2014). Specifically, both (lack of) readability measures are positively associated with higher return volatility, earnings forecast errors, and earnings forecast dispersion even after including both industry and year fixed effects and controlling for common market and complexity variables used in prior research.

Overall, our measures of readability based on plain English principles appear to exhibit construct validity and are significantly associated with market-based proxies for managers' effectiveness in communicating relevant information to investors. More importantly, unlike many of the existing measures of readability, the components of our readability measures (e.g., active voice) are driven by how the information in the document is communicated and, therefore, it is less likely to be confounded by firm complexity. For instance, our analysis reveals that the use of legal words is actually negatively correlated with other good writing attributes. This evidence suggests that clear financial writing is associated with the use of technical terms, which demonstrates that clear writing does not necessarily lead to the 'dumbing down' of financial disclosures. Based on our evidence, we recommend that researchers interested in studying the implications of readability use measures grounded in the principles of plain English.[6]

The paper proceeds as follows. Section 2 provides a review of the accounting and finance literature related to the textual analysis of financial reporting readability. Section 3 discusses our sample selection procedures, documents the sources for the readability and other measures used in our study, and provides descriptive evidence. Section 4 reports the regression results for the components of file size. Section 5 develops our measures of plain English readability and provides descriptive evidence regarding those measures. Section 6 uses the *1998 Plain English Mandate* to provide a validation of the plain English measures and reports regression results for how the plain English measures relate to various outcome measures used in prior literature that gauge the effectiveness of how managers communicate relevant information to investors and analysts. Section 7 concludes by discussing the benefits of using plain English readability measures and addressing potential concerns over the availability of the plain English measures.

2. Background

2.1. Financial Readability and the SEC

Firtel (1999) discusses that the SEC has been concerned with improving financial statement readability and understandability since the Securities Act of 1933. The SEC more formally examined the readability of financial statements in the *Wheat Report*, which it released in 1969 and concluded that prospectuses were too long and complex for the average investor to understand. Based on these concerns, the report recommended against writing that was unnecessarily long, complex or verbose. To combat concerns over unreadable reports the SEC adopted the *1998 Plain English Mandate*, SEC Rule 421(d). Along with this regulation the SEC

[6] To address the potential concern that the plain English measures proposed in this study are costly to calculate, we plan to make our parsing code and detailed plain English attributes for each 10-K filing publicly available.

provided a companion handbook entitled "*A Plain English Handbook; How to create clear SEC disclosure documents*" (SEC 1998). The handbook encourages plain English disclosures in all investor communications and provided several recommendations to mitigate common problems encountered with disclosure documents.

2.2. Related Literature

Most of the early research investigating the readability of financial statements and footnotes employed very small sample sizes. Jones and Shoemaker (1994) provide an assessment of this literature by summarizing that most financial disclosures are not only difficult to read but are "inaccessible" to a large segment of investors. Recent technological advances and the electronic availability of financial filings through the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) database have enabled researchers to analyze the implications of financial reporting readability on a large population of filings.

Research using textual analysis to evaluate readability of financial reports has increased dramatically over the past decade coinciding with Core's (2001) call for greater use of techniques from computational linguistics to capture large-sample measures of disclosure quality. This burgeoning literature has primarily focused on market responses to the readability of financial reports. Bloomfield (2002) predicts that when information is more costly to extract from larger and more complex disclosures it will result in less trading and will be less completely revealed in market prices. Consistent with these predictions Miller (2010) finds that firms with more readable financial reports have more pronounced small investor trading around the 10-K filing date, while DeFranco, Hope, Vyas, and Zhou (2014) show that trading volume reactions increase with the readability of analysts' reports. Similarly, Lawrence (2013) provides evidence consistent with retail investors being more likely to invest in firms with financial reports that are shorter and more readable, while You and Zhang (2009) document that firms with longer 10-K filings have a larger delay in the market reaction to 10-K filings. Finally, other research shows that more readable reports affect sophisticated intermediaries such as analysts (Lehavy, Li, and Merkley 2011; Bozanic and Thevenot 2014) and bond rating agencies (Bonsall and Miller 2014). Most recently, Lang and Stice-Lawrence (2014) perform a large-sample empirical analysis of annual report textual disclosures for non-U.S. companies and show that firms with improvements in financial reporting experience experienced improvements in several economic outcomes. In summary, this literature suggests that less readable reports have market implications stemming from higher processing costs.

Other studies examine financial reporting readability associations with profitability and future investment. For instance, Li (2008) shows that both less readable and longer reports are associated with lower profitability and lower earnings persistence and interprets this association as evidence of managerial obfuscation when performance is poor. Biddle, Hilary, and Verdi (2009) use readability as a measure of financial reporting quality and show that more readable financial reports are associated with lower over and under-investment.

2.3. The Fog Index

Most of the studies discussed previously rely on the Fog Index as a primary measure of financial reporting complexity/readability. Gunning (1952) developed this index based on average sentence length and proportion of complex words. The Fog Index appears to have gained broad appeal largely because it provides a simple and well-known formula for measuring readability. The Fog Index is comprised of two components: 1) average sentence length and 2) percentage of complex words. These components are added together and multiplied by a scalar to predict a reading grade level, where higher values indicate less readable text:

Fog Index = 0.4 (average number of words per sentence + percentage of complex words) (1)

Although the Fog Index has many advantages in other settings, it is not necessarily a good measure of readability in the financial reporting context. Specifically, Loughran and McDonald (2014) point out that business texts have an extremely high percentage of words that are denoted "complex" by the Fog Index because they are multisyllabic, but in fact are well understood by the vast majority of investors and analysts. Consistent with this notion, the authors show that the complex words component of the Fog Index merely adds measurement error as they fail to find any evidence that it explains post-filing stock return volatility, unexpected earnings, and analyst forecast dispersion.

Given the issues with the Fog Index, Loughran and McDonald (2014) ultimately recommend using the file size of a 10-K filing as an alternative measure of readability. The authors argue for the use of file size as a readability proxy because it "is straightforward, is substantially less prone to measurement error, is easily replicated, is strongly correlated with alternative readability measures" (Loughran and McDonald 2014 p.1644). However, the authors do not provide a clear theoretical basis for why, *ex ante*, file size should be expected to be a valid proxy for plain English readability. Further, as we show in the following sections, file size does contain significant measurement error. For these reasons, we introduce more theoretically based measures of readability that are subject to less measurement error.[7]

3. Sample Selection and Data

3.1. Sample Composition

[7] To clarify, Loughran and McDonald (2014) refer to measurement error in the context of researcher measurement error. That is, multiple researchers attempting to measure file size will virtually always obtain the same values. In this sense, we agree that file size is substantially less prone to researcher measurement error than alternatives. However, as we will show in the following sections, file size inherently contains significant measurement error when it comes to the underlying construct of *readability*.

Table 1 provides a detailed description of the sample selection procedure. We begin with all 10-K documents (e.g., 10-K, 10-K405, 10KSB, and 10KSB40) that contain more than 3,000 words and were filed on EDGAR between 1994 and 2012.[8] We then implement similar data restrictions to those outlined by Loughran and McDonald (2014). Specifically, we require a CRSP PERMNO match (dropping 34,100), a price of greater than $3 (dropping 14,546), available business segment data (dropping 11,408), ordinary common equity (dropping 5,291), a positive book to market (dropping 1,856), available macroeconomic volatility (VIX) and return on assets (ROA) data (dropping 740), a filing date greater than 180 days from prior filing (dropping 102), and pre- and post-market model data available (dropping 245). After these exclusions the final sample used in our primary tests examining post-filing stock return volatility consists of 58,118 observations.[9]

In addition to our market-based tests, for our plain English tests we also consider two other outcome proxies for the effective communication of valuation information: standardized unexpected earnings (*SUE*) and analyst dispersion (*DISPERSION*). To calculate these measures, we require two or more analyst forecasts from I/B/E/S in the time period between the 10-K filing date and the firm's next quarterly earnings announcement. Due to data availability related to these dependent variables our sample size for the *SUE* regressions falls to 43,238, and the sample size for the *DISPERSION* regressions falls to 35,967.

[8] We employ parsing procedures similar to that used by Li (2008) and delete all cleaned filings that contain fewer than 3,000 words or 100 lines of text. Specifically, we download all the raw text files from the EDGAR FTP site and capture the parts of the file contained between <DOCUMENT><TYPE>10-K and </DOCUMENT> (the text of Form 10-K) and, where available, between <DOCUMENT><TYPE>EX-13 and </DOCUMENT> for selected parts of a firm's annual report to shareholders incorporated by reference within Form 10-K. After sub-setting the 10-K filing we remove all HTML tags and decode HTML entities. Further we remove the contents of tables by deleting anything between <TABLE> and </TABLE>. To further ensure that we are not capturing table content, we also remove all "paragraphs" that contain more than 50% non-alphabetic characters.

[9] This compares to 50,739 observations in Loughran and McDonald's (2014) Table VIII where the authors include business segment index as a control variable. A substantial part of the additional observations in our sample is attributable to our sample period including 2012, whereas their sample ends in 2011.

3.2. Descriptive Data

Table 2 reports mean summary statistics for our sample variables. For comparative purposes, we divide the sample into time periods following Loughran and McDonald (2014). Columns (1), (2), (3), and (4) report averages for the entire sample period, filings between 1994 and 2002, filings between 2003 and 2011, and filings made in 2012, respectively. In terms of our dependent variables, we find an overall decrease in the post-filing root mean squared error over time, but we find very little variation in *SUE* and analyst dispersion across the sub-periods. We next examine file size and find substantial variation in total file size across the three time periods examined. In fact, total file size increases from 0.39 MB for filings filed between 1994 and 2002 to total file size of 12.48MB in 2012. However, closer examination indicates most of the variation is due to increases in the non-textual portions of the 10-K filing. For instance, while the file size of the 10-K text doubles across the sample periods (from 0.14 MB to 0.28 MB) the non-textual components of the 10-K filing increase by a factor of more than 48 (from 0.25 MB to 12.20 MB).

Panel A of Figure 1 shows that the vast majority of the yearly variation in file size is driven by components other than the textual component of the 10-K filing. Panel B provides a slightly more detailed breakdown showing that the primary 10-K components of text, table contents, exhibits, and other components contribute very little to total file size. Panel C shows that HTML and XBRL are responsible for large increases in file size over time. HTML formatting became widely used in 2001 and became widely adopted over the next several years as service providers began offering software to generate HTML formatted documents. However, the most significant increases in file size occur in the last few years of our sample period and are attributable to XBRL. The SEC mandated XBRL for all large accelerated after June 15, 2009 and all remaining filers during a phase-in period over the next two years (SEC 2009). This XBRL data is redundant information as it is merely a reformulation of the data already included in the 10-K making it unlikely that the additional file size caused by XBRL affects 10-K readability.

To further illustrate, we take a closer look at the non-textual components of 10-K filings for Apple Incorporated (Apple). We begin by separating all the textual components of the 10-K after excluding HTML, XBRL, PDFs, pictures, and any exhibits unrelated to the actual text of the 10-K filing. After separating the file, we find that in 1994, the first year of our sample, the textual component of the 10-K filing made up 70.6 percent of the total file size. In stark contrast, the textual components of Apple’s 2012 10-K filing made up only 2.6 percent, while XBRL and HTML accounted for 87.1 and 8.3 percent of the total file size, respectively.

We next examine the implications of the non-textual components of file size as a measure of total report readability. Our assumption is that total file size is meant to capture the amount of text in a particular document. As such, Table 3 reports a correlation table summarizing total size, the size of the 10-K text, the non-textual components of file size, and the total words included in the document. Panel A provides the correlations across the entire sample period. Consistent with Loughran and McDonald (2014), we find only a 67.3 percent Pearson correlation between the log of total file size and the Log of the file size related to the textual portion of the 10-K filing. In contrast, we find that the correlation between the log of total file size and the log of file size related to the non-textual portion of the 10-K is 97.0 percent. Combined, this evidence suggests that something besides the text of the document is driving a substantial portion of the variation in total file size.

To further illustrate the point that something beyond the textual content of the 10-K is driving much of the variation in total file size, we next examine how the correlations between file size of the entire 10-K and file size of the textual component of the 10-K change over time. Note that the correlation coefficient for the full sample is not the weighted average of the correlation coefficients for the subsamples. Specifically, the covariance between x and y, for a sample, n, is equal to the weighted average of the covariances of the subsamples (n1 and n2) plus the following term:

$$\frac{n_1 n_2}{n^2}\left(\bar{x}_{n_1} - \bar{x}_{n_2}\right)\left(\bar{y}_{n_1} - \bar{y}_{n_2}\right) \tag{2}$$

Consequently, if the means of x and y for the subsamples are considerably different (as they are in our sample), the covariance of the sample will differ significantly from the covariance of the subsamples. If the term from Equation (2) is positive, as is the case when considering total file size and text file size (i.e., there is an increase in both variables), then the correlation between the variables will be larger for the entire sample than for the subsamples. Thus, in cases where two variables are trending in the same directions (e.g., both increasing over time) it may be more useful to examine the correlations over smaller time periods to remove the influence of the additional term.

To address this issue, we examine the correlations across smaller time periods in Panels B through D. Consistent with the effect of the additional term from Equation (2) increasing the correlation coefficient between the total file size and text file size when we consider the entire sample period, when we re-estimate the correlations over smaller sub-periods we find that the correlations are much smaller than what we observe in the overall sample.

In summary, the evidence from Table 3 suggests that analysis of readability using total file size across any time series becomes extremely problematic as the technological composition of the filing (not the readability) accounts for the vast majority of variation in file size over time.

Further, to the extent that different firms within an industry adopt technology at different times, even within year variation in total file size becomes, at best, a noisy proxy for readability.[10]

4. Regression Analysis of File Size Components

The descriptive evidence raises doubts about the appropriateness of using total file size as a measure of financial report readability. In this section, we examine whether the total file size is a valid measure of financial report readability by examining the relationship between file size and subsequent stock return volatility, which prior research assumes to capture information uncertainty attributable in part to readability. Specifically, we follow Loughran and McDonald (2014) and focus on the root mean squared error from a market model estimated using trading days [6, 28] after the 10-K filing date. As pointed out in their study, the use of a market based measure such as subsequent volatility has an advantage over analyst forecasts because it allows for a less restrictive sample size and is a more inclusive test of all investors (not limited to sophisticated market intermediaries). We expect more readable reports to better convey value relevant information to investors and result in lower post-filing stock return volatility.

4.1. Control Variables

We follow Loughran and McDonald (2014) and include the following controls to explain stock return volatility: 1) *Pre-filing alpha* is the intercept from a market model estimated prior to the filing date; 2) *Pre-filing RMSE* is the root mean squared error from a pre-period market model regression; 3) *Abs(Abnormal Return)* is the absolute value of the 2-day buy-and-hold abnormal return around the filing date (0,1); 4) *Log(market capitalization)* is the natural logarithm of market capitalization on the day before the filing date; 5) *Log(book-to-market)* is

[10] For example, the SEC mandated firms to adopt XBRL in three phases based on filing status, with large accelerated filers that have a public common equity float over $5 billion filing XBRL reports for all fiscal periods ending on or after June 15, 2009. During the second phase, all other large accelerated filers (i.e. public common equity float over $700 million) filed using XBRL for fiscal periods ending on or after June 15, 2010. All remaining filers began filing XBRL statements for fiscal periods on or after June 15, 2011. These differences in tagging are unlikely to drive variation in the actual readability across firms.

the log of the book-to-market ratio calculated prior to the filing date; and 6) *NASDAQ dummy* is a dummy variable set equal to one if the firm trades on NASDAQ, else zero.

In addition to these standard controls, we add additional variables to capture other factors that may potentially affect the relationship between report readability and post-filing return volatility. Because firms with more complex business operations may have greater post-filing return volatility, we include *Business Segment Index,* defined as the sum of the squared business segment proportions as reported for the firm in the COMPUSTAT Historical Segment file. Higher values of the index imply less firm-specific complexity. We control for macro-level uncertainty by including the *Market Volatility Index*, which we measure using the Chicago Board of Options Exchange Volatility Index in the period prior to the filing date. To control for potential effects of performance on future volatility we include *Return on Assets*. To control for the level of accounting detail we include *% of Non-Missing Compustat Items*. Finally, we control for various firm-level events that are likely to affect the amount of text in 10-K filings. Specifically, we include dummy variables as controls for the existence of *M&A Activity, Special Items,* and *Discontinued Operations* during the fiscal year covered by the 10-K filing.

We provide detailed variable descriptions in Appendix B. We winsorize all variables throughout the study by year at the 1st and 99th percentile levels to minimize the potential impact of outliers. Additionally, unless otherwise noted we include an intercept, calendar year dummies, and Fama and French (1997) 48-industry dummies in all regressions. Finally, we use two-way clustered standard errors by year and industry with the corresponding *t*-statistics reported in the parentheses.

Column (1) of Table 4 presents results when we consider only the control variables. We find that with the exception of the controls for firm complexity (*Business segment index*) and the firm level events associated with *Discontinued Operations* all independent variables are statistically significant in explaining root mean squared error. Our results indicate that firms with lower growth (higher book-to-market ratio), higher pre-filing stock market and accounting performance, more non-missing Compustat items, and higher market values tend to have lower subsequent volatility. In contrast, firms with higher pre-filing stock return volatility, with larger absolute filing date returns, listed on the NASDAQ exchange, having M&A activity, with filings containing special items, and filing in periods of greater market uncertainty tend to experience higher subsequent stock return volatility.

4.2. File Size and Subsequent Volatility

Columns (2) through (4) of Table 4 include file size and components of file size as explanatory variables. The results in Column (2) are consistent with Loughran and McDonald (2014) as the coefficient on *Log(Total File Size)* is positive and significant (*t*-statistic=3.86). This suggests that file size is associated with volatility after the filing date. However, since file size consists of both the file size related to the 10-K text and that related to the non-textual components of the 10-K filing, it is unclear which portion of file size is driving the relationship in Column (2).

To address this concern, we separate the filing into the textual and non-textual components to examine which components are associated with future volatility. We begin by separating the portion of the file size related to the text of Form 10-K itself or Exhibit 13, Log(*File Size 10-K Text*), from the other components of file size. Column (3) shows that of the two file size components examined only the non-textual component of file size [*Log(File Size Other)*] is significantly associated with future volatility.

To more clearly identify which of the non-textual components of file size are driving our results, we separate the non-textual portion of the file size into the primary components that exist across our sample period. Specifically, we identify all the numerical data (non-text) that is included in the tables of the 10-K filing, which refer to as Log(*File Size of Table Contents*). We also identify the portion of the document file size related to separate exhibits (e.g., compensation contracts, supplier/customer agreements, or bond indentures) that are unrelated to the 10-K filing requirements but often times are attached to the filing, Log(*File Size of Exhibits*). As previously discussed, the remainder of the document is composed of technical components used to format or re-display information (e.g., HTML, XBRL, PDFs, Pictures). We refer to this portion of the filing as Log(*File Size Other Components*). Column (4) shows that the file size related to table contents is the only component that is significantly associated with future volatility. This result is consistent with the notion that excessive quantitative data lead to differential interpretation of financial performance by investors.[11]

Overall, the evidence from Table 4 suggests that file size related to the textual portion of the 10-K filing which is more likely attributable to the readability of the text does not explain future volatility. Because these are joint tests, the lack of evidence could be due to the fact that 1) using the size of the text in the document provides a crude measure of readability that is too noisy to document a significant relationship in the data or 2) there is no relationship between readability and future volatility. We address this concern later in the paper by examining a quasi-exogenous shock that enables us to observe firms' responses to an SEC regulation requiring a subset of firms to make their financial filings more readable. However, before examining the impacts of

[11] In untabulated analysis, we estimate separate regressions for each component of file size included in columns (3) and (4) of Table 9. Our inferences remain unchanged with the exception that the Log(File Size of Table Contents) and the Log(*File Size of Exhibits*) both become statistically significant at the 0.01 level when estimated in isolation.

this regulation, we first develop a measure of readability based on the plain English attributes suggested by the SEC.

5. Plain English Measures

In this section, we use plain English writing methods described in the SEC's *Plain English Handbook* to develop measures of readability. Many definitions of plain English writing exist and describe plain English as a way to use language to effectively communicate information to the reader. Robert Eagleson, a leading expert on plain English, provides the following definition:

> *Plain English is clear, straightforward expression, using only as many words as are necessary. It is language that avoids obscurity, inflated vocabulary and convoluted sentence construction. It is not baby talk, nor is it a simplified version of the English language. Writers of plain English let their audience concentrate on the message instead of being distracted by complicated language. They make sure that their audience understands the message easily.* Robert Eagleson (2014)

Consistent with these definitions, the SEC's *Plain English Handbook* (1998) provides very clear guidance on plain English writing and issues to avoid when communicating in filings. In the handbook, the SEC lists several distinct problems that they have encountered in their review of regulatory filings including: 1) passive voice, 2) weak or hidden verbs, 3) superfluous words, 4) legal and financial jargon, 5) numerous defined terms, 6) abstract words, 7) unnecessary details, 8) lengthy sentences, and 9) unreadable design and layout.[12] Unfortunately, using simplistic measures such as Fog Index or File Size to capture such multidimensional attributes of plain English is nearly impossible. Fortunately, computational linguistics has progressed to a point where it is possible from a technological standpoint to capture these facets of plain English.

We take advantage of recent software developments made possible by a computational linguistics software program, *StyleWriter—The Plain English Editor*, to develop a multifaceted measure of financial report readability based on the writing factors highlighted in the SEC's

[12] Detailed descriptions and examples of these issues from the SEC's A Plain English Handbook: How to create clear SEC disclosure documents (SEC 1998, pp. 17-35) are included in Appendix C.

Plain English Handbook.[13] We use the software to develop two measures of plain English readability. The first measure, *Bog Index*, is a standard comprehensive measure of readability developed by *StyleWriter* software engineers and based on the principles of plain English writing. Our second measure employs factor analysis to combine the attributes of the *Bog Index* with other specific attributes of plain English writing to create a *Plain English Factor*.

5.1. Bog Index

Technological advances in both software and computing power enable us to assess many more features of writing than word complexity and sentence length upon which measures like the Fog Index rely. The *Bog Index* is a comprehensive measure of readability designed to capture writing features that "bog" readers down. Based on plain English attributes similar to those highlighted by the SEC, the *Bog Index* formula is comprised of three components:

$$\textit{Bog Index} = \textit{Sentence Bog} + \textit{Word Bog} - \textit{Pep} \qquad (3),$$

where a higher *Bog Index* equates to a less readable document.[14]

Sentence Bog identifies readability issues stemming from sentence length, which is calculated using two components. The software first identifies the average sentence length for the document. That average sentence length is squared and then scaled by a default long sentence limit of 35 words per sentence resulting in the following formula:

$$Sentence\ Bog = \frac{(Average\ Sentence\ Length)^2}{Long\ Sentence\ Limit} \qquad (4)$$

Word Bog incorporates multiple plain English style problems (e.g., passive verbs) and word difficulty. Style problems include a broad array of plain English writing attributes highlighted in

[13] Prior studies that incorporate detailed measures of financial statement readability based on plain English attributes include Miller (2010) and Rennekamp (2012). We build on these studies to provide comprehensive measures of plain English and validate these measures by showing that they are associated with alternative readability constructs, consistent with the SEC's 1998 Plain English Regulation, and consistently associated with market based measures used in prior literature.

[14] Wright (2009) provides additional detail of how *StyleWriter* uses plain English attributes to analyze readability.

the SEC's *Plain English Handbook* including passive verbs, hidden verbs, and wordy phrases. The index also captures "heavy words", which is based on word familiarity and not the number of syllables. Accordingly, the StyleWriter program overcomes many of the criticisms outlined by Loughran and McDonald (2014) related to the shortcomings of the Fog Index's crude measure of word complexity based on syllable count. Specifically, the program contains a word list of over 200,000 words based on familiarity, and assesses penalties between zero and four points based on a combination of the word's familiarity and precision (abstract words receive higher scores).[15]

To examine whether the familiarity scores appear reasonable in the financial context, we examine the most frequent multisyllabic words contained in 10-K filings. Appendix D shows the top 100 multisyllabic words included in 10-K filings from 1994 – 2012 which account for over 50 percent of the multisyllabic words used in 10-K filings.[16] As noted by Loughran and McDonald (2014), investors are likely to easily understand many of the multisyllabic words in 10-K filings. Consistent with this notion we show that although the Fog Index penalizes words for having multiple syllables, the penalties assessed by the *Bog Index* are far less severe. Specifically, *StyleWriter* assesses a penalty of no greater than two points out of a possible four points for any word on the list in Appendix D. Further, 43 percent of the multisyllabic words are not assessed any penalty, while another 39 percent of words contain a slight penalty of one out of four. As such, the criticism assessed against *Fog Index* for over-penalizing multisyllabic words does not appear to apply as much to the *Bog Index*.

The measure also creates a penalty for overuse of abbreviations or specialist terms that make the document more difficult to process for readers who may not be familiar with these terms.

[15] For example, *StyleWriter* notes that under their graded word list the word 'attractive' receives no penalty (score of zero), while 'pulchritudinous' receives a four-point penalty.

[16] Note that consistent with the Fog Index we use only stem words (after excluding suffixes such as –ing, -ed, etc.) with more than two syllables.

StyleWriter calculates the *Word Bog* score by compiling violations related to style problems, heavy words, abbreviations, and specialist terminology and multiplying these penalties by 250. Finally, the program scales these penalties from plain English violations by the number of words in the document to create the following formula:

$$Word\ Bog = \frac{(Style\ Problems + Heavy\ Words + Abbreviations + Specialist) * 250}{Number\ of\ Words} \tag{5}$$

Pep identifies writing that makes it easier for a reader to understand. Specifically, the measure accounts for good writing by summing the usage of names, interesting words, and short sentences that make writing more interesting. The measure multiplies the sum of these components by 25 (one tenth of the effect from Bog) and scales by the number of words in the document. Finally, the software accounts for variation in sentence length by adding a measure of sentence variety that *StyleWriter* calculates by taking the standard deviation of sentence length, multiplying that number times ten and scaling by the average sentence length. The combination of these good writing factors results in the following *Pep* formula:

$$Pep = \frac{(Names + Interest\ Words + Conversational) \times 25}{Number\ of\ Words} + \text{Sentence Variety} \tag{6}$$

5.2. Plain English Factor

Although *StyleWriter* bases the *Bog Index* on many of the plain English principles outlined by the SEC, it excludes items such as document length that are important aspects of readability and potentially captures some aspects of readability that may not pertain as much to business writing (e.g., Pep). To more fully capture plain English aspects of report readability highlighted by the SEC, we use factor analysis to create a summary measure of plain English readability based on the *Bog Index* and other plain English attributes, *Plain English Factor*.[17]

[17] Many of the plain English problems used to develop the *Plain English Factor* are used in a measure developed by Miller (2010). However, in contrast to Miller (2010) additional factors are taken into account (e.g., number of words

We begin by mapping the plain English problems identified by the SEC with specific plain English characteristics highlighted in *StyleWriter's* plain English software program. The chart below highlights the comparison between the specific problems identified in Chapter 6 of the SEC's *Plain English Handbook* and the corresponding plain English factors identified by the computational linguistic software.

SEC Plain English Problems	StyleWriter Plain English Components
Passive voice	Passive verbs
Weak/Hidden verbs	Hidden verbs
Superfluous words	Wordy Phrases
Legal and financial jargon	Legal words and Jargon Phrases
Numerous defined terms	Specialist Words
Abstract words	Abstract words
Unnecessary details	Bog Index & # of Words
Long sentences	Long Sentences
Unreadable design and layout	N/A

Although it is impossible to perfectly capture all the dimensions of plain English readability, most of the issues raised by the SEC can be mapped to readability constructs provided by *StyleWriter*. Many of the plain English problems overlap (but not perfectly), which makes it undesirable to simultaneously include all of the measures in our analysis. As such, we use factor analysis to reduce ten plain English problem measures (*passive verbs, hidden verbs, wordy phrases, legal words, jargon phrases, specialist words, abstract words, Bog Index, long sentences, and number of words)* into a single summary measure. Since longer documents are likely to have more plain English problems, we scale each of the following components by the number of sentences in the filing: *passive verbs, hidden verbs, wordy phrases, legal words, jargon phrases, specialist words, and abstract words*. Our *Long sentences* variable is defined as the sum of the sentences in the filing that contain more than 35 words. Finally, we include the *Bog Index* and the natural logarithm of the filing's word count. Given that most of our plain

and the *Bog Index*) and instead of simply adding the components together, we use the more sophisticated empirical methodology of factor analysis to develop our method.

English variables are not normally distributed, we follow Lanen and Larcker (1992) and Rogers and Stocken (2005) and use the asymptotic distribution-free factor estimation method based on Browne (1984).[18]

Results of the factor analysis are reported in Table 5. Panel A shows that all factors, with the exception of legal words, load positively (where positive values capture less readable filings). Panel B shows that the aggregate *Plain English Factor* measure is positively correlated (and statistically significant) with most of the plain English problem measures. The factor is most correlated with the *Bog Index*, which provides some validity that the Bog measure is capturing most of the individual components of plain English highlighted as concerns by the SEC. The only measure that is negatively correlated with the factor is *Legal Words*. We view this as an important take-away from our analysis as the negative relationship is likely due to the fact that in the business context legal words often provide more precise descriptions and a higher frequency of legal words are often associated with more readable filings. As such, our evidence suggests that good financial writing is associated with the use of legal terms, which demonstrates that good writing does not necessarily lead to the 'dumbing down' of financial statements.

5.3. Descriptive Data - Plain English Measures

Table 6 provides descriptive statistics related to the readability measures. Panel A shows that there is a slight increase in both the *Bog Index* and the *Plain English Factor* during our sample period, which indicate that financial statement readability has worsened over time. We also show that the number of words increases from around 19,000 words in the period between 1994 and 2002 to over 43,000 words in 2012. At the same time we find very little variation in the *Fog Index*. For comparative purposes, we also include measures of file size previously documented in

[18] For robustness we repeat the factor analysis using a standard factor method (untabulated) and document that the resulting factor scores are 95% correlated with those from our primary method and all results are comparable to those presented.

Table 2. As previously noted, there is a marked increase in the file size related to the textual component of the document, but the majority of the increase stems from an increase in the file size related to other non-textual components of the filing.

5.4. Correlations between Alternative Readability Measures

Panel B of Table 6 reports the correlations among the *Bog Ind*ex, the *Plain English Factor*, the word count, the Fog Index, file size, and the components of file size. All of the readability measures are significant and positively correlated with both the *Plain English Factor* and the *Bog Index*. We find only a 46.4 (48.6) percent Pearson (Spearman) correlation between the Fog Index and the *Plain English Factor* (*Bog Index*), which suggests that the methods based on plain English attributes are capturing factors beyond the Fog Index. We document a slightly smaller correlation of 45.9 (35.5) percent between the *Plain English Factor* (*Bog Index*) and total file size of the 10-K suggesting that file size fails to capture much of the variability in the plain English factors highlighted by the SEC.

6. Plain English Mandate and the Regression Analysis of Plain English Measures

6.1. Plain English Mandate

In this section, we validate our plain English measures of readability by exploiting a quasi-exogenous shock to disclosure provided by the *1998 Plain English Mandate*, which required firms to use plain English in prospectuses filed after October 1, 1998. To the extent that mandated plain English disclosure improvements in firms' prospectuses spill over to firms' annual filings, this SEC mandate provides us with a unique setting to validate our measure. We examine whether issuing firms that were affected by the disclosure requirement tend to have greater improvements in plain English measures of readability relative to issuing firms that were not affected by the regulation.

We begin by identifying all companies that issued debt or equity securities during 1999 – 2000, the first two years following the adoption of the *Plain English Mandate*. We limit our sample to securities issuers to mitigate concerns that firms issuing securities might inherently have incentives to improve the clarity in their financial reporting.[19] Specifically, we use SDC to identify 1,285 unique firms that issue debt or equity securities during the 1999 – 2000 period. We then identify firms that were subject to the plain English regulation because they issued a prospectus in either 1999 or 2000 (i.e., they are not shelf registrants).[20] The final sample consists of 3,577 firm-year observations for issuers of debt or equity securities, where 2,950 of those firm-year observations correspond to issuers that filed a prospectus in 1999 or 2000 and are thus subject to the regulation. We classify the prospectus filers, *Prospectus_Filer*, as one for the firms that filed a prospectus and zero for the firms that did not file a prospectus. We then create a binary variable, *Post_PE_Regulation*, that is equal to one if the 10-K filing was made during the 1999 to 2000 period and zero if the filing was made in either 1996 or 1997. To avoid any transition year confounds, we exclude any filings made during the initial year of the regulation 1998. Finally, we create an interaction variable, *Prospectus_Filer* Post_PE_Regulation*, from these two main effect variables.

Table 7 provides the regression results of our difference-in-differences analysis related to changes in the dependent variables around the SEC's Plain English Regulation across companies that filed a prospectus with the SEC in either 1999 or 2000 relative to the issuing firms that were

[19] Our results (untabulated) are similar to those reported in Table 7 when we re-run our tests using a propensity score matching of prospectus firms to non-prospectus firms based on a model of seasoned equity offering determinants used by Li and Zhao (2006).

[20] For perspective, while all firms in our sample issue debt or equity securities in 1999 or 2000 only a sub-set of those firms (e.g., those without shelf registrations) are required to issue a prospectus for each bond issuance. The 1998 plain English regulation requires only the firms filing a prospectus to use plain-language in their prospectus, so firms with shelf registrations would be exempt. We examine the 10-K filings (as opposed to the prospectuses) as the issuing firms with shelf-registrations by definition do not file prospectuses. As such, we assume that the mandated disclosure improvements made in firm's prospectuses will carry over to the firm's annual filings.

not subject to the regulation because they did not issue a prospectus. Assuming that the adoption of the Plain English regulation improved the readability of firms' prospectuses and, by extension, their 10-K filings, we expect a negative coefficient on *Prospectus_Filer* Post_PE_Regulation*, if a particular proxy captures the construct of readability.

Column (1) of Panel A reports the results when the outcome variable is the *Bog Index*. Consistent with our prediction we find that the coefficient on the interaction term is negative and significant (*t*-statistic of -3.71), which indicates an incremental improvement in plain English readability for prospectus filers after the regulation relative to non-prospectus filers. Similarly, we also find a negative and significant (*t*-statistic of -3.22) coefficient on the interaction term in Column (2), where the *Plain English Factor* is the outcome variable. In contrast to these results, Column (3) reports an insignificant coefficient on the interaction term when *Log(Total File Size)* is the outcome variable.

Overall, the results from Panel A of Table 7 suggest that prospectus filers improved plain English readability after the *1998 Plain English Mandate* when we use the two plain English proxies to measure readability. In contrast, there is no evidence of improved readability when we use file size. Since the SEC required the prospectus firms to improve readability, the lack of evidence on total file size suggests that this measure likely does not capture the plain English readability factors highlighted by the SEC. The (non) results are also illustrative of the measurement error inherent in using total file size as a measure of readability, which could affect inferences from comparisons across time. In contrast, the plain English measures appear to capture the required improvements in plain English readability as prescribed by the SEC.

To provide additional support for our test using the quasi-experiment of the *1998 Plain English Mandate*, we introduce a placebo test where we observe whether firms issuing prospectuses have similar improvements in readability around another period not affected by the regulation. Specifically, we re-estimate the same regression specification on firm-years spanning 1994 – 1997 as a falsification test using January 1, 1996 as a fictitious-shock date. Similarly to our main plain English test, we identify all securities issuing firms during 1996 – 1997, where 3,391 of the 4,117 firm-year observations for issuers of debt or equity securities file a prospectus during 1996 or 1997. We use the variable *Prospectus_Filer_False* to designate these prospectus issuers and *Post_PE_Regulation_False* to designate firm years during 1996 and 1997. In contrast to our tests around the actual regulation, we would not expect to find a significantly negative coefficient on the interaction term *Prospectus_Filer_False* Post_PE_Regulation_False*. Panel B of Table 7 reports the results of these regressions, where we fail to find any evidence of a decrease in the readability measure around the fictitious regulation date. This evidence is consistent with our prior results stemming from the mandated disclosure, and not fundamental differences in readability between issuing firms that are required to file a prospectus and those that are not required to file.

6.2. Regression Analysis of Plain English Measures

We next examine whether the plain English measures explain subsequent stock return volatility, as well as other measures of the information environment that focus on the effect of financial report readability on analysts' ability to assimilate disclosed information into earnings projections. This enables us to examine the effectiveness of the plain English readability measures on both the overall market and sophisticated market intermediaries.

6.2.1. Impact of Plain English Measures in Regressions of Root Mean Squared Error

Table 8 reports the coefficient values and significance levels when we regress root mean squared error after the 10-K filing date on the plain English measures. These regressions are similar to those reported in Table 4, except that we replace the file size measures with the plain

English readability measures. Column (1) reports that the coefficient on *Bog Index* is positive and significant (*t*-statistic of 3.37). Similarly, Column (2) of Table 8 shows that the coefficient on the *Plain English Factor* is also positive and significant (*t*-statistic 2.33).

We next address the economic significance of the results. The standard deviation of the *Post-filing RMSE* is 1.98, while the standard deviations of *Bog Index* and the *Plain English Factor* are 7.76 and 7.88, respectively, The results from the regression indicate that a one standard deviation increase in *Bog Index* leads to a 2.3% increase in the standard deviation of subsequent volatility ((.006*7.76)/1.98). Similarly, a one standard deviation increase in the *Plain English Factor* results in a 1.2% increase in subsequent volatilities standard deviation ((.003*7.88)/1.98). The economic magnitude of the *Plain English Factor* in predicting subsequent volatility is limited, but reasonable when considering that readability is likely not a primary determinant of subsequent stock return volatility. Overall, the evidence from Table 8 is consistent with both plain English measures being associated with future volatility in a manner that suggests that the measures gauge how effectively managers convey value relevant information to investors.

6.2.2. Impact of Plain English Measures in Regressions of Analyst Processing of Information

Loughran and McDonald (2014 p. 23) discuss the root mean squared error as “the most inclusive surrogate for readability.” However, they also examine other information environment measures that focus more on the ability of analysts to process and incorporate disclosed information into earnings projections. For completeness, we consider the same dependent variables, earnings surprises and analyst earnings forecast dispersion, which are based on the ability of analysts to incorporate 10-K filing information.

We begin by examining the impact of the plain English measures on standardized unexpected earnings (*SUE*). We assume that reports written with fewer plain English problems would be expected to have smaller earnings surprises, where we define *SUE* as follows:

$$SUE = \left| \frac{(Actual\ Earnings - Average\ Expected\ Earnings)}{Stock\ Price} \right| \quad (7),$$

where actual earnings and mean analyst forecasts are obtained from the I/B/E/S unadjusted detail files. We use the absolute value of *SUE* (regardless of whether it is a positive or negative surprise) to assess the magnitude of the earnings surprise. The I/B/E/S data requirements reduce our sample observations from 58,118 in our primary tests to 43,238 when we use *SUE* as the dependent variable.

Consistent with our post-filing volatility (RMSE) regressions we continue to include the control variables, as well as industry and calendar year fixed effects. We also control for the number of analysts following a firm since analyst following may affect earnings surprises. Column (1) of Table 9 includes the control variables with the absolute value of *SUE* as the dependent variable. The results indicate that higher absolute earnings surprises are associated with smaller firms that have worse performance, have higher volatility prior to filing, have larger filing returns, are listed on either the NYSE or AMEX exchange, have higher book-to-market ratios, engage in less M&A activity, and have more special items.

In columns (2) and (3) we include the plain English measures of readability. The coefficient on *Bog Index* is positive and weakly significant in one tailed tests (*t*-statistic of 1.59). In contrast, the *Plain English Factor* is both positive and significant at conventional levels (*t*-statistic 2.14). This indicates that more readable 10-K filings, as measured by the plain English factors, have lower absolute earnings surprise. The differences in significance across the two measure of readability are likely due to the more comprehensive nature of the *Plain English Factor* compared to the *Bog Index*.

In terms of economic significance, we find that a one standard deviation increase in the *Plain English Factor* results in approximately a 2.6% increase in the standard deviation of abs(SUE)

((.003*7.88)/0.91). Consistent with our discussion above on post-filing volatility, one would not expect readability to be a major determinant of earnings surprises, but the results indicate that the plain English measures have a modest impact. In sum, our results provide evidence that the measures of plain English readability capture the clarity of the textual component of the filing in terms of analyst earnings surprises, but that the more comprehensive measure of plain English readability appears to be more robust. Based on our validation of the plain English measures using the Plain English Mandate, the results using earnings surprises as the outcome variable (as opposed to overall market measures) are also consistent with readability having a less pronounced impact on sophisticated information intermediaries.

6.2.3. Impact of Plain English Measures in Regressions of Analyst Dispersion

We next follow Lehavy et al. (2011) and Loughran and McDonald (2014) and examine regression equations where analyst dispersion is the dependent variable. Consistent with prior literature, we expect that less readable reports will make it more difficult for analysts to forecast earnings, which will likely result in greater analyst dispersion. The results in column (4) of Table 9 show that higher analyst dispersion is associated with smaller firms that have lower pre-filing performance, have higher prior stock return volatility, have higher book to market ratios, are listed on the NYSE or AMEX exchange, have more special items, and are followed by more analysts.

Column (5) includes *Bog Index* as an additional explanatory variable. The results indicate that the coefficient is positive and marginally significant at conventional rejection levels (*t*-statistic=1.91). Consistent with the tests related to abs(SUE) the significance of the coefficient on *Plain English Factor* in Column (6) is greater in significance (*t*-statistic=2.65). We find that the *Plain English Factor* also has a slightly greater impact in terms of economic significance, where a one standard deviation increase in the *Plain English Factor* leads to an increase in analyst dispersion that is 3.7% of its standard deviation ((.002*7.88)/0.42). By comparison, a one standard deviation increase in *Bog Index* results in 2.8% increase in the standard deviation of analyst dispersion ((.001*7.76)/0.42). Collectively, the results from Table 9 suggest that plain English measures capture the way that firms communicate relevant financial information to analysts but that the more comprehensive measure, *Plain English Factor*, is more robust to the analyst information environment.

7. Conclusion

The SEC's 1998 mandate of plain English writing in prospectuses and their release of guidelines for plain English communication in all public filings has highlighted the need for research on the textual attributes of firm disclosures. Much of the literature using textual analysis has relied on simplistic readability metrics such as the Fog Index to measure financial disclosure readability. Loughran and McDonald (2014) examine the efficacy of using this measure in the context of financial disclosures and provide evidence that components of the Fog index are poorly specified in financial applications. The authors suggest an alternative measure of readability based on the file size of a 10-K document.

We contend that file size is not based on the plain English attributes recommended by the SEC and contains both textual and non-textual components. We show that non-textual components (e.g., HTML, XBRL, PDFs, pictures, etc.) drive a substantial portion of the variation in the file size of 10-K filings over time and that these components appear to explain the market responses to total file size documented by Loughran and McDonald (2014).

Given the shortcomings of the existing readability measures, we suggest measures of readability based on the plain English attributes highlighted by the SEC (e.g., passive voice, hidden verbs, etc.). We show that these plain English measures of readability are associated with the effectiveness of managers' communication of relevant information to investors. We also validate these measures using a quasi-exogenous shock provided by the *1998 Plain English Mandate* requiring firms issuing prospectuses to communicate in plain English. Based on our analyses, we recommend that future researchers consider plain English measures of readability that better capture the financial readability attributes highlighted by the SEC.

The proposed measures of readability that are based on plain English attributes have many appealing advantages as proxies for financial reporting readability. One potential disadvantage of the plain English measures relative to file size is that they are time-consuming to calculate. To address this potential disadvantage, we plan to make publicly available our parsing code and detailed plain English attributes for each 10-K filing. This should allow future researchers to have a common plain English measure without investing inordinate amounts of time calculating the measures for each filing.

Our paper has regulatory implications for the SEC as it contemplates overhauling current disclosure policies through their "disclosure project" (SEC 2014). Loughran and McDonald (2014) suggest that the "the SEC should focus less on style—which is undifferentiated in 10-Ks—and instead encourage managers to write more concisely." We agree that concise writing is an important attribute of clear communication, but we caution against measuring the textual component of a document with crude measures such as file size that can be distorted by non-textual factors. Further, our results show that plain English based measures effectively communicate 10-K information and contribute to the incorporation of that information into stock prices and analyst forecasts. As such, we contend that SEC should not focus on one attribute of plain English writing (potentially measured with error) to the exclusion of other more robust writing attributes.

References

Biddle, G., G. Hilary, and R. Verdi, 2009. How does financial reporting quality relate to investment efficiency? *Journal of Accounting and Economics* 48, 112–131.

Bozanic, Z. and M. Thevenot. (2014). Qualitative Disclosure and Changes in Sell-Side Financial Analysts' Information Environment. Forthcoming *Contemporary Accounting Research*.

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

Bonsall, S. and B. Miller. 2014. The Impact of Narrative Financial Disclosure Complexity on Bond Ratings and Rating Agency Disagreement. The Ohio State University and Indiana University Working Paper.

Browne, M. 1984. Asymptotically Distribution-Free Methods for the Analysis of Covariance Structure. *British Journal of Mathematical and Statistical Psychology*, 37(1): 62–83.

Core, J. 2001. A review of the empirical disclosure literature: A discussion. *Journal of Accounting and Economics,* 31: 441-456.

De Franco, G., O.K. Hope, D. Vyas, and Y. Zhou. 2014, Analyst report readability, *Contemporary Accounting Research*, forthcoming.

Dougal, C., J. Engelberg, D. Garcia, and C. Parsons, 2012, Journalists and the stock market, *Review of Financial Studies* 25, 639–679.

Eagleson, R. 2014. Short Definition of Plain Language. http://www.plainlanguage.gov/whatisPL/definitions/eagleson.cfm

Fama, E., and K. French. 1997. Industry costs of equity. *Journal of Financial Economics* 43, 153–193.

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.

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

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

Lanen, W. and D. Larcker. 1992. Executive Compensation Contract Adoption and in the Electric Utility Industry. *Journal of Accounting Research*, 30(1): 70–93.

Lang, M. and L. Stice-Lawrence. 2014. Textual Analysis and International Financial Reporting: Large Sample Evidence. University of North Carolina working paper.

Lawrence, A. 2013. Individual investors and financial disclosure (2013). *Journal of Accounting & Economics* 56: 130-147.

Lehavy, R., F. Li, and K. Merkley. 2011. The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review*, 85 (6): 2107-2143.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. Journal of *Accounting and Economics,* (August): 221-247.

Li, X. and Zhao, X. 2006. Propensity score matching and abnormal performance after seasoned equity offerings. *Journal of Empirical Finance* 13: 351-370.

Loughran, T. and B. McDonald. 2014. Measuring readability in financial disclosures. *Journal of Finance* 69(4) 1643-1671.

Miller, B.P. 2010. The effects of disclosure complexity on small and large investor trading. *The Accounting Review*, 85 (6): 2107-2143.

Rennekamp, K. 2012. Processing fluency and investors' reactions to disclosure readability. *Journal of Accounting Research*, 50 (5) 1319-1354.

Rogers, J. and P. Stocken. 2005. Credibility of Management Forecasts. *The Accounting Review*, 80(4): 1233–1260.

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. 2013. The path forward on disclosure. Speech given by SEC Chair Mary Jo White to the National Association of Corporate Directors. http://www.sec.gov/News/Speech/Detail/Speech/1370539878806#.UrBX57Rojvc

Securities and Exchange Commission. 2014. Disclosure effectiveness: remarks before the American Bar Association Business Law Section Spring Meeting. http://www.sec.gov/News/Speech/Detail/Speech/1370541479332#.U4Zb-SjrB3s

Wright, N. (2009). Towards a better readability measure – the Bog Index. http://s3-eu-west-1.amazonaws.com/plcdev/app/public/system/files/36/original/TowardsABetterReadabilityMeasure.pdf

You, H. and X. Zhang. 2009. Financial disclosure complexity and investor underreaction to 10-K information. *Review of Accounting Studies*. 14 (4): 559-586.

Appendix A
Illustration of the Components of File Size

This appendix provides an overview of the variety of factors that affect the total file size of a 10-K filing. These include but are not limited to:

1. *The number of words (characters) in included in form 10-K.*

2. *The number of tables.*

3. *The number of words included in any supplemental exhibits*. These exhibits can include something as short as an Auditor's consent letter, or as long as a merger agreement. These exhibits can have a major impact on total file size. For example, in 1999, the size of Apple Inc.'s (Apple) 10-K was 320 KB, but after inclusion of exhibits (e.g., employee stock purchase plan) the total submission file size was 502 KB. In the next year, Apple's 2000 10-K was 305 KB (four percent decrease), but it contained almost no exhibits, so the total submission file size was only 311 KB. Simply using total file size would suggest a 39 percent decrease in "readability" when submission total file size is used as the proxy. Worse, it is likely that exhibits occur more often when significant corporate events occur. Consequently, researchers using file size to measure readability may instead be inadvertently using a proxy for the existence of corporate events that may be correlated with outcome variables of interest (e.g., uncertainty). As such, the use of file size as a measure of readability could lead to erroneous inferences.

4. *The method in which a 10-K is rendered for viewing*. The majority of 10-Ks were filed in text format prior to 2000. After 2000, many companies began submitting filings in HTML. The addition of HTML added significantly to file sizes. For example, Apple's 10-K file size was 311 KB in 2000, but more than doubled to 792 KB in 2001. However, of the 481 KB increase in file size during this year, 467KB was due solely to HTML. A researcher using file size as a measure of readability would misinterpret this as a decrease in readability that was instead merely due to the use of HTML to render the document. Related to this issue, even if firms elect to use HTML, differences in implementation and syntax can lead to differences in file size caused by differences in the amount of HTML content. Continuing with Apple, for example, between 2004 and

2005, Apple’s 10-K increased from 966 KB to 3,186 KB. However, when we separate the document into text and html, we find the entire increase was due to changes in HTML likely stemming from differences in implementation and syntax.

5. *Binary files (e.g., JPEG, PDF, GIF, etc.) are embedded in the filing*. Companies are permitted to include image files (e.g., graphics) or PDF files in a document submission. These files are most often duplicate versions of the 10-K but in an alternative binary format. For example, Northport Network Systems’ 2011 10-K filing includes twenty GIF files that are reproductions of its financial statements and footnotes. While the total file size of the Northport Network Systems is 222,891 KB, 99.8 percent, or 222,535 KB, is related to these GIF files. Because of the relative size of these binary files, researchers using file size as a measure of financial readability would likely misclassify filings Northport Network Systems (and others like this filing) among the most unreadable filing, when in reality the file size increase is merely due to the way that the information is formatted.

6. *Addition of XBRL rendering*. The 2009 mandate requiring firms to tag their financial statement elements has led to exponential growth in file size. The regulation allowed a three-year phase-in period, where larger filers were required to adopt XBRL in their reporting first. The file size of these XBRL related components dwarfs the other components in the filing. For example, the total file size of Apple’s 2012 10-K is 9,860 KB, but 8,588 is attributable to XBRL. Similar to the other non-textual components previously discussed, these dramatic increases in XBRL could be misinterpreted as enormous decreases in readability by researchers using file size as a proxy for readability.

Appendix B
Variable Definitions

Variable Names	Definitions
Readability Measures:	
Total File Size	is the file size (in megabytes) of the complete 10-K filing from EDGAR.
File Size 10-K Text	is the file size (in megabytes) of the cleaned text component of the 10-K filing (excluding appendices not related to the 10-K filing) from EDGAR.
File Size Other	is the file size (in megabytes) of all components of the document that are not related to the text of the 10-K filing from EDGAR.
File Size Table Contents	is the file size (in megabytes) of data contained in the tables (e.g., data included between table tags) of the 10-K filing from EDGAR.
File Size Exhibits	is the file size (in megabytes) of the exhibits (e.g., compensation contracts, supplier/customer agreements, or bond indentures) of the 10-K filing from EDGAR.
File Other Components	is the file size (in megabytes) of other components (e.g., XML, HTML, PDF's, pictures, etc.) contained in the 10-K filing from EDGAR.
Bog Index	is a proprietary measure of readability created by Editor Softwares's plain English software, *StyleWriter* . As descrbed in greater detail in Appendix A, the formula is based on the plain English components of 'Sentence Bog', 'Word Bog', and 'Pep'. Higher values of the index imply lower readability.
Plain English Factor	is the readability measure based on factor analysis of the individual plain English problems highlighted by the Securities and Exchange Commission. The specific components of the factor are derived from Editor Softwares's plain English software, *StyleWriter* . Higher values of the Plain English Factor imply lower readability.
Fog Index	is the readability measure based on the Gunning Fog index, which is calculated as 0.4*(average number of words per sentence + percent of complex words). The Fog Index is based on the cleaned text component of the 10-K filing (excluding appendices not related to the 10-K filing) from EDGAR. Higher values of the Fog Index imply lower readability.
Words	is the number of words (in 000's) contained in the cleaned text component of the 10-K filing (excluding appendices not related to the 10-K filing) from EDGAR.
Dependent Variables:	
Post-filing RMSE	is the root mean squared error from a market model multiplied by 100. The model is estimated using trading days [6,28] relative to the 10-K file date, where a minimum of 10 observations are required to be included in the sample.
Abs(SUE)	is the absolute value of the standardized unexpected earnings (SUE) multiplied by 100. SUE is defined as the [actual earnings-average expected earnings]/stock price where at least one analyst making a forecast is required to be included in the sample. The actual earnings and mean analyst forecasts are obtained from the I/B/E/S unadjusted data files. To avoid stale forecasts, we include only forecasts occurring between the 10-K filing date and the next earnings announcement date. For analysts with more than one forecast reported during this time interval we retain only the forecast closest to the filing date in the sample.
Analyst dispersion	is the standard deviation of analysts' forecasts appearing in the SUE estimate divided by the stock price from before the 10-K filing date multiplied by 100. We retain only firms with at least two analyst forecasts.
Control Variables:	
Pre-filing alpha	is the alpha from a market model multiplied by 100. The model is estimated using trading days [-252, -6] relative to the 10-K file date, where a minimum of 60 observations of daily returns must be available to be included in the sample.
Pre-filing RMSE	is the root mean squared error from a market model multiplied by 100. The model is estimated using trading days [-257, -6] relative to the 10-K file date, where a minimum of 60 observations of daily returns must be available to be included in the sample.
Abs(abnormal return)	is the absolute value of the filing date excess return. The buy-and-hold return period is measured from the filing date (day 0) through day +1 minus the buy-and-hold return of the CRSP value-weighted index over the same 2-day period.
Market capitalization	is the CRSP stock price times shares outstanding on the day prior to the 10-K filing date (in $ millions).
Book-to-market	is the firm's book-to-market, using data from both Compustat (book value from most recent year prior to filing date) and CRSP (market value of equity). Firms with negative book values are removed from the sample.
NASDAQ dummy	is a dummy variable that is set to one if the firm is listed on NASDAQ at the time of the 10-K filing, else zero.
Business segment index	is the sum of the squared business segment proportions reported for the firm in the COMPUSTAT Segment file based on sales data.
Market Volatility Index	is the Chicago Board of Options Exchange Volatility Index, defined as the average level of the index during the 30 days prior to the 10-K filing.
Return on Assests	is the firms return on assets, defined as operating income after depreciation for the fiscal year prior to the 10-K filing from Compustat.
% Non-Missing Compustat Items	is a measure of a firms complexity based on the percentage of non-missing items in Compustat. The measure is calculated as the number of items with non-missing values in Compustat scaled by the total number of items availabe in Compustat.
M&A activity dummy	is a dummy variable that is set to one if the firm engages in any acquisition activity during the fiscal year covered by the 10-K filing, else zero
Special items dummy	is a dummy variable that is set to one if the firm reports any special items in Compustat during the fiscal covered by the 10-K filing, else zero
Discontinued operations dummy	is a dummy variable that is set to one if the firm reports any results for discontinued operations during the fiscal year covered by the 10-K filing, else zero.
# of Analysts	The number of analysts used in the Analyst dispersion calculation.
Regulation Variables:	
Propsectus_Filer	is one if a firm filed a prospectus with the SEC after the regulation (i.e., 1999 or 2000) and zero otherwise.
Propsectus_Filer_False	is one if a firm filed a prospectus with the SEC after the *fictitious* regulation (1996 or 1997) and zero otherwise.
Post_PE_Regulation	is one if the 10-K was filed after the regulation (i.e., 1999 or 2000) and zero if the 10-K was filed during 1996 or 1997.
Post_PE_Regulation_False	is one if the 10-K was filed after the *fictitious* regulation (i.e., 1996 or 1997) and zero if the 10-K was filed during 1994 or 1995.

Appendix C

Examples of Plain English Problems[1]

Long sentences

Sentences that are packed with too much information can be hard to understand. Often, these sentences contain excessive jargon and legalese.

Example problem:

The following description encompasses all the material terms and provisions of the Notes offered hereby and supplements, and to the extent inconsistent therewith replaces, the description of the general terms and provisions of the Debt Securities (as defined in the accompanying Prospectus) set forth under the heading "Description of Debt Securities" in the Prospectus, to which description reference is hereby made. The following description will apply to each Note unless otherwise specified in the applicable Pricing Supplement.

Example rewrite:

We provide information to you about our notes in three separate documents that progressively provide more detail: 1) the prospectus, 2) the prospectus supplement, and 3) the pricing supplement. Since the terms of specific notes may differ from the general information we have provided, in all cases rely on information in the pricing supplement over different information in the prospectus and the prospectus supplement; and rely on this prospectus supplement over different information in the prospectus.

Passive voice

The subject of the sentence is acted upon by a person or object. Often the person or object doing the action is introduced with the word "by." Sometimes, though, that person or object is omitted altogether.

Example problem:

The foregoing Fee Table **is intended** to assist investors in under- standing the costs and expenses that a shareholder in the Fund will bear directly or indirectly.

Example rewrite:

This table describes the fees and expenses that you may pay if you buy and hold shares of the fund.

[1] Definitions and examples come from "A Plain English Handbook: How to create clear SEC disclosure documents," an August 1998 publication by the SEC Office of Investor Education and Assistance.

Appendix C (Con't)
Examples of Plain English Problems

Weak verbs

Sentences with weak verbs take strong verbs and turn them into a noun, usually with the suffix, "-tion." These noun forms of verbs are less vigorous and more abstract than using the verb itself.

Example problem:

There is the possibility of prior Board **approval** of these investments.

Example rewrite:

The Board might **approve** these investments in advance.

Superfluous words

Often a writer can replace phrases in sentences with fewer words that have the same meaning. Lists of adjectives can sometimes be replaced with a single word or phrase.

Example problem:

Drakecorp has **filed** with the Internal Revenue Service **a tax ruling request concerning, among other things,** the tax **consequences** of the Distribution to the United States holders of Drakecorp Stock. It is expected **that the Distribution of Beco Common Stock to the shareholders of Drakecorp** will be tax-free **to such shareholders** for federal income tax **purposes,** except **to the extent** that cash is received for fractional share **interests.**

Example rewrite:

While we expect that this transaction will be tax free for U.S. shareholders at the federal level (except for any cash paid for fractional shares), we have asked the Internal Revenue Service to rule that it is.

Legal and financial jargon

Use of document specific acronyms, industry terms, and legalese can bog a reader down and reduce their understanding of the document.

Example problem:
NLR Insured Mortgage Association, Inc., a Delaware corporation ("NLR MAE"), which is an actively managed, infinite life, New York Stock Exchange-listed real estate investment trust ("REIT"), and PAL Liquidating REIT, Inc., a newly formed, finite life, self-liquidating Delaware corporation which intends to qualify as a REIT ("PAL Liquidating REIT"), hereby jointly offer, upon the terms and subject to the conditions set forth herein and in the related Letters of

of Transmittal (collectively, the "Offer"), to exchange (i) shares of NLR MAE's Common Stock, par value $.01 per share ("NLR MAE Shares"), or, at the option of Unitholders, shares of PAL Liquidating REIT's Common Stock, par value $.01 per share ("PAL Liquidating REIT Shares"), and (ii) the right to receive cash payable 60 days after closing on the first of any Acquisitions (as defined below) but in no event later than 270 days (nine months) following consummation of the Offer (the "Deferred Cash Payment"), for all outstanding Limited Partnership Interests and Depository Units of Limited Partnership Interest (collectively, "Units") in each of PAL Insured Mortgage Investors, a California limited partnership ("PAL 84"), PAL Insured Mortgage Investors - Series 85, A California Limited Partnership, a California limited partnership ("PAL 85"), and PAL Insured Mortgage Investors L.P. - Series 86, a Delaware limited partnership ("PAL 86"). See "THE OFFER."

Numerous defined terms

Defined terms at the beginning of a document discourage a reader from moving into the main body of the document. These terms overwhelm readers and likely reduce the likelihood that they read or comprehend the document.

Abstract words

These are terms that do not easily create an image in a reader's mind. Often a writer can replace an abstraction with a concrete example.

Example problem:

Sandyhill Basic Value Fund, Inc. (the "Fund") seeks **capital appreciation** and, secondarily, income by investing in securities, primarily equities, that management of the Fund believes are **undervalued** and therefore represent **basic investment value.**

Example rewrite:

At the Sandyhill Basic Value Fund, we will strive to increase the value of your shares (capital appreciation) and, to a lesser extent, to provide income (dividends). We will invest primarily in undervalued stocks, meaning those selling for low prices given the financial strength of the companies.

Appendix D
Bog Scores for First Quartile of Most Frequently Occurring Multi-Syllabic Words in 10-Ks

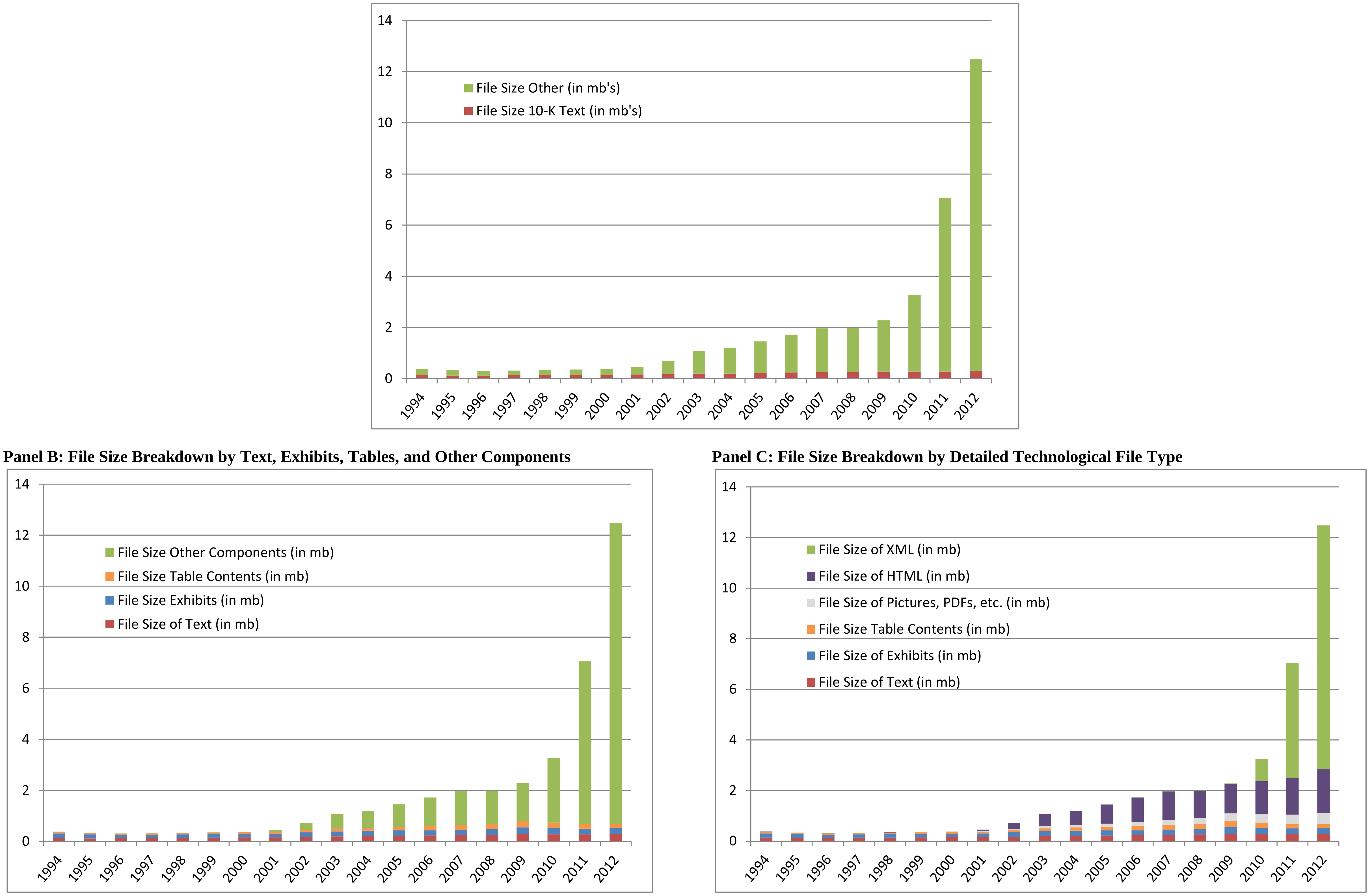
This Appendix provides a summary of stem words with more than two syllables. The sample contains word from 58,104 10-K's filed between 1994 and 2012. The Word Bog Score is a measure of word difficulty derived from the StyleWriter Plain English Software. Words receive a score of 0, 1, 2, 3, or 4, where less familiar / less precise (abstract) receive higher scores.

	Word Bog Score	*Percent of Total Complex Words*	*Cummulative Percentage*
COMPANY	0	5.15%	5.15%
FINANCIAL	0	2.36%	7.51%
STATEMENT	0	1.83%	9.34%
INTEREST	0	1.28%	10.62%
AGREEMENT	1	1.18%	11.80%
OPERATION	2	1.11%	12.91%
BUSINESS	0	1.10%	14.01%
SECURITY	1	1.01%	15.03%
MANAGEMENT	1	0.92%	15.95%
CONSOLIDATE	2	0.88%	16.83%
OPERATE	2	0.84%	17.67%
REVENUE	1	0.82%	18.49%
PERIOD	0	0.79%	19.28%
DIRECTOR	0	0.76%	20.04%
APPROXIMATELY	2	0.70%	20.74%
CAPITAL	0	0.68%	21.43%
INVESTMENT	1	0.68%	22.10%
CUSTOMER	0	0.67%	22.77%
ESTIMATE	0	0.65%	23.41%
PROPERTY	1	0.62%	24.03%
INFORMATION	1	0.59%	24.63%
CONTINUE	0	0.59%	25.21%
OFFICER	0	0.57%	25.79%
SUBSIDIARY	1	0.57%	26.35%
FACILITY	2	0.53%	26.89%
MATERIAL	1	0.51%	27.39%
CORPORATION	1	0.51%	27.90%
DEVELOPMENT	1	0.50%	28.40%
ACQUISITION	2	0.48%	28.87%
COMPENSATION	1	0.47%	29.34%
ADDITIONAL	2	0.47%	29.81%
EXECUTIVE	0	0.46%	30.27%
EQUITY	1	0.46%	30.73%
ANNUAL	1	0.46%	31.19%
INCORPORATE	1	0.46%	31.65%
SIGNIFICANT	1	0.45%	32.10%
CONDITION	1	0.44%	32.54%
TRANSACTION	1	0.44%	32.98%
REGISTRANT	1	0.43%	33.41%
REFERENCE	0	0.43%	33.84%
PARTNERSHIP	0	0.42%	34.25%
ACTIVITY	2	0.41%	34.67%
RESPECTIVELY	1	0.41%	35.08%
OUTSTANDING	0	0.41%	35.49%
PRIMARILY	2	0.41%	35.90%
GENERAL	0	0.41%	36.30%
BENEFIT	0	0.40%	36.70%
LIABILITIES	1	0.40%	37.10%
ADDITION	1	0.38%	37.48%
EXHIBIT	2	0.38%	37.86%
EFFECTIVE	1	0.35%	38.21%
REQUIREMENT	2	0.35%	38.56%
GENERALLY	2	0.35%	38.91%
TECHNOLOGY	1	0.35%	39.26%
FEDERAL	0	0.34%	39.60%
PURSUANT	2	0.34%	39.93%
RECOGNIZE	0	0.34%	40.27%
REGULATION	1	0.33%	40.60%
DETERMINE	2	0.33%	40.93%
LIMITED	0	0.33%	41.26%
AVAILABLE	1	0.32%	41.58%
INSURANCE	1	0.32%	41.91%
EQUIPMENT	1	0.32%	42.22%
PRINCIPAL	0	0.31%	42.54%
AVERAGE	0	0.31%	42.85%
OBLIGATION	2	0.30%	43.14%
DISCLOSURE	1	0.29%	43.43%
STOCKHOLDER	0	0.29%	43.72%
PROVISION	1	0.28%	44.01%
INDUSTRY	0	0.28%	44.29%
POLICY	0	0.28%	44.57%
POSITION	0	0.27%	44.85%
PRESIDENT	0	0.27%	45.12%
DIVIDEND	0	0.27%	45.39%
ABILITY	0	0.27%	45.66%
DISTRIBUTION	1	0.27%	45.92%
INTERNAL	0	0.27%	46.19%
ASSOCIATE	0	0.27%	46.46%
COMMISSION	0	0.26%	46.72%
CURRENTLY	0	0.25%	46.97%
INSTRUMENT	1	0.25%	47.22%
SHAREHOLDER	0	0.25%	47.47%
MANUFACTURE	2	0.25%	47.72%
DEPOSIT	0	0.23%	47.96%
EXERCISE	0	0.23%	48.19%
PERFORMANCE	1	0.23%	48.42%
VARIOUS	0	0.23%	48.66%
DEVELOP	0	0.23%	48.89%
IMPAIRMENT	2	0.23%	49.12%
LIABILITY	1	0.23%	49.35%
ESTABLISH	2	0.23%	49.58%
REPRESENT	0	0.22%	49.80%
MARKETING	0	0.22%	50.02%
ACCORDANCE	1	0.22%	50.24%
CONNECTION	1	0.21%	50.45%
ALLOWANCE	1	0.21%	50.66%
SENIOR	0	0.21%	50.87%
COMMERCIAL	1	0.21%	51.08%
PRODUCTION	1	0.21%	51.29%
EXISTING	0	0.20%	51.49%

Figure 1
Breakdown of 10-K File Size Components by Year

This figure shows the breakdown of the components of 10-K file size by year. Panel A provides a breakdown by year of the file size related to the text of the 10-K filing and non-textual components included in the 10-K filing. Panel B provides a breakdown by year of how the primary components of text, table contents, exhibits, and other components (e.g., HTML, XML, etc.) contribute to file size. Panel C provides a more detailed breakdown of the effects of HTML, XML, Pictures, and PDFs on file size.

Panel A: File Size Attributable to 10-K Text and Other Factors



Panel B: File Size Breakdown by Text, Exhibits, Tables, and Other Components

Panel C: File Size Breakdown by Detailed Technological File Type

Table 1
Sample Creation

This table reports the sample selection criteria for 10-K filings in our sample.

	Dropped	Sample Size
SEC 10-K files 1994 to 2012 with words > 3,000 (after eliminating duplicated within year/CIK)		126,406
CRSP PERMNO match	34,100	92,306
Price on filing date day minus one ≥ $3	14,546	77,760
Business Segment Data Available	11,408	66,352
Reported on CRSP as ordinary common equity	5,291	61,061
Book-to-market COMPUSTAT data available and book value > 0	1,856	59,205
VIX data available	430	58,775
COMPUSTAT data available	310	58,465
File date > 180 days from prior filing	102	58,363
Post-filing date market model RMSE for days [6,28]	148	58,215
At least 60 days data available for market model estimates from event days [-252,-6]	97	58,118

Table 2
Variable Means by Time Period, 1994-2012

This table reports descriptive information for the key variables in our sample. Appendix A provides detailed variable definitions. Our primary regressions examining volatility subsequent to the 10-K filing are based on the entire sample of 58,118 observations. This sample is reduced to 43,238 observations for abs(SUE) and 35,967 for Analyst dispersion and the # of analysts. Detailed variable definitions are provided in Appendix A.

Variable	**1994 to 2012**	**1994 to 2002**	**2003 to 2011**	**2012**
Dependent Variables:				
Post-filing RMSE	2.91	3.55	2.28	1.77
Abs(SUE)	0.47	0.47	0.47	0.46
Analyst dispersion	0.25	0.24	0.25	0.27
Readability Measures:				
Total File Size (in mb)	1.71	0.39	2.35	12.48
File Size 10-K Text (in mb)	0.19	0.14	0.24	0.28
File Size Other (in mb)	1.52	0.25	2.11	12.20
Control Variables:				
Pre-filing alpha	0.06	0.08	0.05	0.01
Pre-filing RMSE	3.18	3.65	2.70	2.33
Abs(abnormal return)	0.03	0.04	0.03	0.03
Market capitalization (in $ millions)	$2,556.39	$1,801.64	$3,259.01	$4,558.73
Book-to-market	0.61	0.62	0.59	0.67
NASDAQ dummy	0.55	0.56	0.54	0.51
Business segment index	0.46	0.60	0.30	0.40
Market Volatility Index	21.13	21.79	20.40	20.78
Return on Assests	0.04	0.02	0.07	0.08
% Non-Missing Compustat Items	0.33	0.29	0.37	0.40
M&A activity dummy	0.14	0.16	0.11	0.13
Special items dummy	0.54	0.45	0.63	0.71
Discontinued operations dummy	0.13	0.07	0.19	0.17
# of Analysts	6.71	6.13	7.11	8.61
Number of Observations	58118	29895	26120	2103
Observations SUE	43238	21180	20279	1779
Observations Analysts	35967	17101	17246	1620

Table 3
Correlation Matrix for Readability Measures

This table reports correlations for various readability measures based on the file size of the 10-K filing, where Pearson (Spearman) correlations are presented below (above) the diagonal. The bolded correlations are significant at less than 1%. Panel A provides correlations for all 58,118 observations in our sample. Panel B provides correlations for the 29,895 observations that occur between 1994 and 2002. Panel C provides correlations for the 26,120 observations that occur between 2003 and 2011. Panel D provides correlations for the 2,103 observations that occur in 2012. Detailed variable definitions are provided in Appendix A.

Panel A: Correlation Matrix for Readability Measures (1994-2012)

	Log(Total File Size)	*Log(File Size 10-K Text)*	*Log(File Size Other)*	*Log(# of Words)*
Log(Total File Size)	1	**0.7074**	**0.9824**	**0.7681**
Log(File Size 10-K Text)	**0.6736**	1	**0.6048**	**0.9580**
Log(File Size Other)	**0.9703**	**0.5781**	1	**0.6787**
Log(# of Words)	**0.7437**	**0.9187**	**0.6621**	1

Panel B: Correlation Matrix for Readability Measures (1994-2002)

	Log(Total File Size)	*Log(File Size 10-K Text)*	*Log(File Size Other)*	*Log(# of Words)*
Log(Total File Size)	1	**0.5750**	**0.9161**	**0.6087**
Log(File Size 10-K Text)	**0.5488**	1	**0.2776**	**0.9556**
Log(File Size Other)	**0.8996**	**0.2545**	<.0001	**0.3367**
Log(# of Words)	**0.5970**	**0.9231**	**0.3354**	0

Panel C: Correlation Matrix for Readability Measures (2003-2011)

	Log(Total File Size)	*Log(File Size 10-K Text)*	*Log(File Size Other)*	*Log(# of Words)*
Log(Total File Size)	1	**0.5280**	**0.9959**	**0.5622**
Log(File Size 10-K Text)	**0.5028**	1	**0.4622**	**0.9468**
Log(File Size Other)	**0.9794**	**0.4171**	1	**0.5029**
Log(# of Words)	**0.5689**	**0.9010**	**0.5007**	1

Notes: Pearson (Spearman) correlations are presented below (above) the diagonal. The bolded correlations are significant at less than 1%.

Panel D: Correlation Matrix for Readability Measures (2012)

	Log(Total File Size)	*Log(File Size 10-K Text)*	*Log(File Size Other)*	*Log(# of Words)*
Log(Total File Size)	1	**0.4941**	**0.9998**	**0.5133**
Log(File Size 10-K Text)	**0.5236**	1	**0.4816**	**0.9677**
Log(File Size Other)	**0.9998**	**0.5100**	1	**0.5015**
Log(# of Words)	**0.5475**	**0.9392**	**0.5343**	1

Table 4
An Analysis of Prior Readability Measures Using Post-Filing Date Market Model Root Mean Squared Error (RMSE) as the Dependent Variable

This table reports regression results where the market model root mean squared error for trading days [6, 28] relative to the 10-K filing date is the dependent variable. All regressions include an intercept, calendar year dummies, and Fama and French (1997) 48-industry dummies. Standard errors are clustered by year and industry. The t-statistics are reported in parentheses below the coefficient. All regressions include 58,118 firm-year observations during 1994 to 2012. Detailed variable definitions are provided in Appendix A.

	(1)	(2)	(3)	(4)
	RMSE	RMSE	RMSE	RMSE
Readability Measures:				
Log(Total File Size)		0.046		
		(3.86)		
Log(File Size 10-K Text)			0.161	0.015
			(1.17)	(1.11)
Log(File Size Other)			0.023	
			(3.36)	
Log(File Size Table Contents)				0.008
				(1.99)
Log(File Size Exhibits)				0.004
				(1.12)
Log(File Size Other Components)				-0.001
				(-0.11)
Control Variables:				
Pre-filing alpha	-0.541	-0.531	-0.531	-0.533
	(-2.50)	(-2.45)	(-2.44)	(-2.46)
Pre-filing RMSE	0.589	0.587	0.586	0.587
	(26.47)	(26.15)	(26.28)	(26.23)
Abs(abnormal return)	4.287	4.285	4.282	4.284
	(15.72)	(15.70)	(15.78)	(15.74)
Log(market capitalization)	-0.083	-0.091	-0.091	-0.089
	(-6.53)	(-7.29)	(-7.20)	(-6.77)
Log (book-to-market)	-0.148	-0.152	-0.152	-0.151
	(-5.13)	(-5.22)	(-5.16)	(-5.25)
NASDAQ dummy	0.212	0.213	0.213	0.212
	(4.17)	(4.17)	(4.15)	(4.15)
Business segment index	-0.020	-0.016	-0.017	-0.018
	(-0.39)	(-0.31)	(-0.33)	(-0.35)
Market Volatility Index	0.032	0.031	0.032	0.032
	(3.03)	(3.00)	(3.02)	(3.02)
Return on Assests	-0.316	-0.313	-0.313	-0.315
	(-11.54)	(-11.12)	(-11.07)	(-11.24)
% Non-Missing Compustat Items	-1.620	-1.585	-1.593	-1.598
	(-2.03)	(-2.01)	(-2.01)	(-2.01)
M&A Activity	0.065	0.061	0.061	0.061
	(2.62)	(2.51)	(2.52)	(2.55)
Special Items	0.037	0.031	0.031	0.033
	(2.73)	(2.34)	(2.48)	(2.38)
Discontinued Operations	0.018	0.011	0.011	0.012
	(1.07)	(0.64)	(0.65)	(0.72)
Industry Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
# of OBS	58118	58118	58118	58118
R^2	0.5666	0.5668	0.5667	0.5667

TABLE 5
Factor Loadings for Summary Plain English Measure

This table provides information on the factor analysis used to create the summary plain English measure. Panel A provides the factor loadings and factor score coefficients. Panel B provides correlation matrices of the *Plain English Factor* and the plain English components, where Pearson (Spearman) correlations are presented below (above) the diagonal. The bolded correlations are significant at less than 1%.

Panel A: Factor Loadings for Summary Plain English Measure

	Factor Loading	**Factor Score Coefficients**
Component		
Passive Verbs	0.4305	0.6029
Hidden Verbs	0.6478	3.5217
Wordy Phrases	0.7575	2.0989
Legal Words	-0.1660	-0.5766
Jargon Phrases	0.8737	3.2177
Specialist Words	0.3919	0.2037
Abstract Words	0.5585	0.7249
Bog Index	0.9499	0.0698
Long Sentences	0.7736	0.6794
Log (# of Words)	0.7412	0.1665

Panel B: Correlation Matrix for Summary Plain English Measure and Plain English Components

	Plain English Factor	*Passive Verbs*	*Hidden Verbs*	*Wordy Phrases*	*Legal Words*	*Jargon Phrases*	*Specialist Words*	*Abstract Words*	*Bog Index*	*Long Sentences*	*Log (# of Words)*
Plain English Factor	1	**0.3899**	**0.8142**	**0.6788**	**-0.1031**	**0.7195**	**0.3146**	**0.4381**	**0.8125**	**0.6505**	**0.6292**
Passive Verbs	**0.3937**	1	**0.2912**	**0.3368**	**0.0355**	**0.1544**	**0.0740**	**0.0561**	**0.3423**	**0.3460**	**0.2772**
Hidden Verbs	**0.8244**	**0.2895**	1	**0.5301**	**-0.0583**	**0.3396**	**0.1551**	**0.2097**	**0.5079**	**0.4971**	**0.4803**
Wordy Phrases	**0.6835**	**0.3555**	**0.5357**	1	**0.1162**	**0.2246**	**0.0249**	**0.1285**	**0.4600**	**0.6997**	**0.4623**
Legal Words	**-0.0968**	**0.0788**	**-0.0445**	**0.1753**	1	**-0.1611**	**-0.0877**	**-0.0315**	**-0.0618**	**0.1659**	**-0.2213**
Jargon Phrases	**0.7299**	**0.1455**	**0.3441**	**0.2116**	**-0.1553**	1	**0.4531**	**0.4964**	**0.8077**	**0.3007**	**0.4571**
Specialist Words	**0.3553**	**0.0427**	**0.2142**	**0.0221**	**-0.0777**	**0.5018**	1	**0.1417**	**0.4155**	**0.0935**	**0.1038**
Abstract Words	**0.4311**	**0.0421**	**0.1847**	**0.1037**	**-0.0948**	**0.4873**	**0.1200**	1	**0.4803**	**0.1796**	**0.2194**
Bog Index	**0.8209**	**0.3525**	**0.5251**	**0.4724**	**-0.0367**	**0.8026**	**0.4582**	**0.4769**	1	**0.4441**	**0.4979**
Long Sentences	**0.6482**	**0.3602**	**0.4986**	**0.7202**	**0.2486**	**0.2855**	**0.0606**	**0.1487**	**0.4447**	1	**0.5165**
Log (# of Words)	**0.6352**	**0.2613**	**0.4799**	**0.4557**	**-0.1669**	**0.4453**	**0.0810**	**0.2028**	**0.5081**	**0.5099**	1

Table 6
Descriptive Statistics for Readability Measures

This table reports descriptive information and correlations for the key readability measures. Panel A provides descriptive evidence for the entire sample of 58,118 observations, as well as subsamples during various time periods. Panel B provides correlations using the entire sample of 58,118 observations, where Pearson (Spearman) correlations are presented below (above) the diagonal. The bolded correlations are significant at less than 1%. Detailed variable definitions are provided in Appendix A.

Panel A: Readability Measures by Time Period, 1994-2012

Variable	**1994 to 2012**	**1994 to 2002**	**2003 to 2011**	**2012**
Plain English Readability Measures				
Bog Index	82.38	80.20	84.57	86.08
Plain English Factor	-0.01	-3.07	3.05	5.52
Other Readability Measures				
# of Words (in 000's)	27.42	19.12	35.63	43.66
Fog Index	19.36	19.14	19.57	19.87
File Size Measures				
Total File Size (in mb)	1.71	0.39	2.35	12.48
File Size 10-K Text (in mb)	0.19	0.14	0.24	0.28
File Size Other (in mb)	1.52	0.25	2.11	12.20
Number of Observations	58118	29895	26120	2103

Panel B: Correlation Matrix for Readability Measures

	Plain English Factor	*Bog Index*	*Log (# of Words)*	*Fog Index*	*Log(Total File Size)*	*Log(Text Size)*	*Log(Other Size)*
Plain English Factor	1	**0.8125**	**0.6293**	**0.4510**	**0.4782**	**0.5810**	**0.4250**
Bog Index	**0.8209**	1	**0.4978**	**0.4741**	**0.3679**	**0.4580**	**0.3264**
Log (# of Words)	**0.6371**	**0.5084**	1	**0.2399**	**0.7681**	**0.9580**	**0.6787**
Fog Index	**0.4639**	**0.4861**	**0.2515**	1	**0.2031**	**0.2135**	**0.1889**
Log(Total File Size)	**0.4597**	**0.3557**	**0.7437**	**0.2130**	1	**0.7074**	**0.9824**
Log(File Size 10-K Text)	**0.5433**	**0.4340**	**0.9055**	**0.2208**	**0.6666**	1	**0.6048**
Log(File Size Other)	**0.2190**	**0.1751**	**0.3966**	**0.1362**	**0.6821**	**0.4206**	1

Table 7
An Analysis of Plain English Measures around SEC Plain English Regulation

This table reports regression results where the dependent variable is report readability measured as the Bog Index, The Plain English Factor, or the log of total file size. All regressions include an intercept, control variables, calendar year dummies, and Fama and French (1997) 48-industry dummies. Standard errors are clustered by year and industry.The t-statistics are reported in parentheses below the coefficient. Panel A provides results for the actual tests related to the plain English Regulation. The sample in Panel A consists of 3,577 firm-year observations representing 1,285 unique firms that issued debt or equity in 1999 or 2000. 1,522 (2,055) of the firm-year observations occured in the pre (post) regulation period, where 1,194 (1,756) of those firm-year observations filed a prospectus (*Prospectus_Filer*). *Post_PE_Regulation* is a binary variable equal to one if a firm-year is after the adoption of the Plain English regulation (i.e., 1999, 2000) and zero otherwise. Panel B provides results for the falsification (placebo) tests. The sample in Panel B consists of 4,117 firm-year observations representing 1,772 unique firms that issued debt or equity in 1996 or 1997. 1,105 (3,012) of the firm-year observations occured in the pre (post) regulation period, where 894 (2,503) of those firm-year observations filed a prospectus (*Prospectus_Filer_False*). *Post_PE_Regulation_False* is a binary variable equal to one if a firm-year is after the fictitious adoption date (i.e., 1994 or 1995) and zero otherwise. Detailed variable definitions for all other variables are provided in Appendix A.

Panel A - Plain English Regulation Shock

	(1) *Bog Index*	(2) *Plain English Factor*	(3) *Log(Total File Size)*
Readability Measures:			
Prospectus_Filer	1.551	1.287	0.010
	(2.24)	(2.05)	(0.24)
Post_PE_Regulation	0.835	1.631	0.284
	(1.04)	(2.09)	(4.55)
*Prospectus_Filer*Post_PE_Regulation*	-2.020	-1.718	-0.053
	(-3.71)	(-3.22)	(-1.16)
Controls	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
# of OBS	3577	3577	3577
R^2	0.332	0.241	0.215

Panel B - Plain English Regulation Falsification Test

	(1) *Bog Index*	(2) *Plain English Factor*	(3) *Log(Total File Size)*
Readability Measures:			
Prospectus_Filer_False	-0.126	0.154	0.051
	(-0.17)	(0.23)	(1.02)
Post_PE_Regulation_False	1.281	0.602	-0.007
	(1.50)	(0.73)	(-0.10)
*Prospectus_Filer_False*Post_PE_Regulation_False*	1.204	1.413	0.019
	(1.99)	(2.46)	(0.41)
Controls	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
# of OBS	4117	4117	4117
R^2	0.387	0.323	0.156

Table 8

An Analysis of Plain English Measures Using Post-Filing Date Market Model Root Mean Squared Error (RMSE) as the Dependent Variable

This table reports regression results where the market model root mean squared error for trading days [6, 28] relative to the 10-K filing date is the dependent variable. All regressions include an intercept, calendar year dummies, and Fama and French (1997) 48-industry dummies. Standard errors are clustered by year and industry. The t-statistics are reported in parentheses below the coefficient. All regressions include 58,118 firm-year observations during 1994 to 2012. Detailed variable definitions are provided in Appendix A.

	(1)	(2)
	RMSE	RMSE
Readability Measures:		
Bog Index	0.006	
	(3.37)	
Plain English Factor		0.003
		(2.33)
Control Variables:		
Pre-filing alpha	-0.530	-0.534
	(-2.45)	(-2.46)
Pre-filing RMSE	0.586	0.587
	(26.02)	(26.20)
Abs(abnormal return)	4.275	4.275
	(15.62)	(15.69)
Log(market capitalization)	-0.087	-0.085
	(-6.98)	(-6.74)
Log (book-to-market)	-0.150	-0.148
	(-5.20)	(-5.13)
NASDAQ dummy	0.209	0.211
	(4.16)	(4.17)
Business segment index	-0.018	-0.022
	(-0.35)	(-0.43)
Market Volatility Index	0.032	0.032
	(3.02)	(3.03)
Return on Assests	-0.314	-0.314
	(-10.52)	(-10.86)
% Non-Missing Compustat Items	-1.476	-1.587
	(-1.88)	(-1.98)
M&A Activity	0.057	0.060
	(2.37)	(2.55)
Special Items	0.030	0.032
	(2.49)	(2.63)
Discontinued Operations	0.010	0.013
	(0.59)	(0.71)
Industry Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
# of OBS	58118	58118
R^2	0.5669	0.5667

Table 9
An Analysis of Readability Measures Using Analyst Measures as the Dependent Variable

This table reports regression results for analysts measures. The dependent variable for the regressions in the first three columns is abs(SUE). The dependent variable for the regressions in the last three columns is Analyst dispersion.All regressions include an intercept, calendar year dummies, and Fama and French (1997) 48-industry dummies. Standard errors are clustered by year and industry. The t-statistics are reported in parentheses below the coefficient. Regression where the dependent variable is abs(SUE) include 43,238 firm-year observations during 1994 to 2012. Regression where the dependent variable is Dispersion include 35,967 firm-year observations during 1994 to 2012. Detailed variable definitions are provided in Appendix A.

	(1)	(2)	(3)	(4)	(5)	(6)
	abs(SUE)	abs(SUE)	abs(SUE)	Dispersion	Dispersion	Dispersion
Readability Measures:						
Bog Index		0.002			0.001	
		(1.59)			(1.91)	
Plain English Factor			0.003			0.002
			(2.14)			(2.65)
Control Variables:						
Pre-filing alpha	-0.598	-0.595	-0.593	-0.355	-0.354	-0.353
	(-9.18)	(-9.16)	(-9.17)	(-7.69)	(-7.68)	(-7.69)
Pre-filing RMSE	0.153	0.151	0.150	0.085	0.084	0.084
	(5.41)	(5.30)	(5.28)	(5.16)	(5.07)	(5.07)
Abs(abnormal return)	1.285	1.280	1.273	0.682	0.678	0.673
	(5.75)	(5.72)	(5.70)	(4.50)	(4.45)	(4.43)
Log(market capitalization)	-0.093	-0.094	-0.093	-0.033	-0.034	-0.033
	(-5.98)	(-6.03)	(-6.02)	(-3.80)	(-3.86)	(-3.86)
Log (book-to-market)	0.124	0.123	0.123	0.066	0.066	0.066
	(5.23)	(5.19)	(5.21)	(5.51)	(5.46)	(5.48)
NASDAQ dummy	-0.124	-0.125	-0.124	-0.067	-0.068	-0.067
	(-6.78)	(-6.96)	(-6.83)	(-4.17)	(-4.26)	(-4.18)
Business segment index	-0.042	-0.040	-0.043	-0.019	-0.018	-0.019
	(-0.93)	(-0.88)	(-0.95)	(-1.02)	(-0.94)	(-1.03)
Market Volatility Index	0.006	0.006	0.006	0.003	0.003	0.003
	(2.50)	(2.47)	(2.50)	(1.38)	(1.37)	(1.38)
Return on Assests	-0.113	-0.112	-0.111	-0.042	-0.042	-0.041
	(-2.09)	(-2.15)	(-2.15)	(-1.10)	(-1.13)	(-1.13)
% Non-Missing Compustat Items	-0.110	-0.058	-0.084	0.030	0.066	0.050
	(-0.19)	(-0.10)	(-0.15)	(0.14)	(0.30)	(0.22)
M&A Activity	-0.069	-0.072	-0.074	-0.014	-0.015	-0.017
	(-4.95)	(-4.99)	(-5.06)	(-1.56)	(-1.83)	(-1.97)
Special Items	0.093	0.091	0.088	0.034	0.033	0.031
	(6.64)	(6.73)	(6.69)	(4.61)	(4.30)	(4.09)
Discontinued Operations	0.076	0.073	0.071	0.015	0.013	0.012
	(3.43)	(3.31)	(3.29)	(1.42)	(1.23)	(1.11)
# of Analysts	-0.001	-0.001	-0.001	0.004	0.004	0.004
	(-0.26)	(-0.24)	(-0.25)	(2.13)	(2.17)	(2.17)
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
# of OBS	43238	43238	43238	35967	35967	35967
R^2	0.1829	0.1831	0.1833	0.2122	0.2127	0.2132