Journal of Accounting and Economics 63 (2017) 329–357

Contents lists available at ScienceDirect

Journal of Accounting and Economics

journal homepage: www.elsevier.com/locate/jae

A plain English measure of financial reporting readability[☆]

Samuel B. Bonsall IV[a,b], Andrew J. Leone[c,*], Brian P. Miller[d], Kristina Rennekamp[e]

[a] The Ohio State University, United States
[b] The Pennsylvania State University, United States
[c] University of Miami, United States
[d] Indiana University, United States
[e] Cornell University, United States

ARTICLE INFO

Article history:
Received 5 February 2015
Received in revised form
8 March 2017
Accepted 21 March 2017
Available online 5 April 2017

Keywords:
Plain English reporting
Textual analysis
Regulation
Mandatory disclosure

ABSTRACT

We propose a new measure of readability, the Bog Index, which captures the plain English attributes of disclosure (e.g., active voice, fewer hidden verbs, etc.). We validate this measure using a series of controlled experiments and an archival-based regulatory intervention to prospectus filing readability. We also demonstrate the importance of understanding the underlying drivers of quantity-based measures of readability. In particular, we caution researchers that a vast amount of the variation in Form 10-K file size over time is driven by the inclusion of content unrelated to the underlying text in the 10-K (e.g., HTML, XML, PDFs).

1. Introduction

A growing body of research in accounting and finance examines whether, and to what extent, qualitative attributes of corporate communication (e.g., tone, readability) affect the decision-making of investors and information intermediaries. Although relatively clear guidance exists for measuring linguistic attributes of financial disclosure such as tone (e.g., Loughran and McDonald, 2011; Henry and Leone, 2016), the most appropriate measure of readability is less clear. Researchers have primarily selected from a limited set of existing readability measures that are based on either writing clarity (e.g., the Gunning Fog Index) or disclosure quantity (e.g., file size of the filing). We extend this literature by introducing a new measure of readability, the Bog Index, which is designed to capture the plain English attributes of disclosure. We then use controlled experiments, a regulatory intervention, and archival-based capital market tests to validate and compare this new measure with existing readability measures.[1]

☆ We are grateful for discussions and helpful comments from Rob Bloomfield, Kimball Chapman, Khrystyna Bochkay, Roman Chychyla, Wayne Guay (editor), Patrick Hopkins, Michelle Hutchens, Robert Libby, Jeff McMullin, Sundaresh Ramnath, Lorien Stice-Lawrence, K.R. Subramanyam, Eric Weisbrod, an anonymous reviewer, and workshop participants at the 11th Annual Conference on Empirical Legal Studies, Baruch College, Chinese University of Hong Kong, 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, Mike Durney, Adam Hadley, Eric Holzman, Sharadh Ramesh, and Barrett Wheeler for research assistance. Brian Miller gratefully acknowledges financial support from the PwC Faculty Fellowship. The parsing code and Bog Index data are available for download on Brian Miller's website at https://kelley.iu.edu/bpm/index.html.

* Corresponding author.
E-mail address: a.leone@miami.edu (A.J. Leone).

[1] The Bog Index data are available for download on Brian Miller's website at https://kelley.iu.edu/bpm/index.html.

http://dx.doi.org/10.1016/j.jacceco.2017.03.002
0165-4101/

Although the exact definition varies, there is general agreement that "readability" refers to the ease with which a reader can process and comprehend written text. In terms of financial disclosure readability, the Securities and Exchange Commission (SEC) provides some very specific guidance in recommending that managers employ plain English attributes by avoiding writing constructs like passive voice, weak or hidden verbs, superfluous words, legal and financial jargon, numerous defined terms, abstract words, unnecessary details, lengthy sentences, and unreadable design and layout in their financial disclosures (SEC, 1998b). Lower frequencies or the absence of these constructs are consistent with clear writing that language experts have asserted is critical to users' understanding of written materials in a variety of contexts including the military, healthcare, and law (DuBay, 2004). Despite the strong support for plain English from linguists and regulators, few studies employ plain English measures as proxies for readability.

We begin by evaluating how well alternative measures of readability capture the plain English attributes recommended by linguistic experts and the SEC. Most readability studies use the Gunning (1952) Fog Index as a proxy for readability when testing the causes and consequences of financial reporting readability.[2] Given this measure is based on average sentence length and the proportion of words with three or more syllables, the measure does in theory capture two broad plain English attributes. However, recent research by Loughran and McDonald (2014a) raises valid criticisms that the complex word component of the Fog Index treats all words with three or more syllables as "complex" even though the meaning of many of these multisyllabic words (e.g., Company) would be well understood by even the least sophisticated investors. As such, only the sentence length portion of the Fog Index appears to map directly into the SEC's plain English guidelines.

In addition to measuring the writing clarity component of readability using the Fog Index, many other studies rely on the quantity of textual disclosure to measure readability. Although historically these quantity-based measures have focused on the number of words contained in a financial filing, more recently Loughran and McDonald (2014a, p. 1644) advocate the use of "file size of the 10-K as an easily calculated proxy for document readability." It is important to note that both of these quantity-based measures are limited as measures of plain English readability because they only capture a single plain English attribute: superfluous words. Further, the SEC cautions that quantity of disclosure measures may have some shortcomings in capturing writing clarity. Specifically, the SEC notes that there may be a trade-off between writing clarity and the quantity of disclosure: "writing a disclosure in plain English can sometimes increase the length of particular sections…" (SEC, 1998a). In summary, both the Fog Index and quantity-based measures of readability capture some attributes of plain English disclosures, but they are not comprehensive.

Given the limitations of these measures in capturing plain English attributes, we propose an alternative multi-faceted measure of disclosure clarity, the Bog Index, which is based on plain English writing principles and captures the spirit of almost all of the SEC's guidelines regarding clear communication with investors. The Bog Index is derived from a commercial software program, *StyleWriter*, which captures attributes specifically mentioned in the SEC Plain English Handbook including sentence length, passive voice, weak verbs, overused words, complex words, and jargon (SEC, 1998b). Among these many features of the Bog Index, one unique aspect of the measure stems from the way in which word complexity is determined. Rather than assuming all multi-syllabic words are complex, as is done in computing the Fog Index, word complexity is instead determined by word familiarity based on a proprietary list of over 200,000 words. Thus, the Bog Index measure of writing clarity overcomes the major criticism of the Fog Index related to capturing word complexity based on syllable counts alone.

Prior research has employed a subset of the plain English attributes used in the Bog Index. For instance, Miller (2010) uses output from earlier and much more limited versions of the same software used to generate the Bog Index to create a somewhat ad hoc measure of plain English reporting.[3] More recently, Loughran and McDonald (2014b) creates a measure of plain English writing (*LM PE Index*), based on a small subset of the plain English attributes highlighted by the SEC. As discussed in more detail in subsequent sections, compared to these earlier measures, the Bog Index provides a much more comprehensive set of factors and is calculated using a pre-programmed algorithm that eliminates researcher discretion related to how to calculate the measure.

One of the challenges that prior researchers have faced in validating readability measures is that archival-based capital market tests, which are used for proxy validation, are joint tests that the proxy captures the construct of interest and that the construct is related to the outcome. For instance, in our setting it is possible that the Bog Index is a valid proxy for readability, but that readability does not affect a capital market outcome such as stock return volatility. As such, we use several complementary approaches to validate the Bog Index as a proxy for readability including an experiment, a quasi-experiment, and an archival examination of a readability-related regulatory intervention.

The experiment and the quasi-experiment enable us to examine the effects of the Bog Index on surrogates for investors while holding constant other potentially influential variables (Libby et al., 2002). The evidence shows that participants who receive a more readable disclosure, as measured by the Bog Index, rate the disclosure to be significantly easier to read than participants who receive a less readable disclosure. This suggests that the Bog Index captures financial statement readers' internal evaluations of readability.

[2] Numerous studies have examined the impact of financial reporting complexity on retail investors (Lawrence, 2013; Rennekamp, 2012; Miller, 2010), sell-side equity analysts (Bozanic and Thevenot, 2015; Lehavy et al., 2011), rating agencies (Bonsall and Miller, 2017), and the voluntary disclosure behavior of firms (Guay et al., 2016).

[3] This pre-cursor plain English measure of readability created from components of the earlier versions of *StyleWriter* is far less comprehensive than the components calculated by the more recent software package. As such, in this study, we focus on the more comprehensive standardized Bog Index measure as it subsumes this prior measure.

Our regulatory intervention analysis allows us to observe how the Box Index changes around the 1998s *Plain English Mandate*, which required registrants to prepare their prospectuses with plain English writing. We find that readability improved for all three clarity-of-writing measures examined: the Bog Index, the Fog Index, and the LM PE Index, but find no such evidence for the quantity-based readability measures of total file size or total words. Further, we document that around the mandate the Bog Index captured a significantly greater improvement in readability relative to the Fog Index and the LM PE Index. The Bog Index changed by approximately 1.22 standard deviations, which is more than double the change in the standard deviation of the Fog Index and about ten percent greater than the corresponding change in the LM PE Index.

We also implement a series of placebo tests using pseudo regulation dates both before and after the actual regulatory intervention. Our results show that while the improvement in the Bog Index and Fog Index was unique to the regulation period, the LM PE Index also improves after the regulation. This suggests that the LM PE Index response was not solely due to the regulatory intervention. Overall, the evidence from our examination of a readability-related regulatory intervention suggests that among the writing clarity measures of readability, the Bog Index best captures the plain English attributes mandated by the SEC.

Following these validation tests, we turn our attention to examining whether the readability measures are associated with the capital market outcomes examined in Loughran and McDonald (2014a) (i.e., stock market volatility and equity analysts' earnings forecast properties) after including firm fixed effects.[4] We find that with the exception of the LM PE Index, all measures of readability are positively associated with post-10-K filing stock return volatility (i.e., less readable filings are associated with greater future volatility). Although the regression coefficients on most of the readability measures are statistically significant, it is important to point out that the Bog Index has the largest association with future stock market volatility of any of the readability measures examined. In particular, the magnitude of the association between the Bog Index and future stock return volatility is nearly twenty-five percent greater than those exhibited by the other measures of readability.

We next examine whether the various measures of readability are associated with equity analysts' earnings forecast properties, which should provide insights into whether readability affects sophisticated market intermediaries. We find no evidence of an association between analyst dispersion and any of the readability measures when we include firm fixed effects. This suggests that time-invariant firm characteristics are likely driving any association between readability and analyst dispersion documented in prior research. With respect to analyst accuracy, we only find evidence of an association between the quantity-based measures of total file size and total words and analyst accuracy. Although we cannot completely rule out an effect of readability on equity analysts based on the insignificant statistical results obtained in our tests, these results are consistent with quantity of information, but not readability, having an influence on more sophisticated market participants.

In our final set of tests, we show that breaking down the quantity-based measures can help disentangle what annual report content is driving the associations documented in our analyses of total file size. For instance, we show that the file size related to the text of Form 10-K has no association with stock return volatility, but instead the file size related to additional exhibits and more tabulated information drives the association. Further, in contrast to the previously discussed insignificant association between total file size and analyst dispersion, we show that after separating out the non-textual components, the file size components related to the actual text of Form 10-K and exhibits are significantly associated with analyst dispersion. These two cases demonstrate that the use of total file size as a proxy for the readability of the text of the 10-K filing may lead to over/under rejection depending on the influence of the non-textual components. Finally, we show that total file size increases dramatically over our sample period and that most of the variation in that measure 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). As such, we provide a note of caution related to comparing total file size over time.

The remainder of the paper proceeds as follows. Section 2 reviews the academic accounting literature on readability. Section 3 discusses the different readability measures used in prior academic studies. Section 4 presents the validation of the Bog Index based on evidence from experiments and archival tests around the *Plain English Mandate*. Section 5 tests whether plain English readability affects capital market participants. Section 6 concludes.

2. Background

2.1. Academic accounting readability literature

Research on the readability of financial reports has increased dramatically over the past decade, coinciding with the call by Core (2001) for greater use of techniques from computational linguistics to capture large-sample measures of disclosure quality. This burgeoning literature primarily focuses 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 be less completely revealed in market prices. Consistent with these predictions, Miller (2010) finds that

[4] As discussed in more detail later in the paper, in untabulated results we replicate and find consistent results to those documented in Loughran and McDonald (2014a) when using industry effects. However, when we replace the industry fixed effects with firm fixed effects we find that certain outcome variables (e.g., analyst dispersion) are no longer significant. This suggests that in these instances time-invariant firm characteristics are likely influencing the previously documented associations with readability.

firms with more readable financial reports have more pronounced small investor trading around the 10-K filing date, while De Franco et al. (2015) shows 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) documents 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 equity analysts (Lehavy et al., 2011; Bozanic and Thevenot, 2015) and bond rating agencies (Bonsall and Miller, 2017). Lang and Stice-Lawrence (2015) performs a large-sample empirical analysis of annual report textual disclosures for non-U.S. companies and shows that firms with improvements in financial reporting experience improvements in several economic outcomes. Relatedly, Guay et al. (2016) documents that firms use voluntary disclosure to mitigate the negative effects associated with less readable filings.

Other studies test whether financial reporting readability is related to profitability and future investment. 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. In addition, Biddle et al. (2009) uses readability as a measure of financial reporting quality and shows that more readable financial reports are associated with lower over- and under-investment.

2.2. Readability, plain English, and the SEC

Research over the past century has documented the importance of readability in contexts as wide-ranging as the military, healthcare, and the law (DuBay, 2004). More recently, experts in multiple fields have focused on specific plain English attributes (e.g., active voice, shorter sentences, avoiding jargon) that reflect fundamentally sound habits for clear writing. Whiteman (2000) discusses how the scientific community increasingly emphasizes plain English communication with the non-scientific public, as well as between scientists across disciplines, in order to improve the understanding of scientific findings. In the legal profession, the international organization Clarity has members in over 26 countries who advocate for the use of plain English language in place of unintelligible legalese. In healthcare, the U.S. Department of Health and Human Services (HHS) promotes the use of plain English for improving patients' health literacy and outcomes and provides guidelines and resources for achieving those goals.

Many definitions of plain English writing exist, but language researchers generally 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).

Regulators and professional accounting organizations have also embraced the importance of incorporating plain English attributes into financial disclosures. To combat concerns over unreadable prospectus filings the SEC adopted the 1998 *Plain English Mandate*, SEC Rule 421(d). Along with this regulation the SEC provided a companion handbook entitled "*A Plain English Handbook: How to create clear SEC disclosure documents*" (SEC, 1998b). The Handbook encourages plain English disclosures in all investor communications and provides several recommendations to mitigate common problems encountered with disclosure documents. More recently in 2013, the AICPA launched the Center for Plain English Accounting, with the goal of helping auditors and practitioners to more clearly explain the technical accounting literature to their clients by providing plain English interpretations (Averett, 2013).

The SEC's Plain English Handbook (1998b) provides very clear guidance on plain English writing and issues to avoid when communicating in regulatory filings. In the handbook, the SEC lists several distinct problems that they encountered in their review of regulatory filings: (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.[5] Fortunately, computational linguistics has progressed to a point where it is possible from a technological standpoint to capture these plain English attributes. Despite the strong support for plain English by linguistic experts and regulators, few studies employ plain English measures as proxies for readability.

3. Readability measures

In this section, we provide an overview and definitions of the readability measures used in the existing accounting and finance literatures and introduce the Bog Index measure of plain English readability. We provide all related variable definitions in Appendix B, while we describe the specific procedures we use to parse SEC filings and calculate the readability measures in Appendix C.

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

3.1. Fog Index

Most prior research in accounting and finance relies on the Fog Index, developed by Gunning (1952), as a primary measure of financial reporting complexity or readability. 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 (i.e., words of three or more syllables). The Fog Index formula adds these components together and multiplies the resulting sum by a scalar to estimate a reading grade level for which higher values indicate less readable text. Formally, we calculate the *Fog Index* as follows:

$$Fog\ Index = 0.4(average\ number\ of\ words\ per\ sentence + percentage\ of\ complex\ words) \tag{1}$$

Although the Fog Index provides a straightforward measure that captures two of the plain English attributes recommended by linguistic experts and regulators, Loughran and McDonald (2014a) identifies a significant shortcoming of the measure in the context of financial reporting. Specifically, Loughran and McDonald (2014a) points out that business texts have an extremely high percentage of words that are denoted ‘complex’ by the Fog Index because they contain three or more syllables but are, in fact, well understood by the vast majority of investors and analysts (e.g., “depreciation,” “liability”). Loughran and McDonald (2014a) suggests that the complex words component of the Fog Index merely adds measurement error since the study fails to find any evidence that multisyllabic words explain post-filing stock return volatility, unexpected earnings, and analyst forecast dispersion.

3.2. Words and file size measures of readability

While most prior studies rely on the Fog Index to capture the writing clarity component of readability, other studies use measures of writing attributes such as document length (i.e., number of words in filing) to capture the quantity of disclosure component of readability (You and Zhang, 2009). Formally, we define *Total Words* as the number of words contained in the entire 10-K filing. More recently, Loughran and McDonald (2014a) recommends an alternative disclosure quantity measure based on file size of the 10-K. We calculate *Total File Size* as the number of megabytes used by the entire 10-K filing as recorded on the EDGAR filing system.

Both of these quantity-based readability measures are based on the notion of overwriting, where documents are written in a manner that is too detailed and long for readers to easily process. As such, they are consistent with a limited subset of the plain English guidance (e.g., superfluous words). However, the potential downside of these quantity-based measures is that they could capture constructs other than readability. For instance, measuring the quantity of disclosure in a 10-K setting inadvertently includes separate exhibits that are unrelated to the annual 10-K filing requirements (e.g., compensation contracts, supplier/customer agreements, or bond indentures) but are attached to the filing. To the extent these exhibits are related to changes in a firm's underlying business model, it is difficult to disentangle whether the increase in overall quantity of disclosure is driven by the clarity of writing or fundamental uncertainty surrounding the economic changes taking place in the firm. In addition, with respect to the file size measure of readability, other factors such as the introduction of HTML, XML, PDF, and picture format file attachments can lead to substantial variation in 10-K file size over time but are unrelated to the underlying text in the 10-K that relates to the economic performance of a firm over the prior fiscal year. As such, these quantity-based measures of disclosure are particularly susceptible to concerns that the content of the disclosure, as opposed to readability, explains statistical associations between quantity-based measures of readability and capital market outcomes.

3.3. Bog Index

Given the fact that prior measures are somewhat limited in capturing plain English attributes, we propose an alternative measure of readability, the Bog Index, which is designed to capture a broader set of plain English attributes. Specifically, we take advantage of recent software developments made possible by a computational linguistics software program, *StyleWriter—The Plain English Editor*, which computes a multifaceted measure of readability, the Bog Index. The Bog Index captures the plain English writing attributes recommended by linguistic experts and highlighted in the SEC's Plain English Handbook (1998b).[6] The Bog Index, which summarizes those writing attributes that tend to *bog readers down*, is computed as the sum of three multifaceted components:

$$Bog\ Index = Sentence\ Bog + Word\ Bog - Pep, \tag{2}$$

where a higher Bog Index equates to a less readable document.

The first component, *Sentence Bog,* identifies readability issues stemming from sentence length, where longer sentences lead to a higher Bog Index. Specifically, the software identifies the average sentence length across the entire document. That average sentence length is then squared and scaled by a standard long sentence limit of 35 words per sentence.

The second component, *Word Bog,* is comprised of two main subcomponents: (1) plain English style problems and

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

(2) word difficulty. Specifically, *Word Bog* is calculated as the sum of plain English style problems and word difficulty multiplied by 250 and divided by the number of words. For perspective, the plain English style problems component of *Word Bog* is a combination of issues highlighted in the SEC's Plain English Handbook (1998b): passive verbs, hidden verbs, overwriting, legal terms, clichés, abstract words, and wordy phrases.

The second component of *Word Bog* is calculated based on the word difficulty for general vocabulary (i.e., heavy words), abbreviations, and specialist terms.[7] In contrast to the Fog Index, which calculates word difficulty based on syllable counts, the Bog Index measures word difficulty using a proprietary 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).

For illustration purposes, Panel A of Appendix D provides scores for the top 100 multisyllabic words included in 10-K filings from 1994–2011.[8] This list of words accounts for over 50 percent of the multisyllabic words appearing in 10-K filings over the sample period and, as noted by Loughran and McDonald (2014a), most investors are likely to understand many of these common multisyllabic words. Consistent with this notion, we show that although the Fog Index penalizes these words for having multiple syllables, the penalties assessed by the Bog Index are far less severe. Specifically, over eighty percent of these words receive a penalty of no greater than one point, and none of these words receives a penalty greater than two points out of a possible four points. For comparison, in Panel B of Appendix D, we also obtain a random sample of 100 of the most abstract words used in 10-K filings that received the full penalty of four points. We contend that most readers would likely judge these words as being more complex than the words receiving lower penalties in Panel A of Appendix D. Overall, this evidence supports the assertion that the Bog Index captures word complexity. Thus, the criticism about the Fog Index over-penalizing multisyllabic words appears to be significantly mitigated when using the Bog Index.

The final component of the Bog Index, *Pep,* identifies writing attributes that facilitate understanding of texts by readers. Specifically, this component of the Bog Index accounts for good writing by summing the usage of items such as names and interesting words, which tend to make writing more interesting. *Pep* is calculated as the sum of these components multiplied by 25 (one tenth of the effect from *Word Bog*) and scaled by the number of words in the document plus sentence variety (i.e., the standard deviation of sentence length multiplied by ten and scaled by the average sentence length).

To more clearly see the contribution of different writing issues to the calculation of the Bog Index, we follow a similar approach to Bonsall and Miller (2017) and regress the values of the Bog Index on standardized (mean=0, variance=1) transformations of *StyleWriter*'s primitive outputs. We present the results from the estimation of this regression in Appendix E and group the outputs by the higher-level *StyleWriter* categories of *Sentence Bog*, *Word Bog*, and *Pep*. These standardized outputs explain 92 percent of the variation in the Bog Index and highlight that, as predicted, the software penalizes plain English style violations (e.g., passive voice), difficult general words (e.g., complex words, specialist terms), and rewards conversational expressions.[9] Further, the results show that deviations in these primitive factors are meaningful. For instance, in our sample a one standard deviation in average sentence length results in a three-point higher value of the Bog Index.

It is worth noting that other studies have also relied on prior versions of the *StyleWriter* software to create plain English measures of readability (Miller 2010; Rennekamp 2012). These pre-cursor constructs created by the earlier versions of *StyleWriter* were far less comprehensive than the components calculated by the more recent software package. Further, because the summary Bog Index measure was unavailable these researchers relied on more *ad hoc* frequency-based summary measures of plain English violations. In this study, we focus on the more comprehensive standardized Bog Index measure as it subsumes these prior measures.

3.4. Loughran and McDonald plain English index

Loughran and McDonald (2014b) provides an alternative measure of readability based on some plain English attributes. They use this measure to document that firms complied with the 1998 s *Plain English Mandate* by improving the readability of their prospectuses and Form 10-K filings. We re-create the Loughran and McDonald (2014b) measure by parsing each filing in our sample to identify the six different components used by that study to generate a comparable plain English measure: average sentence length, average word length, the ratio of passive verbs to total words, the ratio of legal terms to total words using legal words from Loughran and McDonald (2011), the ratio of personal pronouns to total words, and the ratio of other plain English violations (e.g., negative phrases, superfluous words) to total words. We then standardize each of these components to have a mean of zero and standard deviation of one in our sample. Finally, to create the summary *LM PE Index*, we sum sentence length, word length, passive voice, legalese, and other items, and subtract personal pronouns. This

[7] Some examples of heavy words include 'ultimo,' 'pari-mutuel,' 'proximal,' 'postprandial,' and 'perpetual.' Some examples of abbreviations include AHP (Affordable Housing Program), GSE (Government Sponsored Entities), LTIP (Long Term Incentive Plan), and ESOP (Employee Stock Ownership Plan). Some examples of specialist terms include 'ontology,' 'continuant,' 'filial,' 'sublingual,' and 'syntactical.'

[8] It is worth noting that our word list differs slightly from that used in Loughran and McDonald (2014b). These differences are driven by the fact that we remain consistent with the Fog Index and use only stem words with three or more syllables (after excluding suffixes such as –ing, -ed, etc.). In contrast, their list contains words with suffixes.

[9] Appendix E is only as an illustration of the components of the Bog Index. Specifically, the primitive components are often transformed and do not receive equal weights in the actual calculation of the Bog Index, which clarifies why they do not explain 100% of the variation in the measure.

approach creates a measure that is comparable with the other measures of readability, where higher scores are associated with lower readability. It is worth pointing out that this transformation is the equivalent of multiplying the Loughran and McDonald (2014b) measure by negative one.[10]

While the LM PE Index appears to capture many of the attributes of plain English writing highlighted by the SEC, it has several shortcomings relative to the more comprehensive Bog Index. First, it incorporates word complexity based on average word length, which is similar to syllable count, the component of the Fog Index that Loughran and McDonald (2014a) directly criticize. Second, the LM PE Index uses a limited set of twelve phrases to capture technical jargon usage. Similarly, the measure uses limited lists of legal jargon, negative phrases, and superfluous words. In contrast, the Bog Index scores word complexity based on frequency of contemporary usage and identifies other style problems, such as jargon and superfluous words based on a corpus of over 200,000 words.

Consistent with these differences we find (untabulated) that across the 10-K filings in our sample, *LM PE Index* is only about five percent correlated with *Bog Index* and eighteen percent correlated with *Fog Index*. Further, *LM PE Index* is significantly negatively correlated with both Log(*Total File Size*) and Log(*Total Words*) (−.241 and −.131, respectively). In contrast *Bog Index* is significantly positively correlated with all the other measures of readability, such as, *Fog Index*, Log(*Total File Size*), and Log(*Total Words*) (.460, .336, and .340, respectively). Combined, these correlations suggest that LM PE Index captures something fundamentally different from the Bog Index and other measures of readability.

4. Results

4.1. Validation of the Bog Index

One of the challenges that prior researchers have faced in validating readability measures is that archival-based capital market tests, which are used for proxy validation, are joint tests that the proxy captures the construct of interest and that the construct is related to the outcome. For instance, in our setting it is possible that the Bog Index is a valid proxy for readability, but that readability does not affect a capital market outcome such as stock return volatility. Alternatively, it is also possible to falsely conclude that the Bog Index is a valid proxy for readability if it is correlated with another unobservable variable that is correlated with the outcome variable examined (e.g., complexity). Thus, we begin our validation of the Bog Index as a proxy for readability by conducting an experiment, a quasi-experiment, and an archival examination of a readability-related regulatory intervention. The experiment and the quasi-experiment enable us to examine the effects of the Bog Index on surrogates for investors while holding constant other potentially influential variables (Libby et al., 2002), while the regulatory intervention allows us to observe how the Bog Index changes around an SEC requirement to prepare prospectus documents using plain English principles.

4.1.1. Experimental validation of the Bog Index

We begin by providing experimental evidence on the validity of the Bog Index as a measure of readability. Specifically, we implement a 1×2 between-subjects experiment, in which we manipulate whether a firm disclosure is more or less readable, based on the Bog Index. Participants are randomly assigned to receive only one of the two disclosures. To isolate the impact of the Bog Index from other measures of readability, we hold constant across both disclosures (1) length, in words (244 words each), (2) file size, (3) formatting, and (4) the *Fog Index* score.[11] We manipulate the disclosures so that the more readable disclosure has a *Bog Index* score of 50 (Fair), and the less readable disclosure has a *Bog Index* score of 86 (Poor). Since the *LM PE Index* is also based on plain English attributes, not surprisingly we find that the more readable disclosure is also associated with a lower *LM PE Index* score.

The more and less readable disclosures are presented in Appendix F. This appendix also highlights some of the linguistic attributes that contribute to a worse Bog Index score (e.g., abstract words, jargon phrases, wordy phrases, passive voice, long sentences). After reading the disclosure, participants are asked, "How difficult or easy is it to read the information in the disclosure above?" with responses ranging from 1 = very difficult to 9 = very easy. Participants conclude by answering debriefing and demographic questions.

In total, we recruited 102 participants from Amazon's Mechanical Turk (AMT) platform. AMT has been used in prior accounting research (e.g., Rennekamp, 2012; Koonce et al., 2015), has been shown to replicate results found in existing accounting studies (Krische, 2015), and has also been shown to be more representative of the U.S. population than alternative student pools from which we could have recruited our participants (Paolacci et al., 2010; Horton et al., 2011). Prior

[10] Loughran and McDonald (2014b) make their plain English measure publicly available on Bill McDonald's website http://www3.nd.edu/~mcdonald/Data/Plain%20English_LoughranMcDonald.txt. However, their posted sample data ends in 2007 and only covers 10-K filings (not IPO filings). As such, the plain English measures used in our paper are replications of their measure of plain English readability. In untabulated tests, we find that the correlation between our replication and their scores are around 75%. Further, we re-analyze (untabulated) all of our regressions and figures using their downloaded data and find virtually identical results to those reported.

[11] Despite our best effort to hold the *Fog Index* constant, the score varies slightly across the two disclosures. However, we designed our materials so that the more readable condition has a slightly *worse Fog Index* score (more readable = 15.83; less readable = 15.70). As such, the difference should bias against our finding a result on the *Bog Index*.

Table 1
Experimental Validation of Bog Index.
This table reports results from our experimental validation of the Bog Index. Specifically, we use a controlled 1×2 between-subjects experiment and manipulate whether participants receive a more or less readable disclosure, as measured by the Bog Index. Between conditions we hold constant the disclosures' (1) length, in words, (2) formatting, and (3) Fog Index to rule these out as alternative explanations for our results. See Appendix D for images of the more and less readable disclosures. Results presented below are for our direct measure of participants' perceptions of readability, where we ask participants, "How difficult or easy is it to read the information in the disclosure above?" with responses ranging from 1 = very difficult to 9 = very easy. The reported p-value is one-tailed, to be consistent with our directional prediction.

	More Readable Condition	*Less Readable Condition*	*t-stat*	*p-value*
Mean Readability Rating	6.50	5.67	2.28	0.012
Std. Deviation	(0.26)	(0.25)		
N	50	52		

studies have also shown that under the AMT platform participants often exert just as much (or more) effort (e.g., take the task as seriously) as other student participation pools commonly used (Paolacci et al., 2010). In sum, recruiting from AMT for our studies is consistent with the idea that experiments should use participants with sufficient knowledge and work ethic to complete the tasks at hand but avoid using more sophisticated subjects than is necessary to meet the goals of the research (Libby et al., 2002).

In terms of our participant pool we find that, on average, participants are 31.77 years old with 11.24 years of full-time work experience. 32.35 percent report having directly invested in stocks in the past and 26.47 percent report having invested indirectly (e.g., via mutual funds or other retirement accounts). These demographics are in line with the populations used in prior studies to test for effects on individual investors (Krische, 2015) and do not significantly differ across the experiments in our study.

The results support the notion that the Bog Index captures investors' assessments of financial text readability. Specifically, we find that participants who receive the lower Bog Index disclosure rate the disclosure to be significantly easier to read than participants who receive the higher Bog Index disclosure. Table 1 shows that participants rate, on average, disclosures in the lower *Bog Index* condition as 6.50 on the nine-point scale. In contrast, participants in the higher *Bog Index* condition rate their disclosure, on average, as a 5.67 on the same scale. Further, the ratings in the higher *Bog Index* condition are significantly lower than the lower *Bog Index* condition (t=2.28). In sum, the results from Table 1 provide support that the Bog Index captures aspects of readability that are incremental to other non-plain English measures of readability used in prior research such as document length, formatting, and the Fog Index.

4.1.2. Quasi-experimental validation of Bog Index

Our first experiment enables us to design a disclosure in which we can hold several other attributes of the disclosure constant. However, it is possible that these disclosures abstract from actual financial reports and, as such, may generate external validity concerns. To address issues of external validity, we provide a separate analysis in which we use a quasi-experiment to examine whether the Bog Indices for excerpts from actual firm disclosures correspond with investors' perceptions of disclosure readability. We start by hand-collecting excerpts of approximately 200 words from the beginning of the Letter to Shareholders for all 2014 Fortune 100 firms; there is slight variation in the exact length of each excerpt to prevent cutting off disclosures mid-sentence.

In contrast to our experimental evidence, in this quasi-experiment we rely on actual disclosures. As such, we do not have the ability to alter the disclosure to manipulate the plain English attributes. To ensure that readability differences are sufficiently large for participants to distinguish, we select the top three most readable (Wells Fargo, Kroger, and PepsiCo) and least readable (Phillips 66, International FC Stone, and United Technologies) excerpts, as captured by their *Bog Index* measures.[12] We then disguise each firm by replacing the actual company names with generic names like Gamma or Theta, which allows us to reduce the likelihood that participants' responses will be affected by their knowledge or perceptions of the actual firms.

We randomly assign each participant one of the lower *Bog Index* disclosures and one of the higher *Bog Index* disclosures, and then, as in our experimental study, ask participants to rate "How difficult or easy is it to read the information in the disclosure above?" with responses ranging from 1 = very difficult to 9 = very easy. Examples are provided in Appendix G for one of the lower *Bog Index* (i.e., Wells Fargo) and one of the higher *Bog Index* (i.e., Phillips 66) disclosures.

[12] These cut-offs were selected because there were substantial gaps in the distribution of the *Bog Index* measure around these excerpts. An added benefit of limiting our sample to the top and bottom three disclosures is that it allows us to mask the identity of the fourth most readable disclosure, where the actual disclosure included information that made the firm clearly-identifiable (i.e., a discussion of Amazon Prime). Masking the disclosure by editing the content would violate the spirit of the study of using actual firm disclosures (with only firm names changed), while electing not to edit the text would potentially confound the results as participants would be able to identify the firm and their responses might be driven by feelings about Amazon rather than disclosure readability.

Table 2
Quasi-Experiment Validation of Bog Index.

This table reports results from our quasi-experimental validation of the Bog Index. We hand-collect excerpts of approximately 200 words from the beginning of the Letter to Shareholders for all 2014 Fortune 100 firms. We then measure the Bog Index for each excerpt in our collected sample, and keep the three most-readable disclosures (Wells Fargo, Kroger, and PepsiCo) and three least-readable disclosures (Phillips 66, International FC Stone, and United Technologies) for our study. Appendix G provides a sample of these disclosures. Each participant is randomly assigned to read both one of the more readable disclosures, and one of the less readable disclosures and then rate, "How difficult or easy is it to read the information in the disclosure above?" with responses ranging from 1 = very difficult to 9 = very easy. Panel A presents the comparison of average ratings between the more and less readable excerpts, collapsed across the three disclosures. Panels B and C present the average ratings for each of the individual more and less readable excerpts, respectively. Panel D presents information on within-participant differences in ratings between the more and less readable disclosure that they viewed. The reported t-statistic (signed rank value) is associated with testing whether the mean (median) difference between ratings of the more vs. less readable disclosure is significantly greater than zero. Panel D also reports the proportion of within-participant differences in ratings that are non-negative (i.e., where the more readable disclosure is rated to be as or more readable than the less readable disclosure). Reported p-values are one-tailed, to be consistent with the directional predictions.

Panel A - Readability Ratings Collapsed Across the More Readable and Less Readable Excerpts

	More Readable Excerpts (Collapsed)	*Less Readable Excerpts (Collapsed)*	*t-stat*	*p-value*
Mean Readability Rating	7.24	6.32	4.27	< 0.001
Std. Deviation	(0.15)	(0.15)		
N	154	154		

Panel B - Readability Ratings for the More Readable Individual Excerpts

	Wells Fargo	*Kroger*	*PepsiCo*
Bog Index	2	14	19
Mean Readability Rating	7.25	7.69	6.78
Std. Deviation	(0.24)	(0.20)	(0.28)
N	52	51	51

Panel C - Readability Ratings for the Less Readable Individual Excerpts

	Phillips 66	*International FC Stone*	*United Technologies*
Bog Index	89	93	81
Mean Readability Rating	6.47	5.33	7.13
Std. Deviation	-0.25	-0.3	-0.22
N	51	51	52

Panel D - Within-Participant Differences

	Mean Difference in Readability Rating Between More and Less Readable Disclosure	*Median Difference in Readability Rating Between More and Less Readable Disclosure*	*Proportion of Differences in Readability Ratings Between More and Less Readable Disclosure that are Non-Negative*
Value	1.12	0.50	79.22% (122 of 154)
T-Stat	6.23		
Signed Rank		2174.5	
p-value	< 0.001	< 0.001	
N	154	154	

In total, we recruit 154 participants from the AMT platform. On average, participants are 35.67 years old, with 14.26 years of full-time work experience, and 29.22 (30.52) percent report having directly (indirectly) invested in the past. We note that these demographics are once again similar to prior studies (Krische, 2015) and not substantially different across other experiments in our study.

Panel A of Table 2 shows that the average ratings of the lower *Bog Index* excerpts are significantly higher than that of the higher *Bog Index* excerpts, in terms of ease of readability (t=4.27), controlling for individual participants. Panel B reports the *Bog Index* and participant readability ratings for the three most readable disclosures from the 2014 Fortune 100 firms, where all three are classified in *StyleWriter's* "Excellent" category. Panel C reports the *Bog Index* and participant readability ratings for the three least readable disclosures from the 2014 Fortune 100 firms, where all three are classified in *StyleWriter's* "Poor" category. In Panel D, we also examine differences in judgments within participants. We take each participant's rating for the more readable excerpt and subtract his or her rating for the less readable excerpt. We find that the mean and median differences are significantly greater than zero, and that nearly 80 percent of the readability rating differences are non-negative, which further supports the notion that individual participants, on average, believe the lower *Bog Index* disclosures are easier to read than the higher *Bog Index* disclosures.[13]

4.1.3. Analysis of plain English regulatory intervention

Both the controlled experiment and quasi-experiment provide evidence consistent with the Bog Index being a valid proxy for plain English readability. In this section, we examine the response of the Bog Index to an SEC regulatory intervention that required firms to improve the plain English readability of firms' SEC filings. Specifically, we re-examine the results documented in Loughran and McDonald (2014b) related to the 1998 *Plain English Mandate*. This mandate required firms to use plain English in prospectuses filed after October 1, 1998. For this analysis, we examine the changes in five measures of readability that cover both the clarity and quantity of disclosure: *Bog Index*, *Fog Index*, *LM PE Index*, Log(*Total File Size*), and Log(*Total Words*).

We begin by identifying prospectus filings (Forms S-1, S-3, or S-4) during 1996, 1997, 1999, and 2000, the two years before and after the adoption of the *Plain English Mandate*. To ensure comparability across entire period, we require a firm to make a prospectus filing both before and after the mandate. This results in a final sample of 772 prospectus filings. As previously discussed, we calculate each of the readability measures on these prospectuses after applying the parsing procedures outlined in Appendix C. Table 3 presents averages for the readability measures across the pre- and post-intervention periods. Although all of the writing clarity measures of readability (*Bog Index*, *Fog Index*, and *LM PE Index*) significantly decrease from the pre- to post-period, there is no evidence of a significant decrease in the two quantity-based measures of readability (Log(*Total File Size*), and Log(*Total Words*)) after the mandate. This evidence suggests that the quantity-based measures of readability are not capturing the aspects of readability mandated by the SEC.

Table 3 also provides descriptive information on the control variables we incorporate in our multivariate regressions. In particular, we follow Loughran and McDonald (2014a) and include the following controls: (1) *Pre-filing alpha,* the intercept from a market model estimated prior to the filing date; (2) *Pre-filing RMSE*, the root mean squared error from a pre-period market model regression; (3) *Abs(Abnormal Return),* the absolute value of the 2-day buy-and-hold abnormal return around the filing date (0,1); (4) *Log(market capitalization)*, the natural logarithm of market capitalization on the day before the filing date; (5) *Log(book-to-market)*, the natural logarithm of the book-to-market ratio calculated prior to the filing date; and (6) *NASDAQ dummy*, a dummy variable set equal to one if the firm trades on NASDAQ and zero otherwise.[14] Across these control variables, we find that only pre-filing volatility, the absolute returns, market capitalizations, and book-to-market values are significantly higher in the post-period.

We next implement regression analysis where we control for these variables and include firm fixed effects to mitigate any concerns that these increases are influencing our results. We formally test the change in each of the readability measures after the regulatory intervention using a binary variable, *Post_PE_Regulation,* that is equal to one if the prospectus was filed during the 1999–2000 period and zero if it was filed in 1996–1997, as the variable of interest. Assuming general compliance with the regulation, we expect a negative coefficient on *Post_PE_Regulation* if a particular proxy captures the construct of plain English readability.

Table 4 presents the results of our tests of changes in the different readability measures around the *Plain English Mandate*. In order to compare across the different readability measures, we transform all variables in our regression models to be mean zero and standard deviation of one. Consistent with Loughran and McDonald's (2014b) finding of a decrease in their measure of plain English readability after the mandate, Table 4 shows that *Bog Index*, *Fog Index*, and *LM PE Index* all improve following the regulation. In contrast, we are unable to document a change in Log(*Total File Size*) or Log (*Total Words*) following the regulation, suggesting that these two measures do not capture the plain English improvements prescribed by the SEC.

In terms of relative significance, we document that the change in readability around the regulation for *Bog Index* is approximately 1.21 standard deviations. In comparison to the magnitudes of the estimated coefficients on the other two clarity measures of readability, *Bog Index* increases slightly less than 10 percent more than *LM PE Index* and 129 percent more than *Fog Index*. Overall, these results suggest that while none of the quantity-based measures of disclosure readability change around the SEC *Plain English Mandate*, all three of the writing clarity measures of readability appear to improve after the mandate with *Bog Index* best capturing the plain English attributes mandated by the SEC.

To support the conclusion that the results documented in Table 4 are driven by improvements related to the mandate and not a broad trend toward improved readability, we next perform a set of placebo tests. In particular, we conduct two falsification tests with prospectuses around other periods not affected by the regulation, both before and after the real regulation. We begin by re-estimating the same regressions as in Table 4 using prospectuses filed during

[13] In contrast to the experiment where we alter the firm's disclosure to hold constant the other readability measures, in this case we rely on actual disclosure excerpts from the firm. While total words and file size are held constant across the disclosures, the more readable *Bog Index* excerpts also have more readable *Fog Index* and *LM PE Index* scores. As such, our quasi-experimental evidence suggests that all of the writing clarity measures (i.e., *Bog Index*, *Fog Index*, and *LM PE Index*) capture readability that is incremental to the quantity based measures of disclosure.

[14] To remain consistent with prior research, we include this same set of control variables in all of our analyses. All of the results formally presented in the paper are robust to the inclusion of additional controls for: business complexity (squared business segment sales proportions), market volatility (VIX), performance (ROA), accounting detail (percent of non-missing Compustat items) and dummy variables to control for various firm-level events (M&A activity, special items, and discontinued operations).

Table 3
Variable Means for Prospectus Filings.

This table reports descriptive information for the key variables in our sample of prospectus filings around the 1998 s Plain English Mandate. The sample consist of consists of 772 firm-issuances of debt or equity in 1996, 1997, 1999, or 2000. All readability measures are based on the readability of the prospectus filing. Cluster robust t-test statistics are clustered by firm. Detailed variable definitions are provided in Appendix B.

Variable	Full Sample	Pre-Period 1996 to 1997	Post-Period 1999 to 2000	Pre - Post Differences	t-stat of Difference
Measures of Readability:					
Bog Index	94.98	103.79	86.27	−17.52	(−18.43)
Fog Index	22.52	23.21	21.84	−1.36	(−6.67)
LM PE Index	0.90	2.61	−0.80	−3.41	(−24.76)
Total File Size (in mb)	0.43	0.43	0.44	0.01	(0.49)
Total Words (000's of words)	55.34	55.51	55.18	−0.34	(0.92)
Control Variables:					
Pre-filing alpha	0.04	0.05	0.02	−0.02	(−1.61)
Pre-filing RMSE	2.57	2.11	3.03	0.92	(14.55)
Abs(abnormal return)	0.03	0.02	0.04	0.02	(7.07)
Market capitalization (in $ millions)	7455.10	4997.32	9887.54	4890.22	(7.07)
Book-to-market	0.46	0.44	0.48	0.04	(2.06)
NASDAQ dummy	0.23	0.24	0.22	−0.02	(−1.28)
Number of Observations	772	384	388		

Table 4
An Analysis of Plain English Measures around SEC Plain English Mandate.

This table provides evidence of narrative disclosure measures related to prospectus filings made around the 1998 Plain English Mandate. All OLS regression results include an intercept, filing form fixed effects, and firm fixed effects. The t-statistics are reported in parentheses below the coefficient, where standard errors are clustered by firm. For ease of interpretation all regression variables are standardized with a mean of zero and standard deviation of one. The sample for both panels consist of 772 firm-issuances of debt or equity in 1996, 1997, 1999, or 2000. 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 if the issuance is made in 1996 or 1997. Detailed variable definitions for all other variables are provided in Appendix B.

Readability Measures	(1) *Bog Index*	(2) *Fog Index*	(3) *LM_PE_Index*	(4) *Log(Total File Size)*	(5) *Log(Total Words)*
Post_PE_Regulation	-1.210	-0.528	-1.118	-0.018	-0.031
	(-11.67)	(-4.56)	(-15.95)	(-0.20)	(-0.33)
Control Variables:					
Pre-filing alpha	0.030	-0.033	0.003	-0.036	-0.036
	(0.61)	(-0.63)	(0.08)	(-0.77)	(-0.75)
Pre-filing RMSE	-0.137	-0.155	-0.135	-0.145	-0.168
	(-1.38)	(-1.29)	(-1.81)	(-1.52)	(-1.69)
Abs(abnormal return)	0.005	0.096	-0.023	0.110	0.119
	(0.11)	(1.85)	(-0.57)	(2.40)	(2.49)
Log(market capitalization)	0.378	0.417	-0.048	-0.002	-0.009
	(2.20)	(2.08)	(-0.38)	(-0.01)	(-0.06)
Log (book-to-market)	0.132	0.114	-0.044	-0.134	-0.126
	(1.53)	(1.21)	(-0.65)	(-1.52)	(-1.36)
NASDAQ dummy	0.208	0.267	-0.077	-0.353	-0.349
	(0.71)	(1.09)	(-0.32)	(-1.57)	(-1.52)
Filing Type Fixed Effects	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes
# of OBS	772	772	772	772	772
R^2	0.6201	0.5445	0.7405	0.6464	0.5389

1994–1997 with January 1, 1996 as a pseudo-regulation date. We identify prospectus filings during 1994–1997 and use *Post_PE_Regulation_False_Prior* to designate observations during 1996 and 1997. We require a firm to make a prospectus filing both before and after the pseudo-regulation date, which results in a sample of 390 prospectus filings. In contrast to our tests around the actual regulation, we do not expect to find a significantly negative coefficient on any of the 'false' post regulation variables. Panel A of Table 5 reports the results of these regressions. Consistent with the SEC motivation that prior to the mandate filings were becoming increasingly less readable, we

Table 5

Plain English Measures around Placebo Regulation Dates (Falsification Tests).

This table provides a series of falsification tests related to changes in narrative disclosure measures related to prospectus filings in periods before and after the 1998 Plain English Mandate. Our expectation is that the measures should not change around these placebo dates. Panel A provides results for the falsification (placebo) tests prior to the actual regulation, where the sample consists of 390 firm-issuances of debt or equity in 1994–1997. Post_PE_Regulation_False_Prior is a binary variable equal to one if a firm-year is after the adoption of the false plain English regulation date (i.e., 1996, 1997) and zero if the issuance is made in 1994 or 1995. Panel B provides results for the falsification (placebo) tests after the actual regulation, where the sample consists of 971 firm-issuances of debt or equity in 1999–2002. Post_PE_Regulation_False_After is a binary variable equal to one if a firm-year is after the adoption of the false plain English regulation date (i.e., 2001, 2002) and zero if the issuance is made in 1999 or 2000. The t-statistics are reported in parentheses below the coefficient, where standard errors are clustered by firm. For ease of interpretation all regression coeficients are standardized with a mean of zero and standard deviation of one. All OLS regressions include an intercept, control variables, filing form fixed effects, and firm fixed effects. Detailed variable definitions for all other variables are provided in Appendix B.

Panel A - Plain English Prior Regulation Falsification Test

Readability Measures	(1) *Bog Index*	(2) *Fog Index*	(3) *LM_PE_Index*	(4) *Log(Total File Size)*	(5) *Log(Total Words)*
Post_PE_Regulation_False_Prior	0.337	0.155	-0.088	0.061	0.064
	(2.05)	(0.85)	(-0.97)	(0.51)	(0.51)
Control Variables	Yes	Yes	Yes	Yes	Yes
Filing Type Fixed Effects	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes
# of OBS	390	390	390	390	390
R^2	0.5477	0.5285	0.584	0.6305	0.6235

Panel B - Plain English After Regulation Falsification Test,

Readability Measures	(1) *Bog Index*	(2) *Fog Index*	(3) *LM_PE_Index*	(4) *Log(Total File Size)*	(5) *Log(Total Words)*
Post_PE_Regulation_False_After	0.024	0.147	-0.164	0.237	0.155
	(0.65)	(3.25)	(-4.24)	(4.32)	(2.80)
Control Variables	Yes	Yes	Yes	Yes	Yes
Filing Type Fixed Effects	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes
# of OBS	971	971	971	971	971
R^2	0.6459	0.6332	0.7412	0.6387	0.6384

find evidence of an increase in the *Bog Index*. We are unable to find similar evidence that any of the readability measures appear to substantially change around the pseudo-regulation date prior to the real regulation.

We next examine a second pseudo-regulation date that occurs after the real regulation date (i.e., January 1, 2000). Specifically, in Panel B of Table 5 we identify prospectus filings during the 1999–2002 period and use *Post_PE_Regulation_False_After* to designate observations during 2001–2002. After reducing the sample to include only observations where firms filed before and after this second pseudo-regulation date, our sample consists of 971 prospectus filings. Consistent with the improvement in the Bog Index being observable only in the period around the mandate, we are unable to find a significant change in *Bog Index*. Additionally, we show significant increase in *Fog Index* (t=3.25), as well as both quantity based measures of readability (t=4.32 and 2.80, respectively), suggesting that improvements in these measures were short-lived and reverse after the mandate. Most concerning, we find that the coefficient estimate on *LM PE Index* continues to be negative in the period after the mandate ($t=-4.32$), which suggests a general trend of improved readability using the LM PE Index after the mandate.

Overall, our examination of the 1998 *Plain English Mandate* suggests that clarity-based measures of readability captured the changes in readability around the regulatory intervention. In terms of relative magnitude and evidence from the falsification tests, the Bog Index appears to best capture the changes in plain English surrounding the regulation. Further, there is some evidence of an overall trend in improved readability using the LM PE Index. Finally, neither quantity-based measure of readability (file size and total words) appears to change around the regulation, suggesting that quantity based measures do not capture plain English attribute improvements mandated by the SEC.

5. RMSE and analyst tests

In Section 4, we provide evidence that the Bog Index is a valid measure of plain English readability. In this section, we turn our attention to the question of whether plain English readability affects capital market participants. To do so, we follow the tests used in Loughran and McDonald (2014a): stock return volatility, analyst forecast dispersion, and analyst earnings forecast accuracy. When we conduct these tests, we measure *Bog Index*, *Fog Index*, *LM PE Index*, Log(*Total File Size*) or Log(*Total Words*) on a sample of 10-K filings spanning 1994–2011. To mitigate endogeneity concerns raised by Li (2010), we replace the industry fixed effects used in Loughran and McDonald (2014a) with firm fixed effects that control for time-invariant firm characteristics.

5.1. Sample composition

We begin with all 10-K documents (e.g., 10-K, 10-K405, 10KSB, and 10KSB40) that contain more than 3,000 words and have filing dates on EDGAR between 1994 and 2011. Using this initial sample of 10-K filings, we parse and calculate the readability of each filing using the procedures described in Appendix C. We then implement the same data restrictions used by Loughran and McDonald (2014a) to be consistent with that study. Specifically, we require a CRSP PERMNO match (dropping 32,451), a stock price of greater than $3 (dropping 13,990), ordinary common equity (dropping 5,391), a positive book-to-market (dropping 2,237), a filing date greater than 180 days from the prior filing (dropping 113), and pre- and post-market model data available (dropping 320). After these exclusions, the final sample used in our primary tests examining post-filing stock return volatility consists of 66,173 observations.[15]

In addition to our stock return volatility tests, we also conduct tests using properties of analysts' earnings forecasts, as in Loughran and McDonald (2014a). The two outcome variables used are analyst earnings forecast dispersion (*Dispersion*) and analyst earnings forecast accuracy (*Accuracy*). 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, the sample size for the *Dispersion* regressions falls to 37,642 and for the *Accuracy* regressions falls to 46,424.

5.2. Descriptive data

Table 6 reports mean summary statistics for the variables used in our capital market tests. For comparative purposes, we divide the sample into two distinct time periods following Loughran and McDonald (2014a). Columns (1), (2), and (3) report averages for filings between 1994 and 2002, filings between 2003 and 2011, and the entire sample period, respectively. Our three outcome variables are *Post-filing RMSE*, *Dispersion*, and *Accuracy*. *Post-filing RMSE* is the root mean squared error multiplied by 100 from a market model estimated over trading days +6 to +28 relative to the 10-K filings date. *Dispersion* is the standard deviation of analysts' earnings forecasts for the first quarter following the 10-K filing, scaled by stock price immediately before the 10-K filing, using only forecasts after the 10-K filing date. *Accuracy* is the absolute value of the stock-price-scaled difference between actual earnings and the mean analyst forecast for the first fiscal quarter following the 10-K filing using forecasts after the 10-K filing date. From the early part of our sample period to the later part the average *Post-filing RMSE* declines from 3.38 to 2.24; however, both the average values of *Dispersion* and *Accuracy* increase over this same period by 0.05 and 0.02, respectively.

We next examine how our readability measures calculated on our 10-K filings change over the sample period. Fig. 1 shows how the readability measures change over the entire sample period. Both *Bog Index* and *Fog Index* appear to increase prior to the 1998 *Plain English Mandate*, then fall after the regulatory intervention before increasing through the remainder of the sample period. Both quantity of disclosure measures, Log*(Total File Size)* and Log*(Total Words)*, steadily increase throughout the sample period, suggesting the systematic worsening of readability over the 1994–2011 period. In contrast to all the other readability measures and pundits' concerns that readability has gotten worse over the past few decades, *LM PE Index* decreases throughout the sample period. This suggests that based on *LM PE Index* the average readability of 10-K filings has steadily improved from 1994–2011. This pattern in *LM PE Index* is consistent with the post-regulation falsification test detailed in the prior section, where the measure rejected the null hypothesis around the pseudo-regulation dates.

5.3. Impact of readability measures in regressions of root mean squared error

We begin the capital market tests by examining the association between the readability measures and *Post-filing RMSE*. To the extent that readability affects stock market investors, we expect a positive association between the readability measures and *Post-filing RMSE*. Following Loughran and McDonald (2014a), we include all the control variables discussed in our previous model. Table 7 reports the results of our regression estimations with *Post-filing RMSE* as the dependent variable and including firm fixed effects. *Bog Index* and *Fog Index* are positively associated with *Post-filing RMSE*, suggesting that less readable annual reports generate greater stock return volatility following the public release of the annual report. Similarly, the quantity-based measures, *Log(Total File Size)* and *Log(Total*

[15] Our final sample of 66,173 is similar to the 66,707 observations used Loughran and McDonald (2014a). Our descriptive statistics from Table 6 are also similar to those reported in Loughran and McDonald (2014a).

Table 6
Variable Means for 10-K Filings.

This table reports descriptive information for the key variables in our sample. Our primary regressions examining volatility subsequent to the 10-K filing are based on the entire sample of 66,173 observations. This sample is reduced to 46,424 observations for analyst forecast accuracy (Accuracy) and 37,642 for analyst forecast dispersion (Dispersion). Detailed variable definitions for all other variables are provided in Appendix B.

Variable	Full Sample	1994 to 2002	2003 to 2011
Dependent Variables:			
Post-filing RMSE	2.84	3.38	2.24
Dispersion	0.47	0.45	0.50
Accuracy	0.24	0.23	0.25
Measures of Readability:			
Bog Index	81.63	79.71	83.78
Fog Index	19.35	19.13	19.60
LM PE Index	0.00	0.72	-0.81
Total File Size (in mb)	1.33	0.40	2.37
Total Words (000's of words)	29.04	21.35	37.64
Control Variables:			
Pre-filing alpha	0.06	0.08	0.05
Pre-filing RMSE	3.08	3.49	2.63
Abs(abnormal return)	0.03	0.04	0.03
Market capitalization (in $ millions)	2254.94	1679.45	2899.05
Book-to-market	0.63	0.63	0.63
NASDAQ dummy	0.59	0.60	0.58
# of Analysts	6.55	6.11	7.00
Number of Observations	66173	34948	31225
Observations Analyst Accuracy	46424	46424	46424
Observations Analyst Dispersion	37642	37642	37642

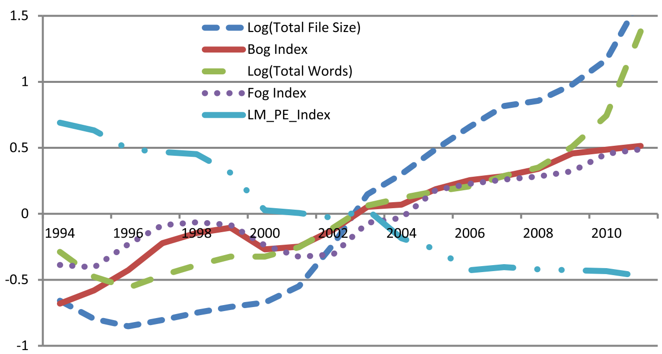


Fig. 1. Trend of 10-K Readability Measures by Year. This figure shows trend in each of the readability measures by year over our entire sample period. For ease of interpretation all regression coeficients are standardized with a mean zero and standard deviation of one.

Words), are also positively associated with future stock return volatility, which is consistent with Loughran and McDonald (2014a). A comparison across these measures shows that *Bog Index* has a nearly twenty-five percent greater association with future stock return volatility than any other measure of readability. Finally, the coefficient on *LM PE Index* is statistically indistinguishable from zero in both panels, which would lead a researcher to conclude that readability has no effect on stock market volatility if he or she used *LM PE Index* as the proxy for readability.

5.3.1. Impact of readability measures in regressions of analyst processing of information

Following our capital market tests using stock return volatility as the outcome variable of interest, we test whether readability affects more sophisticated market participants such as sell-side equity analysts. We use both analyst forecast dispersion (*Dispersion*) and analyst accuracy (*Accuracy*) as our outcome variables of interest for this set of tests. In addition to the inclusion of all the control variables from our previous models, we also follow Loughran and McDonald (2014a) and

Table 7
An Analysis of Narrative Disclosure Measures Using Post-Filing Date Market Model.
This table reports regression results for the narrative disclosure measures, where the market model root mean squared error (RMSE) for trading days [6, 28] relative to the 10-K filing date is the dependent variable. All OLS regressions include an intercept, year fixed effects, and firm fixed effects. For ease of interpretation all regression coeficients are standardized with a mean of zero and standard deviation of one. The t-statistics are reported in parentheses below the coefficient, where standard errors are clustered by firm and year. Regressions include 66,173 firm-year observations during 1994 to 2011. Detailed variable definitions are provided in Appendix B.

	(1) *RMSE*	(2) *RMSE*	(3) *RMSE*	(4) *RMSE*	(5) *RMSE*
Narrative Disclosure Measure:					
Bog Index	0.035				
	(4.28)				
Fog Index		0.022			
		(3.39)			
LM PE Index			-0.002		
			(-0.21)		
Log(Total File Size)				0.028	
				(2.89)	
Log(Total Words)					0.022
					(4.25)
Control Variables:					
Pre-filing alpha	-0.048	-0.048	-0.048	-0.048	-0.047
	(-2.13)	(-2.14)	(-2.16)	(-2.13)	(-2.12)
Pre-filing RMSE	0.436	0.436	0.437	0.436	0.436
	(20.56)	(20.66)	(20.67)	(20.59)	(20.63)
Abs(abnormal return)	0.000	0.000	0.092	0.092	0.092
	(16.93)	(16.91)	(16.94)	(16.97)	(16.98)
Log(market capitalization)	-0.147	-0.145	-0.147	-0.151	-0.151
	(-2.27)	(-2.23)	(-2.28)	(-2.32)	(-2.35)
Log (book-to-market)	-0.071	-0.070	-0.070	-0.071	-0.071
	(-5.58)	(-5.45)	(-5.47)	(-5.57)	(-5.58)
NASDAQ dummy	0.082	0.081	0.081	0.081	0.082
	(3.26)	(3.20)	(3.19)	(3.18)	(3.22)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes
# of OBS	66173	66173	66173	66173	66173
R^2	0.6617	0.6616	0.6614	0.6616	0.6616

include the number of analysts in both of our analyst related analyses.

Panels A and B of Table 8 report the results from estimation of regressions with *Dispersion* and *Accuracy* as the respective dependent variables. In Panel A, we do not find any evidence of a statistically significant association between any of the readability measures and analyst forecast dispersion. While these results are seemingly inconsistent with the significantly positive association between file size and analyst dispersion shown in Loughran and McDonald (2014a), the difference stems from our use of firm fixed effects. In fact, when we substitute (untabulated) our firm fixed effects with the industry fixed effects used in Loughran and McDonald (2014a), *Bog Index*, *Log(Total File Size)*, and *Log(Total Words)* are all positively associated with the dispersion in analysts' earnings forecasts (t=2.07, 3.99, and 4.10, respectively). Combined, these findings indicate that time-invariant firm characteristics are likely driving the association between readability and analyst dispersion document in prior literature.

In Panel B of Table 8, we report results from firm fixed effect regressions that use *Accuracy* as the dependent variable and are only able to document a positive association between *Log(Total File Size)* and *Log(Total Words)* and analysts' forecast accuracy (t=2.54 and t=3.16, respectively). These results are consistent with those documented in Loughran and McDonald (2014a). In sum, the results of Panel B suggest that the quantity of disclosure affects sophisticated information intermediaries, but not the clarity of disclosure.

5.4. Components of quantity-based measures

While our evidence thus far is consistent with the quantity of disclosure affecting sophisticated information intermediaries such as sell-side equity analysts, it is important to understand the sources of disclosure quantity. With respect to the file size of the filing, Panel A of Fig. 2 shows that a significant amount of 10-K file size relates to something other than the

Table 8
An analysis of narrative disclosure measures using analyst outcomes as the dependent variables.

Panel A - Analyst Dispersion

	(1) *Dispersion*	(2) *Dispersion*	(3) *Dispersion*	(4) *Dispersion*	(5) *Dispersion*
Narrative Disclosure Measure:					
Bog Index	0.005				
	(0.42)				
Fog Index		0.001			
		(0.17)			
LM PE Index			-0.004		
			(-0.41)		
Log(Total File Size)				0.010	
				(0.66)	
Log(Total Words)					0.009
					(0.76)
Control Variables	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes
# of OBS	37642	37642	37642	37642	37642
R^2	0.4937	0.4937	0.4937	0.4937	0.4937

Panel B - Analyst Accuracy

	(1) *Accuracy*	(2) *Accuracy*	(3) *Accuracy*	(4) *Accuracy*	(5) *Accuracy*
Narrative Disclosure Measure:					
Bog Index	0.012				
	(0.92)				
Fog Index		-0.010			
		(-0.90)			
LM PE Index			0.004		
			(0.38)		
Log(Total File Size)				0.030	
				(2.54)	
Log(Total Words)					0.023
					(3.16)
Control Variables	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes
# of OBS	46424	46424	46424	46424	46424
R^2	0.4544	0.4544	0.4544	0.4546	0.4546

text of Form 10-K itself, particularly in the sample years after 2001.

Panel B of Fig. 2 further illustrates the sources of variation in file size. In particular, exhibits, such as material contracts or bond indentures, and tabular data comprise nearly as much disclosure quantity as the 10-K text itself. Furthermore, as firms began to file their 10-Ks in HTML format and later with XBRL tags, there is a very large increase in disclosure quantity related to these markup tags, as well as decoded pictures and spreadsheet files. Appendix H illustrates the detailed components of file size. Given this substantial variation in the file size measure over time is driven by the inclusion of content unrelated to the underlying text in the 10-K (e.g., HTML, XML, etc.), we caution researchers to be careful in making inferences using file size measures of readability across time.

Further, while prior research primarily uses aggregate measures of file size to measure the quantity-based readability of the 10-K filing (Loughran and McDonald, 2014a), it is plausible that the file size related to other parts of the filing could also be important drivers of the observed associations. To explore this issue in greater depth, we disaggregate total file size of the 10-K filing into its components. Appendix C outlines the process we use to disaggregate the file size measure into its components.

We re-estimate the three regression models previously examined (*Post-filing RMSE*, *Dispersion*, and *Accuracy*), but now separately examine the components of file size. For comparative purposes, we also include the total file size specification

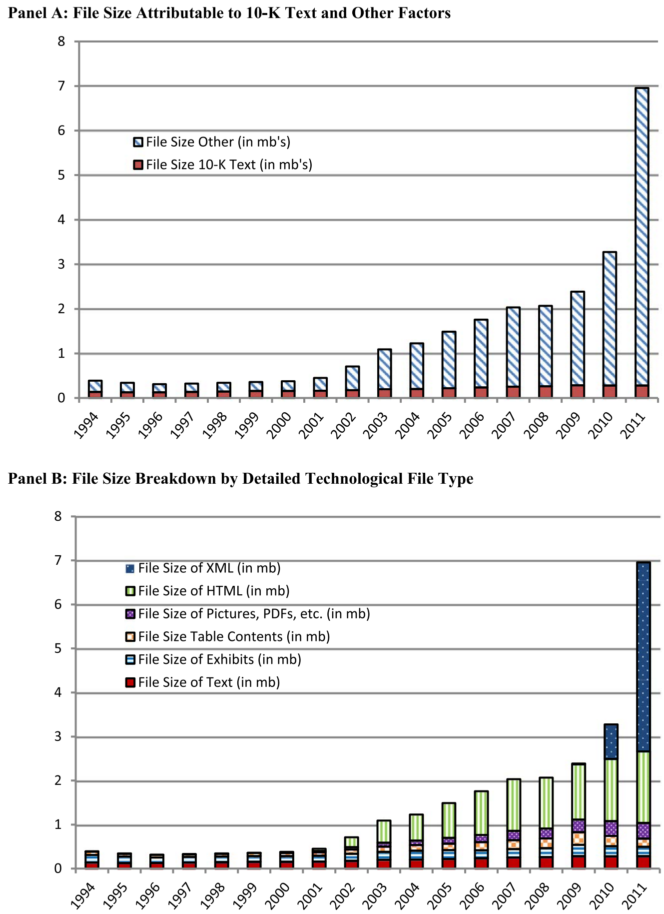


Fig. 2. Breakdown of 10-K File Size Components by Year. This figure shows the breakdown of the components of 10-K file size by year. Panel A provides a breakdown by year of the file size related to the text of the 10-K filing and non-textual components included in the 10-K filing. Panel B provides a breakdown by year of the primary components of text, exhibits, table contents, Pictures and PDFs, HTML, and XML, Pictures on file size.

previously reported as a reference point for interpreting the disaggregated results. Table 9 reports the results for the models including file size. As previously documented, *Log(Total File Size)* is associated with *Post-filing RMSE* (t=2.89) and *Accuracy* (t=2.54), but not *Dispersion* (t=0.66).

To conduct our disaggregated analysis, we separate the non-textual portion of the file size into the primary components that exist across our sample period to examine which components are associated with each capital market outcome variable. The components include the text of Form 10-K (primarily the risk factors discussion, MD&A, and footnotes), tabular data, attached exhibits (e.g., compensation contracts, supplier/customer agreements, or bond indentures), and other non-textual content. The “other” component primarily reflects technical components used to render information (e.g., HTML, XBRL, PDFs, Pictures). Column (2) in Table 9 shows that there is no significant evidence that file size derived from the text of Form 10-K is associated with stock return volatility. Instead, the results indicate that file size derived from the text in exhibits and tables are the only components that remain significantly associated with future stock return volatility across both regression specifications at conventional significance levels.[16] These results suggest that changes in the business environment reflected by additional exhibits (e.g., compensation contracts, supplier/customer agreements, bond indentures) and more quantitative data are driving the association with future volatility.

Further, in contrast to the previously discussed insignificant association between *Log(Total File Size)* and *Dispersion*, Column (4) shows that after separating out the non-textual components, the file size components related to the actual text of Form 10-K and exhibits are significantly associated with *Dispersion*. Similarly, the evidence in Column (6) suggests that

[16] Loughran and McDonald (2014a) note that their results are essentially the same using net file size (i.e., the text of the 10-K filing). When we combine the text from the 10-K and the text from the exhibits, we also find (untabulated) that total text in the filing is associated with greater future volatility. This evidence highlights the importance of considering the exhibits often appended to 10-K filings when calculating quantity based readability measures.

Table 9
An Analysis of Components of Content Based Measures of Readability.
This table reports regression results for breakdown of the File Size measure, where the dependent variables are RMSE, Dispersion, and Accuracy. For ease of interpretation all regression variables are standardized to have a mean of zero and standard deviation of one. All OLS regressions include an intercept, controls, firm fixed effects, and year fixed effects. Standard errors are clustered by firm and year. The t-statistics are reported in parentheses below the coefficient. Regressions where RMSE is the dependent variable, consist of 66,173 firm-year observations during 1994 to 2011. Regressions where Dispersion (Accuracy) is the dependent variable consist of 37,642 (46,424) firm-year observations during 1994 to 2011. Detailed variable definitions are provided in Appendix B.

	(1)	(2)	(3)	(4)	(5)	(6)
	RMSE	*RMSE*	*Dispersion*	*Dispersion*	*Accuracy*	*Accuracy*
Log(Total File Size)	0.028		0.010		0.030	
	(2.89)		(0.66)		(2.54)	
Breakdown of File Size Based Measure:						
Log(File Size 10-K Text)		0.009		0.071		0.083
		(0.98)		(4.90)		(5.37)
Log(File Size Table Contents)		0.012		-0.006		-0.006
		(2.99)		(-0.90)		(-0.94)
Log(File Size Exhibits)		0.009		0.017		0.010
		(2.52)		(3.06)		(1.58)
Log(File Size Other Components)		0.000		-0.004		0.010
		(-0.03)		(-0.32)		(0.82)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
# of OBS	66173	66173	37642	37642	46424	46424
R^2	0.6616	0.6616	0.4937	0.4948	0.4546	0.4558

the text of the 10-K filing appears to influence analyst accuracy.

Overall, the evidence from Tables 7–9 illustrates that both quantity-based measures of readability and writing clarity-based measures of readability have validity in certain instances. As such, while we recommend that researchers use the Bog Index when studying the effects of writing clarity, we also note that quantity-based measures of readability may also be appropriate in certain settings. However, the evidence from Columns (2) and (4) of Table 9 suggests that the use of total file size as a proxy for the readability of the text of the 10-K filing without considering the other components of file size may lead to over/under rejection depending on the influence of the non-textual components. As such, we recommend that researchers carefully examine the relationships between their outcome variables of interest and the components of quantity-based measures of readability to reduce the likelihood of Type I and Type II errors in their analyses.

6. Conclusion

In response to the call by Core (2001) for greater use of techniques from computational linguistics to capture large-sample measures of disclosure quality, there has been a substantial increase in the literature devoted to textual analysis. Most of the research that has focused on the effects of readability selects from a rather limited set of existing readability measures that reflect either writing clarity (e.g., Fog Index) or disclosure quantity (e.g., file size of the filing). We attempt to fill this void by introducing a new measure of readability, the Bog Index, which is designed to capture the plain English attributes of disclosure (e.g., active voice, fewer hidden verbs, etc.). We validate this new measure based on a combination of regulatory guidance, experimental validation, and archival tests surrounding a regulatory intervention related to the readability of prospectus filings. In particular, our controlled experiments show that participants who receive the more readable disclosure, as measured by the *Bog Index*, rate the disclosure to be significantly easier to read than participants who receive a less readable disclosure. Further, the significant improvements in the *Bog Index* relative to other readability measures around the 1998 *Plain English Mandate* suggests that the *Bog Index* best captures the writing attributes enforced by regulators after the regulation.

Following our validation tests, we test whether readability affects capital market outcomes including future stock market volatility and equity analysts' earnings forecast properties. We find that while most readability measures are associated with future stock market volatility, the *Bog Index* has nearly a twenty-five percent greater association than the next closest readability measure. In contrast, we find that only quantity-based measures of disclosure (e.g., total words and file size) are significantly associated with analyst forecast accuracy. Combined, this evidence suggests various types of financial statement users are affected by different readability attributes.

We recommend that researchers use the Bog Index when studying the implications of writing clarity in firms' financial disclosures. At the same time, our archival evidence suggests that in certain instances quantity-based measures of readability may also be useful in explaining certain economic outcomes. As such, since a vast amount of the variation in Form10-K file size over time is driven by the inclusion of content unrelated to the underlying text in the 10-K (e.g., HTML, XML, etc.), we recommend that when using a quantity-based measure of readability researchers carefully examine the relationships between their outcome variables of interest and the components of the quantity-based measure. For example, we recommend that researchers examine the relationships between the components of quantity-based measures of readability, such as Log*(Total File Size)*, and their variables of interest to reduce the likelihood of incorrect non-rejection of null hypotheses. In particular, we caution researchers against comparing total file size over time as the vast amount of the variation in the file size measure over time is driven by the inclusion of content unrelated to the underlying text in the 10-K (e.g., HTML, XML, PDFs).

Appendix A. Examples of Plain English problems[17]

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.

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.

[17] 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.

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 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 B. Variable definitions

Variable Names	Definitions
Readability Measures:	
Bog Index	is a proprietary measure of readability created by Editor Software's plain English software, *StyleWriter*. The formula is based on several plain English factors such as sentence length, passive voice, weak verbs, overused words, complex words, and jargon. Higher values of the index imply lower readability.
Fog Index	is the Gunning (1952) Fog Index calculated using Perl's Lingua::EN::Fathom module
LM PE Index	is a measure of plain English readability based on Loughran and McDonald (2014b), which penalizes documents for longer sentences, longer words, more passive voice, more legalese and other poor writing conventions and rewards the use of more personal pronouns
Total File Size	is the file size (in megabytes) of the complete 10-K filing (including appendices) from EDGAR.
Total Words	is the number of words (in 000's) contained in the complete 10-K filing (including appendices) from EDGAR.
Plain English Regulation Variables:	
Post_PE_Regulation	is one if the 10-K was filed after the 1998 Plain English Mandate (i.e., 1999 or 2000) and zero if the 10-K was filed during 1996 or 1997.
Post_PE_Regulation_False_Prior	is one if the 10-K was filed after the *pseudo* regulation date (1/1/1996) prior to the actual regulation (i.e., 1996 or 1997) and zero if the 10-K was filed during 1994 or 1995.
Post_PE_Regulation_False_After	is one if the 10-K was filed after the *pseudo* regulation date (1/1/2001) following the actual regulation (i.e., 2001 or 2002) and zero if the 10-K was filed during 1999 or 2000.
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.
Accuracy	is the absolute value of analyst accuracy multiplied by 100. Analyst accuracy 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.
Dispersion	is the standard deviation of analysts' forecasts appearing in the dispersion 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.
Market 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.
# of Analysts	is the number of analysts.

Appendix C. Procedures to parse and calculate readability measures

This appendix provides an overview of the procedures used to: (1) parse the prospectus and 10-K filings from the Section, (2) calculate the readability measures, and (3) compute the file size and word components of filings.

Parsing of prospectus filings

To clean the text of prospectus filings (S-1, S-3, S-4) from the SEC's EDGAR system, we follow the approaches of Li (2008) and Loughran and McDonald (2014a, 2014b) for 10-K filings.[18] Specifically, after downloading the raw text files from the EDGAR FTP site, we complete the following parsing steps:

1. Remove <DIV>, <TR>, <TD>, and <FONT> tags.

[18] Specifically, we follow the parsing outline provided on Bill McDonald's website, which can be accessed at http://www3.nd.edu/~mcdonald/Word_Lists_files/Documentation/Documentation_StageOne_10-X_Parse.pdf.

2. Remove all content between <XBRL> and </XBRL> tags.
3. Remove all tables with more than 15 percent numeric characters. Tables are initially identified using text enclosed between <TABLE> and </TABLE> tags.
4. Remove carriage returns.
5. Remove all markup tags (e.g., HTML).
6. Re-encode HTML entities. These are codes in HTML that correspond to certain symbols such as an ampersand (&).
7. Several other textual expressions are removed:
   a. Newline characters
   b. The token "and/or" is replace by "and or"
   c. Sequences of two or more hyphens, periods, or equal signs followed by spaces
   d. Underscore characters
   e. Sequences of three or more blanks are replace by a single blank
   f. Sequences of three or more newline characters, possibly followed by spaces, are replaced with two newline characters.
   g. Newline characters that are not preceded by a newline character and not followed by a space or a newline character are replaced with a single space.

Parsing of 10-K filings

To clean the text of 10-K filings from the SEC's EDGAR system, we follow the approaches of Li (2008) and Loughran and McDonald (2014a, 2014b).[19] Specifically, after downloading the raw text files from the EDGAR FTP site, we complete the following parsing steps:

1. Remove all documents in the 10-K filing except <DOCUMENT> <TYPE> 10-K and <DOCUMENT> <TYPE> EX-13 (Exhibit 13 is used for incorporating sections of the annual report to shareholders by references such as the MD&A).
2. Remove <DIV>, <TR>, <TD>, and <FONT> tags.
3. Remove all content between <XBRL> and </XBRL> tags.
4. Remove all tables with more than 15 percent numeric characters. Tables are initially identified using text enclosed between <TABLE> and </TABLE> tags.
5. Remove carriage returns.
6. Remove all markup tags (e.g., HTML).
7. Re-encode HTML entities. These are codes in HTML that correspond to certain symbols such as an ampersand (&).
8. Several other textual expressions are removed:
   a. Newline characters
   b. The token "and/or" is replace by "and or"
   c. Sequences of two or more hyphens, periods, or equal signs followed by spaces
   d. Underscore characters
   e. Sequences of three or more blanks are replace by a single blank
   f. Sequences of three or more newline characters, possibly followed by spaces, are replaced with two newline characters.
   g. Newline characters that are not preceded by a newline character and not followed by a space or a newline character are replaced with a single space.

Calculation of the readability measures

We implement the following procedures to calculate the various readability measures used in our paper.

Bog Index

We use the software package StyleWriter to calculate the Bog Index. StyleWriter is an add-on program for Microsoft Word, so we need to open each of the parsed 10-K filings separately in Microsoft Word, open StyleWriter, and complete the readability analysis of the text. To automate this process, we use a keyboard/mouse click macro to open and score each document. The Bog Index is based on the annual report portion (Form 10-K and Exhibit 13) of each parsed 10-K filing.

Fog Index

We calculate the Gunning (1952) Fog Index using Perl's Lingua:EN:Fathom package from the Comprehensive Perl Archive Network (CPAN). Fog is based on the annual report portion (Form 10-K and Exhibit 13) of each parsed 10-K filing.

LM PE Index

We follow the procedure used in Loughran and McDonald (2014b) to compute our measure of LM plain English readability. Specifically, using the parsed 10-K filing we compute six different components: average sentence length, average word length,

[19] Specifically, we follow the parsing outline provided on Bill McDonald's website, which can be accessed at http://www3.nd.edu/~mcdonald/Word_Lists_files/Documentation/Documentation_StageOne_10-X_Parse.pdf.

ratio of passive verbs to total words, ratio of legal terms to total words using legal words from Loughran and McDonald (2011), ratio of personal pronouns to total words, and ratio of other plain English violations (e.g., negative phrases, superfluous words) to total words. We standardize each of these components to have a mean of zero and standard deviation of one. Finally, we sum sentence length, word length, passive voice, legalese, and other and subtract personal pronouns to create the summary LM plain English readability index. Note that this approach creates a measure that increases in complexity (poorer readability) and is the equivalent of multiplying the Loughran and McDonald (2014b) measure by negative one. The LM plain English measure is based on the annual report portion (Form 10-K and Exhibit 13) of each parsed 10-K filing.

Total file size

We calculate file size based readability as the natural logarithm of the number of megabytes used by the entire 10-K filing on computer storage. Below we describe how we divide the raw filings into components for the calculation of component file size based readability.

Total words

We use the Lingua:EN:Fathom Perl package to count the number of words in the entire 10-K filing. We also use the same package to count the number of words in the 10-K text based on the parsing procedure described above and then separately on the remaining exhibits eliminating during the basic parsing procedure.

Breakdown of file size for components of the 10-K filings

We use Perl regular expressions to identify and separate and calculate the file size of different components of 10-K filings. Below we describe the process for measuring the file size of each file segment:

File size of text – To calculate the file size of the text of the 10-K filing, we use parsed 10-K filing described above, where the file size of text is based on the annual report portion (Form 10-K and Exhibit 13) of each parsed 10-K filing.

File Size of Exhibits – To calculate the file size of the text of the 10-K filing, we calculate the size of the additional 10-K filing documents appended to the 10-K filing. In particular, we extract the textual information (i.e., not pictures, PDFs, or XML), of those documents (i.e., text contained between <DOCUMENT> <TYPE> EX-## and </DOCUMENT>) to calculate the file size of exhibits.

File Size of Table Contents – To calculate the file size of the tables, we search for all text between <TABLE.*?> and </TABLE> in either the Form 10-K and Exhibit 13 (if available). After removing all HTML tags, the remaining content serves as the basis for calculating the file size of table contents.

File Size of Other Components – To calculate the file size of the other components contained in the 10-K filing we combine the file size of the pictures, PDFs, etc., HTML tags, and XML.

Pictures, PDFs, etc. - To calculate the size of the PDFs, we search all documents within the 10-K filing and locate any file name which includes ".pdf" or the tag <PDF>. We then extract the text of that document to calculate the file size. To calculate the size of pictures, we search for the file name extensions jpeg, jpg, tif, gif, and png to capture pictures included with filings. If we match any of these extensions, we extract the text of that document to calculate the file size of other components.

HTML – To calculate file size related to HTML, we locate all tags from anywhere in the 10-K filing and determine the file size of those tags.

XML – To calculate file size related to XML filings, we search for the file name extensions xbrl, xml, zip, xsd, js, css or for the document type (denoted by the <TYPE> tag) XBRL or XML to capture XML data. For filings after 2010, we also search for file name extensions of xls and xlsx. If we match any of these search criteria, we extract the text of that document to calculate the file size of XML.

Breakdown of word counts for components of the 10-K filings

We use Perl regular expressions to identify and separate the text of the original 10-K filing from other text in the filing related to components of the 10-K filing such as exhibit documents that are often appended to the filing.

10-K Text Words – To calculate the text attributable to words related to the original 10-K filing, we count the number of words appearing in the parsed 10-K filing that is comprised of Form 10-K itself and Exhibit 13 when applicable.

10-K Other Words – To calculate the other words contained in the filing, we start with the raw 10-K filing and drop <DOCUMENT> <TYPE> 10-K … </DOCUMENT> and <DOCUMENT> <TYPE> EX-13 … </DOCUMENT>. We then apply our parsing procedure to the remaining exhibits and use Perl's Lingua:EN:Fathom to calculate the number of words in these other exhibits.

Appendix D. Bog scores for multi-syllabic words from 10-Ks

This appendix provides a summary of Bog Index scoring for stem words with more than two syllables. Panel A provides a list of the 100 most frequently used words in 10-K filings. Panel B provides a list of 100 random words that score as the least precise (more abstract) words. The sample contains words from 66,173 10-K's filed between 1994 and 2011. The Bog Index scoring is based on word difficulty derived from the StyleWriter Plain English Software, where words are classified al words receive a score of 0, 1, 2, 3, or 4, where less familiar / precise (i.e., more abstract) words receive higher scores.

Panel A - Word Bog Scores for Most Frequently Used Words

	Word Bog Score		*Word Bog Score*
COMPANY	0	*AVERAGE*	0
FINANCIAL	0	*GENERALLY*	2
STATEMENT	0	*RECOGNIZE*	0
INTEREST	0	*TECHNOLOGY*	1
OPERATION	2	*EFFECTIVE*	1
BUSINESS	0	*REQUIREMENT*	2
AGREEMENT	1	*AVAILABLE*	1
CONSOLIDATE	2	*EQUIPMENT*	1
SECURITY	1	*REGULATION*	1
REVENUE	1	*DETERMINE*	2
MANAGEMENT	1	*MANUFACTURE*	2
OPERATE	2	*GENERAL*	0
CUSTOMER	0	*PROVISION*	1
PERIOD	0	*INDUSTRY*	0
APPROXIMATELY	2	*OBLIGATION*	2
ESTIMATE	0	*POLICY*	0
INVESTMENT	1	*DIVIDEND*	0
INFORMATION	1	*DEPOSIT*	0
CAPITAL	0	*INTERNAL*	0
SUBSIDIARY	1	*PURSUANT*	2
CONTINUE	0	*POSITION*	0
FACILITY	2	*ASSOCIATE*	0
DIRECTOR	0	*PRINCIPAL*	0
CORPORATION	1	*INSTRUMENT*	1
MATERIAL	1	*PRESIDENT*	0
INCORPORATE	1	*STOCKHOLDER*	0
ACQUISITION	2	*DISCLOSURE*	1
PRIMARILY	2	*VARIOUS*	0
BENEFIT	0	*COMMERCIAL*	1
DEVELOPMENT	1	*IMPAIRMENT*	2
REFERENCE	0	*PERFORMANCE*	1
ANNUAL	1	*ABILITY*	0
RESPECTIVELY	1	*COMMISSION*	0
CONDITION	1	*ALLOWANCE*	1
SIGNIFICANT	1	*LIABILITY*	1
COMPENSATION	1	*DEVELOP*	0
OFFICER	0	*MARKETING*	0
ADDITIONAL	2	*CURRENTLY*	0
EQUITY	1	*ESTABLISH*	2
ACTIVITY	2	*REPRESENT*	0
PROPERTY	1	*DISTRIBUTION*	1
REGISTRANT	1	*SHAREHOLDER*	0
EXHIBIT	2	*EXISTING*	0
ADDITION	1	*PRODUCTION*	1
LIABILITIES	1	*REGULATORY*	2
EXECUTIVE	0	*INVENTORY*	1
TRANSACTION	1	*APPLICATION*	1
INSURANCE	1	*PORTFOLIO*	1
OUTSTANDING	0	*EXERCISE*	0
FEDERAL	0	*EXPERIENCE*	0

List of most complex multi-syllabic words

This appendix provides a summary of Bog Index scoring for stem words with more than two syllables. Panel A provides a list of the 100 most frequently used words in 10-K filings. Panel B provides a list of 100 random words that score as the least precise (more abstract) words. The sample contains words from 66,173 10-K's filed between 1994 and 2011. The Bog Index scoring is based on word difficulty derived from the StyleWriter Plain English Software, where words are classified as Specialist, Heavy, or Abbreviations. All words receive a score of 0, 1, 2, 3, or 4, where less familiar / precise (i.e., more abstract) words receive higher scores.

Panel B - Word Bog Scores for Least Precise (Abstract) Words

	Word Bog Score		Word Bog Score
ALPINIST	4	*HIDDENITE*	4
ANAMORPHIC	4	*HYDROLYZE*	4
APERIODIC	4	*IMAGO*	4
ARAUCARIA	4	*IMMUNOCHEMICAL*	4
ARBORVITAE	4	*INHOMOGENEOUS*	4
ARCHAEAN	4	*INTEGUMENT*	4
ARIOSO	4	*IODIDE*	4
ARTERIOSCLEROSIS	4	*KINESTHETIC*	4
ATROPHIC	4	*LAPIDARY*	4
ATTENUATOR	4	*MACULA*	4
AXONAL	4	*MAXILLARY*	4
BIOTITIC	4	*METAPHYSICS*	4
CARBURIZE	4	*METASTASIS*	4
CENTAVO	4	*MICROCRYSTALLINE*	4
CHALYBEATE	4	*MUTAGENIC*	4
CHOLINESTERASE	4	*MYELITIS*	4
CHROMATIN	4	*NANOMETER*	4
CHROMATOGRAPHICALLY	4	*NEPHROPATHY*	4
CILIATED	4	*NUCLEOLUS*	4
COLEOPTERA	4	*ORGANISMAL*	4
COLLOIDAL	4	*OXYTOCIN*	4
COLORIMETRIC	4	*PARACLETE*	4
COLORIMETRY	4	*PAREGORIC*	4
CRYPTOGAM	4	*PATHOGENIC*	4
CYTOMEGALOVIRUS	4	*PHILOLOGY*	4
CYTOSINE	4	*PHOTOMETRY*	4
CYTOSKELETON	4	*PHYLLOXERA*	4
DEHISCENCE	4	*PNEUMONECTOMY*	4
DENATURANT	4	*PNEUMOTHORAX*	4
DENSITOMETER	4	*POLYCHROMIC*	4
DISAMBIGUATE	4	*POLYPLOIDY*	4
DISSONANCE	4	*POLYURIA*	4
ECTOPIC	4	*PREPUBERTAL*	4
ENDOSCOPY	4	*PSYCHOPHARMACOLOGICAL*	4
ENDOTHELIUM	4	*RABBINIC*	4
ENOLOGY	4	*RESORCINOL*	4
ENTERIC	4	*RETICULATE*	4
ENUCLEATION	4	*SCRIVENER*	4
ESCALADE	4	*SYNAPTIC*	4
ESOPHAGITIS	4	*SYNCHROTRON*	4
ESOPHAGOSCOPE	4	*SYNOVITIS*	4
EXACTA	4	*SYNTACTICAL*	4
EXTREMUM	4	*TANTALUM*	4
FIBRINOGEN	4	*TEGUMENT*	4
FLOCCULATE	4	*THROMBOCYTOPENIA*	4
FUSIBLE	4	*TRANSCRIPTASE*	4
GENITOR	4	*TRIGEMINAL*	4
GEODESIC	4	*TRIGLYCERIDE*	4
HAFNIUM	4	*UNMYELINATED*	4
HETEROCYCLIC	4	*VASOCONSTRICTOR*	4

Appendix E. Components of Bog Index for 10-K filings

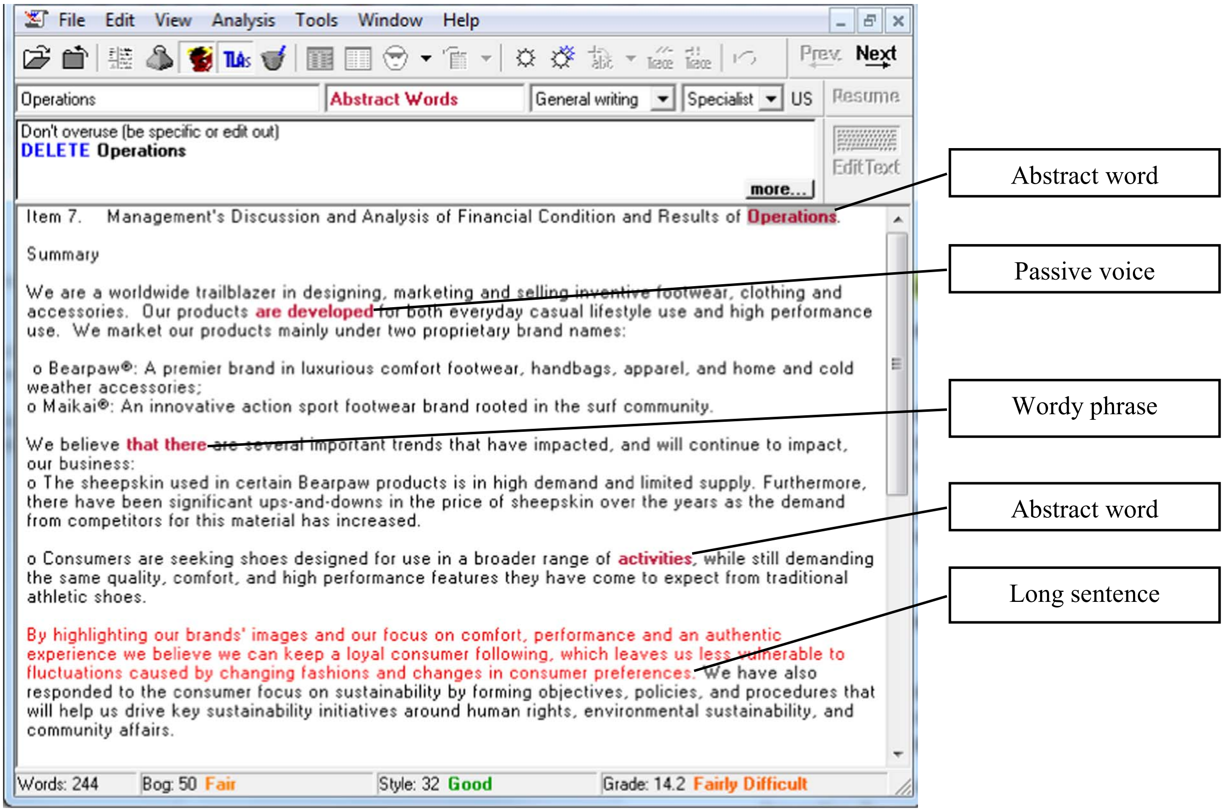
This appendix provides a breakdown of the underlying components that are used to create the Bog Index. Each underlying component is categorized under its associated summary component category. To facilitate comparisons of the estimated coefficients, the OLS regression results present explanatory variables that have been standardized to have a mean of zero and a standard deviation of one. It is important to note that this appendix presents the raw underlying writing style components, which do not necessarily correspond direct with some of the transformed variables used to create the actual Bog Index. All '%' variables are computed as the number of style issues divided by the total number of words in the document.

	Pred. Sign	Bog Index
Constant:	+	81.637
		(3671.61)
Sentence Bog Factors:		
Avg. Sentence Length	+	3.172
		(50.00)
Word Bog Factors:		
% Passive Verbs	+	0.982
		(43.96)
% Hidden Verbs	+	0.455
		(20.59)
% Complex Words	+	1.614
		(62.71)
% Abstract Words	+	1.145
		(43.65)
% Overused Words	+	0.451
		(17.55)
% Legal Words	+	0.450
		(20.19)
% Cliches	+	0.046
		(2.55)
% Business Cliches	+	0.080
		(3.96)
% Wordy Phrases	+	0.799
		(26.13)
% Difficult Words	+	0.166
		(4.45)
% Overwriting Phrases	+	0.519
		(23.59)
% Foreign Words	+	-0.031
		(-1.30)
% Unusual Words	+	0.213
		(8.88)
% Abbreviations	+	2.409
		(89.59)
% Jargon Phrases	+	1.961
		(49.72)
% Specialist Terms	+	1.671
		(39.12)
Pep Factors:		
% Names	–	-0.896
		(-35.60)
% Interesting Words	–	-0.461
		(-16.54)
% Conversational Expressions	–	-0.718
		(-35.46)
Std. Dev. of Sent. Length / Avg. Sent. Length	–	-0.982
		(-37.54)
# of OBS		66,173
R^2		0.9201

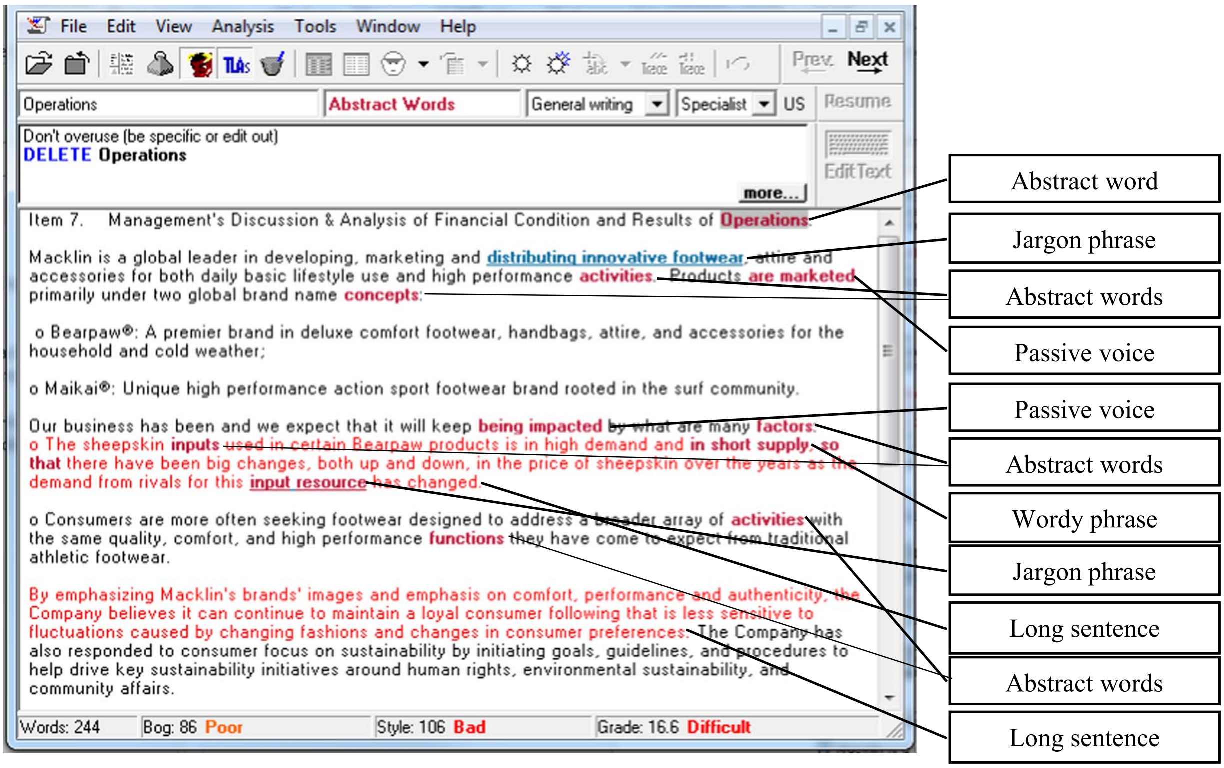
Appendix F. More and less readable conditions related to experimental evidence provided in Table 1

This appendix presents both the more (Panel A) and less (Panel B) readable versions of the disclosures that we provide to participants in our experimental validation of the Bog Index. Participants are randomly assigned to receive one of the two disclosures, and then rate them on their readability. Both disclosures are of equal length (244 words), are held constant in terms of their formatting, and are roughly equivalent on the Fog Index (15.83 for the more readable condition vs. 15.70 for the less readable condition – which actually indicates that our more readable plain English condition fared slightly worse on the Fog Index). The more readable version scores lower on the Bog Index (50) than the less readable version (86) suggesting that it is more readable across the plain English attributes captured by the Bog Index. Panels A and B also indicate examples of some of the linguistic characteristics (e.g., abstract words, jargon phrases, wordy phrases, passive voice, and long sentences) that lead to a lower Bog Index.

Panel A. More Readable Condition



Panel B. Less Readable Condition



Appendix G. Examples of more and less readable excerpts related to quasi-experimental evidence provided in Table 2

This appendix presents examples of one of the more readable (Panel A) and less readable (Panel B) excerpts from real companies that we present to participants in our quasi-experiment, with evidence presented in Table 2. We hand-collect excerpts of approximately 200 words from the beginning of the Letter to Shareholders for all 2014 Fortune 100 firms. We then measure the Bog Index for each excerpt in our collected sample, and keep the three most-readable disclosures (Wells Fargo, Kroger, and PepsiCo) and three least-readable disclosures (Phillips 66, International FC Stone, and United Technologies) for our study. Company names are disguised in all disclosures (e.g., replaced with generic names like Gamma, Theta, etc.) to reduce the likelihood that participants' responses will be affected by their knowledge or perceptions of the actual firms, as opposed to disclosure characteristics. Each participant is randomly assigned to read both one of the more readable disclosures, and one of the less readable disclosures and then, as in the first experiment, rate, "How difficult or easy is it to read the information in the disclosure above?" with responses ranging from 1 = very difficult to 9 = very easy.

Panel A More Readable Example – Wells Fargo

To Our Owners –

I have always believed that culture is the most important part of a company's success. It is the heart of any organization and a significant contributor to long-term performance and stability.

This is certainly true for Gamma, Inc. Since 1852, culture has been a focus of ours, beginning with how we served the Gold Rush-era customers who trusted us to transport their money and valuables on our stagecoaches. In those early days, we held a belief that still holds true today: "Our merchandise is courtesy, willingness, and human ability." Today, I sum up Gamma's culture with this word: "Relationships." It captures the passion we all share for serving our key stakeholders — customers, communities, investors, and team members.

To earn their trust, we strive to do the right thing and act under the highest ethical standards where honesty, trust, and integrity matter. Cultures take years to establish and mature, a lesson I learned while growing up on a family farm in a small town in central Minnesota. Those years taught me that the best harvests come only after years of thoughtful planning, planting, and nurturing. It's no different at Gamma, Inc. The culture our people enjoy today is the result of those who served before us — through a civil war, two world wars, the Great Depression, and a Great Recession that remains fresh in our memories.

Panel B Less Readable Example – Phillips 66

Dear Stakeholders –

As a new company with a rich history, Theta Company has an unparalleled foundation for success. We are starting off our life as an advantaged downstream company with three winning businesses, a strong heritage and enduring commitment to responsible and reliable operations, safety and care for the environment.

Among our attributes are an experienced leadership team, a talented global work force, a dedication to operating excellence and a strong balance sheet. Through these capabilities, Theta intends to enhance and grow our business and return value to shareholders, while maintaining as our highest priorities the operating integrity, safety and environmental performance of our operations. We believe true sustainability requires that we reduce our impact on the environment and contribute to the wellbeing of society. Theta is dedicated to enhancing the sustainability of the communities in which we operate. We believe that supporting healthy local communities provides residents with a high-quality of life and enhances our own ability to succeed.

At Theta Company we recognize that stakeholders have different perspectives, and we seek to establish positive and productive engagement to hear their views, help us find common ground and identify viable solutions. For example, at most of our refineries, we have Community Advisory Panels or CAPs to facilitate closer communication and understanding.

Appendix H. Illustration of 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)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.

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.

References

Averett, W., 2013. AICPA Launces Center for Plain English Accounting. ⟨http://warrenaverett.com/2013/12/02/aicpa-launches-center-for-plain-english-accounting/⟩.
Biddle, G., Hilary, G., Verdi, R., 2009. How does financial reporting quality relate to investment efficiency? J. Account. Econ. 48, 112–131.
Bloomfield, R., 2002. The "incomplete revelation hypothesis" and financial reporting. Account. Horiz. 16, 233–243.
Bonsall, S., Miller, B., 2017. The impact of narrative financial disclosure complexity on bond ratings and the cost of debt capital. Rev. Account. Stud. (forthcoming).
Bozanic, Z., Thevenot, M., 2015. Qualitative disclosure and changes in sell-side financial analysts' information environment. Contemp. Account. Res. 32 (4), 1595–1616.
Core, J., 2001. A review of the empirical disclosure literature: discussion. J. Account. Econ. 31, 441–456.
De Franco, G., Hope, O.K., Vyas, D., Zhou, Y., 2015. Analyst report readability. Contemp. Account. Res. 32 (1), 76–104.
DuBay, W., 2004. The Principles of Readability. Impact Information, Costa Mesa, CA.
Eagleson, R., 2014. Short Definition of Plain Language. ⟨http://www.plainlanguage.gov/whatisPL/definitions/eagleson.cfm⟩.
Guay, W., Samuels, D., Taylor, D., 2016. Guiding through the fog: financial statement complexity and voluntary disclosure. J. Account. Econ. 62 (2-3), 234–269.
Gunning, R., 1952. The Technique of Clear Writing. McGraw-Hill International Book Co., New York, NY.
Henry, E., Leone, A.J., 2016. Measuring qualitative information in capital markets research: comparison of alternative methodologies to measure disclosure tone. Account. Rev. 91 (1), 153–178.
Horton, J.J., Rand, D.G., Zeckhauser, R.J., 2011. The online laboratory: conducting experiments in a real labor market. Exp. Econ. 14, 399–425.
Koonce, L., Miller, J., Winchel, J., 2015. The effects of norms on investor reactions to derivate use. Contemp. Account. Res. 32 (4), 1529–1554.
Krische, S., 2015. The impact of individual investors' financial literacy on assessments of conflicts of interest. Working paper.
Lang, M., Stice-Lawrence, L., 2015. Textual analysis and international financial reporting: large sample evidence. J. Account. Econ. 60 (2-3), 110–135.
Lawrence, A., 2013. Individual investors and financial disclosure. J. Account. Econ. *56*, *130–147*.
Lehavy, R., Li, F., Merkley, K., 2011. The effect of annual report readability on analyst following and the properties of their earnings forecasts. Account. Rev. 85 (6), 2107–2143.
Li, F., 2008. Annual report readability, current earnings, and earnings persistence. J. Account. Econ., 221–247.
Li, F., 2010. Textual analysis of corporate disclosures: a survey of the literature. J. Account. Lit. 29, 143–165.
Libby, R., Bloomfield, R., Nelson, M.W., 2002. Experimental research in financial accounting. Account. Organ. Soc. 27, 775–810.
Loughran, T., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. J. Financ. 66, 35–65.
Loughran, T., McDonald, B., 2014a. Measuring readability in financial disclosures. J. Financ. 69 (4), 1643–1671.
Loughran, T., McDonald, B., 2014b. Regulation and financial disclosure: the impact of plain English. J. Regul. Econ. 45 (1), 94–113.
Miller, B.P., 2010. The effects of disclosure complexity on small and large investor trading. Account. Rev. 85 (6), 2107–2143.
Paolacci, G., Chandler, J., Ipeirotis, P.G., 2010. Running experiments on amazon mechanical turk. Judgm. Decis. Mak. 5 (5), 411–419.
Rennekamp, K., 2012. Processing fluency and investors' reactions to disclosure readability. J. Account. Res. 50 (5), 1319–1354.
Securities and Exchange Commission, 1998a. Staff Legal Bulletin No. 7. ⟨https://www.sec.gov/interps/legal/slbcf7.htm⟩.
Securities and Exchange Commission, 1998b. A Plain English Handbook: How to Create Clear SEC Disclosure. SEC Office of Investor Education and Assistance Washington, District of Columbia ⟨http://www.sec.gov/pdf/handbook.pdf⟩.
Whiteman, L., 2000. Signs of intelligible life. Sci. Careers. ⟨http://sciencecareers.sciencemag.org/career_magazine/previous_issues/articles/2000_11_24/nodoi.3947963619245338704⟩.
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., Zhang, X., 2009. Financial disclosure complexity and investor underreaction to 10-K information. Rev. Account. Stud. 14 (4), 559–586.