Measuring Readability in Financial Disclosures
Author(s): TIM LOUGHRAN and BILL MCDONALD
Source: *The Journal of Finance*, Vol. 69, No. 4 (AUGUST 2014), pp. 1643-1671
Published by: Wiley for the American Finance Association
Stable URL: https://www.jstor.org/stable/43611199
Accessed: 22-11-2018 21:50 UTC

THE JOURNAL OF FINANCE • VOL. LXIX, NO. 4 • AUGUST 2014

Measuring Readability in Financial Disclosures

TIM LOUGHRAN and BILL MCDONALD*

ABSTRACT

Defining and measuring readability in the context of financial disclosures becomes important with the increasing use of textual analysis and the Securities and Exchange Commission's plain English initiative. We propose defining readability as the effective communication of valuation-relevant information. The Fog Index—the most commonly applied readability measure—is shown to be poorly specified in financial applications. Of Fog's two components, one is misspecified and the other is difficult to measure. We report that 10-K document file size provides a simple readability proxy that outperforms the Fog Index, does not require document parsing, facilitates replication, and is correlated with alternative readability constructs.

"Just as the Black-Scholes model is a commonplace when it comes to compliance with the stock option compensation rules, we may soon be looking to the Gunning-Fog and Flesch-Kincaid models to judge the level of compliance with the plain English rules."

—SEC Chairman Christopher Cox, speech at USC Marshall School of Business, March 23, 2007

MANAGERS OF PUBLICLY traded firms are required to produce public documents that provide a comprehensive review of the firm's business operations and financial condition. An important financial disclosure document created by managers to communicate with investors and analysts is the annual report filed pursuant to the Securities Exchange Act of 1934, Form 10-K. Both financial researchers and government regulators have struggled with the notion of how to define and measure the readability of mandated disclosures.

What is meant by "readability" is difficult to define precisely and its measure has evolved predominantly in the process of grade-leveling school textbooks, insurance contracts, and the understandability of instructions in military applications. In the accounting and finance literature, trending with the recent increase in text-based analysis, researchers often use the Fog Index

*Tim Loughran and Bill McDonald are at the Mendoza College of Business, University of Notre Dame. We thank Jeff Burks, Peter Easton, Paul Gao, Campbell Harvey (Editor), Stephannie Larocque, Jennifer Marietta-Westberg, Richard Mendenhall, Paul Tetlock, two anonymous referees, an anonymous Associate Editor, and seminar participants at Michigan State University, Rice University, and The University of Notre Dame for helpful comments. We are grateful to Jianfeng Zhu and Manisha Goswami for research assistance.

DOI: 10.1111/jofi.12162

as a measure of document readability. The Fog Index, also sometimes labeled Gunning-Fog after its founder, is defined as a linear combination of average sentence length and the proportion of complex words (words with more than two syllables).[1] Recent examples of papers using the Fog Index include Li (2008), Biddle, Hilary, and Verdi (2009), Miller (2010), Lehavy, Li, and Merkley (2011), Dougal et al. (2012), and Lawrence (2013). Biddle, Hilary, and Verdi (2009) go as far as to define the Fog Index as "a measure of financial statement readability."

Even the Securities and Exchange Commission (SEC) contemplated elevating the importance of traditional readability measures like the Fog Index for company filings. As noted in the quote above, SEC Chairman Christopher Cox suggests that the Fog Index can be used to gauge compliance with the SEC's plain English initiatives. Although traditional readability measures such as the Fog Index are, at first glance, conceptually appealing, to what extent do they measure effective communication of information used in valuing a firm's stock and estimating its earnings? In the context of financial disclosures, we believe that the goal of "readability" should be the effective communication of valuation-relevant information, whether it is directly interpreted by individual investors or assimilated and distributed by professional analysts.

Because the notion of readability, as highlighted by the SEC's plain English initiative of 1998, focuses on the general public (versus analysts), our initial results consider post-filing stock return volatility in the month following the 10-K filing date as a broad-based measure of the information environment. Market-based measures like subsequent volatility have the additional advantage of more comprehensive samples since data screens for earnings are not needed. Our assumption is that better written documents produce less ambiguity in valuation, as reflected by the lower price volatility of the stock in the period immediately following the filing (controlling for other variables, including the historical level of volatility). Since analysts use 10-Ks to assess the firm's current and future operations, for robustness, we also consider analyst forecast error (i.e., standardized unexpected earnings (SUEs)) and analyst forecast dispersion as measures of the information environment.

Our paper makes several contributions. First, we show that traditional readability measures like the Fog Index are poorly specified when used to evaluate financial documents. By its very nature, business text has an extremely high percentage of complex words—one of Fog's two components—that are well understood by investors and analysts. Second, we recommend using the file size of the 10-K as an easily calculated proxy for document readability. The proposed measure is straightforward, is substantially less prone to measurement

[1] Two other popular readability measures, the Flesch–Kincaid measure and the Flesch Reading Ease Score, use the same two components as Fog, but, rather than create a binary classification of complex words based on syllables, an explicit count of syllables is used. Both the Fog Index and Flesch Reading Ease Score are scaled combinations that produce numeric estimates of grade level, while the Flesch–Kincaid measure creates a 0–100 scaled measure. The correlation in our sample between the binary and integer measures of syllables is 0.96. Because the Fog Index is more commonly used, we focus on this measure in our paper; however, the Flesch variants are clearly very similar.

error, is easily replicated, is strongly correlated with alternative readability measures, and, based on results, appears to better gauge how effectively managers convey valuation-relevant information to investors and analysts.

Hundreds of readability measures exist, having evolved from the early development of these formulas in the 1930s.[2] We focus on the Fog Index because it is one of the most popular readability measures across all fields and it also appears to be the measure of choice in financial research in particular. Through a series of tests, we provide evidence that the Fog Index is not an appropriate measure of readability in financial documents. The first component of the Fog Index, "average words per sentence," provides reasonable empirical correlations with other measures of readability. However, measuring sentence length in the context of financial disclosures is substantially less precise than measuring sentence length in traditional prose.

More importantly, we show that the second component in the Fog Index, "complex words," is a poorly specified measure in business documents. The Fog Index indicates that an increase in the number of complex words (more than two syllables) decreases readability, with this factor accounting for half of the measure's inputs. Business text, however, commonly contains multisyllable words used to describe operations. Words like *corporation*, *company*, *agreement*, *management*, and *operations* are predominant complex words occurring in 10-Ks, yet are presumably easy for investors to comprehend. One of the longest words occurring with reasonable frequency in 10-Ks is *telecommunications*, a word not likely to force most readers to consult their dictionaries.

The frequency count for any reasonable subset of words is characterized by Zipf's law, the widely documented empirical result whereby the frequency of any word is approximately inversely proportional to its rank in the frequency table, that is, a very small number of words will dominate the frequency counts for a given set of words. In our case, 52 complex words out of more than 45,000 complex words appearing in the 10-K sample account for more than 25% of the complex word count. More importantly, virtually all of these words are simple, common business terms.

We show that, based on the frequency of occurrence, all of the top quartile of multisyllable words would likely be known to a typical investor or analyst reading a 10-K. Even if we ignore the most common multisyllable words, few of the remaining complex words are ones that an average reader would stumble over. Our evidence shows that syllable counts are a poor measure of readability in the context of firms' business disclosures.

Consistent with the premise that complex words merely add measurement error, we find that the Fog Index has insignificant predictive power in explaining both unexpected earnings and analyst dispersion. Interestingly, although the Fog Index is linked with analyst dispersion during 1995 to 2006, as shown by Lehavy, Li, and Merkley (2011), we find that this relation does not persist when the time period is expanded to 1994 to 2011.

[2] In an early landmark study, Gray and Leary (1935) consider 228 elements that affect readability. DuBay (2007) provides a useful history of the literature on readability.

Having identified problems with the Fog Index, a natural question to ask is whether other measures can capture the readability of financial documents. We propose using the file size of the 10-K complete submission text file as a readability measure. The 10-K file size is exceptionally easy to determine and is not prone to the substantial measurement errors of other textual procedures requiring parsing of the 10-K documents. Avoiding the imprecision of parsing algorithms has the additional advantage of facilitating replication. More importantly, as shown in our empirical results, the simple measure of file size compares well with our measures of the information environment and is highly correlated with alternative measures of readability. Writing style—the central focus of the Fog Index—is but one dimension of readability and is an aspect less differentiated in financial documents (versus books from various grade levels). File size can be viewed as an omnibus measure capturing the many dimensions of readability.

We appreciate that file size of a 10-K is a gross proxy for readability and one can imagine many exceptions to the inverse relation between 10-K size and readability as we have defined it in the context of valuation. We find, however, that the relation between 10-K file size and both market and analyst measures of valuation ambiguity are consistent with our definition of readability. We argue that, if firms are trying to obscure mandated earnings-relevant information, they are less likely to use sesquipedalian words or complex rhetoric, and more likely to bury the results in longer documents.[3] Additionally, litigation risk will incent managers to disgorge information whether it is useful or not. Of course, to some extent document size can be a simple artifact of the firm's structural complexity. In discussing the comprehensibility of 10-Ks, You and Zhang (2009) refer to complexity of the document, which is presumably linked to firm complexity. Although we consider empirical tests that attempt to control for firm complexity using business segment data, ultimately we do not believe that these two factors—readability and firm complexity—can be entirely disentangled.

As the average 10-K contains more than 38,000 words, investors and analysts must read through hundreds of 10-K pages while gathering information to enhance their valuation estimates of the firm. We find that larger 10-Ks are significantly associated with high return volatility, earnings forecast errors, and earnings forecast dispersion, after controlling for other variables such as firm size, book-to-market, past volatility, industry effects, and prior stock performance.[4]

[3] We informally polled a small sample of partners of major accounting firms and asked how they would legally attempt to obscure information whose disclosure was required. The accountants immediately identified the strategy of burying the awkward revelation in an overwhelming amount of uninformative text and data.

[4] The positive relation between file size and volatility, earnings surprises, and analyst dispersion is inconsistent with the alternative explanation that file size is a proxy for disclosure (see Leuz and Schrand (2009)). If file size proxies for firm disclosure, one would expect larger documents to be negatively (not positively) related to volatility and analyst dispersion.

The paper proceeds as follows. Section I defines the Fog Index, reviews the finance and accounting literature relating to measuring readability, and develops our definition of readability in financial disclosures. In Section II, we describe how 10-Ks are parsed, discuss the data sources for other variables included in the study, document the sample formation process, and provide descriptive results. Section III reports the initial regression results, while Section IV considers alternative measures of the information environment. Section V provides additional evidence on different measures of readability and robustness tests. Section VI concludes and discusses arguments for using file size as a proxy for readability.

I. Background

A. The Fog Index

First published in Gunning (1952), the Fog Index's popularity is primarily attributable to its ease of calculation and adaptability to computational measure. The Fog Index is a simple function of two variables: average sentence length (in words) and complex words, defined as the percentage of words with more than two syllables. As is common with many readability measures, the two factors are combined in a manner that is intended to predict grade level:

$$\begin{aligned} \textit{Fog index} = 0.4(&\text{average number of words per sentence} \\ &+ \text{percent of complex words}). \end{aligned} \tag{1}$$

Lower values of the Fog Index indicate more readable text. Our paper is not the first to criticize the application of the Fog Index to business text. In an early survey paper, Jones and Shoemaker (1994, p. 172) state that traditional tests of reading difficulty "are dubious instruments for adequately assessing the readability of accounting narratives that are adult oriented and specialist in nature." The authors, however, do not empirically identify the weakness of the Fog Index nor do they offer an alternative readability measure for researchers to apply to business text.

B. Related Literature

Background literature for our paper comes from two areas of recent research: textual analysis and the use of readability measures in accounting and finance. The first area, textual analysis, has examined the tone or content of popular newspaper columns (Tetlock (2007) and Dougal et al. (2012)), internet message board postings (Antweiler and Frank (2004)), 10-Ks (Li (2008), You and Zhang (2009), Jegadeesh and Wu (2013), and Lawrence (2013)), and newspaper articles (Tetlock, Saar-Tsechansky, and Macskassy (2008)). The recent surge in textual research is an artifact of better computer technology allowing massive text collections to be parsed, methodological advances attributable to research in web-based text search, and availability of large textual corpuses such as the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) database.

The growth in textual research makes measuring readability an important tool in assessing financial documents. A number of recent papers use the Fog Index or number of words as readability/complexity measures. In his examination of the link between 10-K readability and earnings persistence, Li (2008) uses both the Fog Index and a simple word count to assess readability. He finds that 10-Ks with higher Fog Index values (less readable text) and longer document length have lower subsequent earnings. The evidence in Li (2008) suggests that firm managers may try to hide poor future earnings from investors by increasing the complexity of their written documents.[5] In contrast, our paper examines the incorporation of 10-K information into the current stock price. Greater difficulty in properly valuing the firm should lead to higher short-term volatility. Our paper does not examine whether 10-K readability is linked with higher or lower subsequent earnings.

The extant literature provides evidence that investors are affected by 10-K readability. For example, using U.S. discount-brokerage data, Lawrence (2013) finds that retail investors are more likely to invest in firms with shorter, more readable 10-Ks (as measured by the Fog Index), while Miller (2010) reports that firms with better written documents (using the Fog Index as one of his readability measures) have more pronounced small investor trading activity around the filing date.

The readability of analyst reports also seems to affect trading volume. De Franco et al. (2013) document a positive relation between analyst report readability (proxied by a combination of three traditional readability measures: Fog, Flesch-Kincaid, and Flesch Reading Ease) and the trading volume reaction to the reports. Dougal et al. (2012) use the Fog Index as a control for the readability of the "Abreast of the Market" column. You and Zhang (2009), using the number of 10-K words as a measure of firm complexity, find that firms above the annual median word count have a delayed stock market reaction over the following 12 months.

In related work that also links 10-K readability to analyst coverage and dispersion, Lehavy, Li, and Merkley (2011) find that 10-Ks with less readable text (as measured by the Fog Index) have more analysts following the stock, greater analyst dispersion, and lower accuracy. The authors argue that, as the processing cost increases for 10-Ks with less readable text, more analysts are needed to cover the stock to meet investors' demand for information. They find that 10-Ks with higher Fog Index values are associated with larger levels of analyst dispersion. While they focus on the relation between readability and earnings forecasts, we focus on the methodological issue of how to best measure readability.

[5] Bloomfield (2008), however, notes that there are potentially many explanations for why firms might produce longer and more complex documents. While, in some cases, the intent might be to somehow diffuse bad news ("obfuscation," "attribution," or "misdirection"), some firm events could simply require longer and more detailed explanation.

C. Defining Readability

Readability is not a precisely defined construct and the preferred measure of readability in a specific application depends critically on how the concept is defined. As noted by DuBay (2004, p. 3), the readability definition offered by Klare (1963)—"the ease of understanding or comprehension due to the style of writing"—tends to focus on writing style versus content, coherence, and organization. Measures cast from this perspective, such as the Fog Index, focus primarily on sentence length and polysyllabic words and work well in the context of grade-leveling texts where these two factors are clearly distinguishing features. For example, if one's goal is to compare *The Cat in the Hat* with *The Iliad*, sentence length and word complexity are, in this case, useful discriminators.

Other authors in the readability literature broaden the concept to emphasize the importance of the targeted audience in determining readability. For example, McLaughlin (1969; in DuBay (2004)) defines readability as "the degree to which a given class of people find certain reading matter compelling and comprehensible," and Davison and Kantor (1982, p. 187) emphasize that the "background knowledge assumed in the reader" is more important than "trying to make a text fit a level of readability defined by a formula." Given the targeted audience for financial disclosures, it is not clear whether the short sentences of Hemingway would be more effective than the long sentences of Faulkner.

Because most financial documents are not differentiable based on their use of ponderous, polysyllabic words, and heterogeneous writing styles, we anchor our definition of readability in broader definitions from the literature. For example, the definition of Dale and Chall (1948), referred to by Tekfi (1987, p. 262) as the classic definition and by DuBay (2007) as the most comprehensive definition, specifies that "In the broadest sense, readability is the sum total (including interactions) of all the elements with a given piece of printed material that affects the success which a group of readers have with it." Similarly, Tekfi (1987, p. 262) concludes that readability is "ensuring that a given piece of writing reaches and affects its audience in the way the author intends." In our application, we define readability as the ability of individual investors and analysts to assimilate valuation-relevant information from a financial disclosure. Thus, our proposed measure of readability, file size, is cast in a broader context than the two focal points of the Fog Index.

II. Data

A. 10-K Parsing Procedure

Our starting point is downloading all 10-K documents available on EDGAR during the 1994 to 2011 period.[6] We follow the parsing procedure prescribed in Loughran and McDonald (2011) with two minor exceptions. First, Loughran

[6] Our initial sample includes all 10-K, 10-K405, 10KSB, and 10KSB40 filings. We use the term "10-K" to refer to all of these variants. We do not include amended filings in the sample.

and McDonald remove all tables contained in a filing. Tables are identified by HTML tags. In some cases, the filers use table tags simply to identify paragraphs; thus, Loughran and McDonald examine the text in each table segment and exclude only those tables where more than 25% of the nonblank characters are numbers. We find that a lower threshold of 10% is more consistent in correctly identifying tabular data. Second, we remove company names in the parsing process.

To parse for sentences, we first remove abbreviations, headings, and numbers, and then assume that the remaining periods are sentence terminations. Average of words per sentence is then the number of words tabulated in the document divided by the number of sentence terminations. Alternatively, a regular expression can be used to identify sentences but the outcome from this process tends to provide more volatile estimates.

Although our criticism of the Fog Index as a metric of readability is not centered on measuring sentence length, note that this measure adds an additional source of noise when applied to financial filings. Identifying sentences is an imprecise process where the determination of a sentence relies critically on accurate disambiguation of sentence boundaries, which, in turn, requires disambiguation of capitalized words and abbreviations. In the computational linguistics literature, Palmer and Hearst (1994) and Mikheev (2002) provide a useful discussion highlighting the difficulties in parsing a document for sentences. Measuring sentence length in a standard novel is relatively accurate, where the text format of traditional prose and its presentation structure are essentially one dimensional. Financial disclosures are replete with itemized lists, headings, nonstandard methods for structuring the document (especially before 2002 when the use of HTML was less common), and myriad abbreviations. Thus, correctly identifying sentences in this context is challenging and parsing errors can have a disproportionately large impact on the final estimate of average words per sentence.

While some research focuses solely on the management discussion and analysis (MD&A) section of the 10-K (see Feldman et al. (2010) and Li (2010)), we analyze the entire 10-K document to assess readability.[7] In measuring sentiment, Loughran and McDonald (2011) report that focusing on the MD&A section does not provide more powerful statistical tests. In the current study, we assume that investors' assessment of firm valuation goes well beyond Section 7 of the 10-K filing. We focus on 10-Ks and not 10-Qs because 10-Ks are more informative to investors. For example, Griffin (2003) reports stronger market responses to 10-Ks than 10-Q filings. Typically, 10-Qs are much shorter in length and report unaudited financial statements.

Form 10-K filing day returns are not considered, since it is impossible to separate the effect of new information from the comprehensibility of information when it is first released. That is, when the information is announced, it is impossible to separate the signal (i.e., the information contained in the filing) from noise (i.e., the accessibility of the information via readability). In a

[7] In an early survey of the literature, Jones and Shoemaker (1994) find mixed empirical evidence on whether certain sections of the 10-K differ in terms of readability.

Table I
Sample Creation

This table reports the impact of various data filters on the initial 10-K sample. CRSP PERMNO is the permanent issue identification number assigned by the Center for Research in Security Prices, RMSE is the root mean square error from a market model regression for days [6, 28] following the 10-K filing, and CIK is the Central Index Key assigned by the SEC.

	Dropped	Sample Size
SEC 10-K files 1994 to 2011		188,413
Eliminate duplicates within year/CIK	2,930	185,483
Drop if file date < 180 days from prior filing	585	184,898
Drop if number of words < 2,000	8,298	176,600
CRSP PERMNO match	88,800	87,800
Reported on CRSP as ordinary common equity	4,376	83,424
Price on filing date minus one ≥ $3	13,338	70,086
Book-to-market COMPUSTAT data available and book value > 0	2,874	67,212
Post-filing date market model RMSE for days [6, 28]	346	66,866
At least 60 days' data available for market model estimates from event days [−252, −6]	147	66,719
Returns for days 0–1 in event period	12	66,707

competitive market, investors will respond to information conditional on its ambiguity. We thus expect the firm's stock to immediately incorporate information conditional on its comprehensibility, with subsequent stock volatility reflecting any ambiguity in the information.

B. Sample Creation

Table I documents the sample formation process. We start with a total of 188,413 10-K firm-year observations from EDGAR during 1994 to 2011.[8] In total, 10 different data screens are applied to the initial 10-K sample. For example, we require that the firm have a CRSP Permanent ID match (dropping 88,800 observations), be ordinary common stock according to CRSP (removing 4,376 observations), have a stock price of at least $3 to minimize market microstructure effects (losing 13,338 observations), and have at least 2,000 words in the 10-K (removing 8,298 observations).[9] Our final sample used in the initial tests consists of 66,707 observations. When we subsequently examine earnings-related dependent variables and business segment data, we note the change in sample size.

To test the robustness of our initial results based on market-based measures, we also consider both SUEs and analyst dispersion as measures of the communication effectiveness of valuation information. For these subsamples, we

[8] A data set with the SEC's Central Index Key (CIK) number, filing date, form type, and file size for all 10-K filings is provided at http://www.nd.edu/~mcdonald/Word_Lists.html.

[9] We require at least 2,000 words in the 10-K to eliminate filings that merely mention why the firm is not filing a full 10-K at that point in time. Li (2008) requires firms to have at least 3,000 words to enter his 10-K sample.

Table II

Variable Means by Time Period, 1994 to 2011

See the Appendix for detailed variable definitions. In subsequent regressions, a logarithmic transformation is used for *File size, Size,* and *Book-to-market.* For *Abs(Sue)*, there are 28,434 observations, with 12,390 for the first subperiod and 16,044 for the second subperiod. For *Analyst dispersion* and *# of analysts*, the total sample size is 17,960 with 6,985 observations for the first subperiod and 10,975 for the second subperiod.

Variable	(1) 1994 to 2002	(2) 2003 to 2011	(3) 1994 to 2011
Readability measures:			
Fog index	18.44	18.94	18.68
Average words per sentence	22.82	23.27	23.04
Percent complex words	23.28%	24.09%	23.67%
File size (in megabytes)	0.42	2.51	1.43
Dependent variables:			
Post-filing RMSE	3.45	2.26	2.87
Abs(Sue)	0.27	0.39	0.34
Analyst dispersion	0.14	0.21	0.19
Control variables:			
Pre-filing alpha	0.08	0.05	0.06
Pre-filing RMSE	3.54	2.65	3.11
Abs(filing period abnormal return)	0.04	0.03	0.03
Size (market capitalization) in $ millions	$2,257.56	$3,680.13	$2,946.42
Book-to-market	0.66	0.67	0.66
NASDAQ dummy	0.60	0.58	0.59
# of analysts	4.19	5.60	5.05
Number of observations	34,405	32,302	66,707

require two or more analyst forecasts from I/B/E/S in the period between the 10-K filing date and the firm's next quarterly earnings announcement. To clarify, consider by way of example Google's 10-K for the fiscal year ending December 31, 2007 and filed on February 15, 2008. Google's first earnings announcement (first quarter 2008 earnings results) following the 10-K filing date was on April 17, 2008. A total of 17 different analysts initiated or updated their first quarter earnings forecast for Google between the file date of February 15 and the earnings announcement date of April 17, 2008. If a given analyst makes more than one forecast in this time window, we use the forecast closest to the filing date for that analyst. Forecasts from these 17 analysts are used to create the earnings expectation and analyst dispersion variable. Overall, our earnings-based measures should capture the valuation-relevant information gained by analysts from reading the 10-K in estimating the next quarter's earnings. Controlling for other factors, better written 10-Ks should have smaller absolute SUE and analyst dispersion values.

C. Descriptive Results

Mean summary statistics for the sample variables are reported in Table II. The first two columns divide the sample period in half while the last column of the table lists the averages for the entire period. The average values for the Fog Index, average words per sentence, and percent of complex words are similar between the 1994 to 2002 and 2003 to 2011 periods. For example, the Fog Index is 18.44 in the earlier period compared to 18.94 in the latter period. Since Fog Index values greater than 18 are generally classified as unreadable, this implies that the average 10-K is exceptionally difficult to read. Although 10-Ks use technical business language, it is unlikely that investors with some investment experience would view the average document as unreadable.

Generally, there is little variation in the Fog Index across all 10-Ks in our sample. The 10th percentile has a Fog Index of 17.13 compared to 20.26 for the 90th percentile. Li (2008) also documents the limited variability in the Fog Index. This highlights another concern with the use of the Fog Index to measure readability. Given the consistently high Fog Index values for over 90% of the sample (needing a post-college graduate education to understand the text in a first reading), almost all 10-Ks should be viewed as exceptionally difficult to read.

Table II reports a strong trend in 10-K file size over our sample period. Filings later in the sample have significantly more words, tables, pictures, graphics, and HTML code compared to earlier documents. Li (2008) also shows a rise in 10-K size over time. In 2010, a few firms started using XBRL, which also contributes to a larger file size.[10] For the three dependent variables, post-filing root mean square error (RMSE) has a larger mean value (3.45) in the early subperiod than in the post-2002 period (2.26). Both the absolute value of SUE and analyst dispersion have higher values in the later subperiod.

III. Regression Results

In this section, we initially examine whether the Fog Index is an appropriate measure of business text readability and focus on subsequent stock price volatility to capture uncertainty in the information environment attributable to readability. The volatility of returns on or immediately surrounding the 10-K file date is affected by both the information signal and its uncertainty. We believe the uncertainty component, which we are interested in, is more likely to persist beyond the announcement date. Thus, we focus on the RMSE from a market model estimated using trading days [6, 28] relative to the 10-K file date.

Focusing on a market-based measure like subsequent volatility has an advantage over using analyst forecasts due to dramatically larger sample sizes. In addition, since the SEC, through its plain English initiative, is interested in 10-K readability for all investors, a market-based measure like subsequent

[10] eXtensible Business Reporting Language (XBRL) is an XML variant that encapsulates financial data in tags that allow direct computational access. Arguably, one could use the net file size as a similar proxy, where only text content was included. The log transformation of gross file size, however, is correlated at a level greater than 0.7 with net file size. The results from our regressions remain essentially the same if we use net file size as the readability measure. However, this approach would require parsing of each filing to calculate net file size.

Table III

An Analysis of *Fog Index* and Its Components Using Post–Filing Date Market Model Root Mean Square Error (RMSE) as the Dependent Variable

The dependent variable in each regression is the market model RMSE for trading days [6, 28] relative to the 10-K filing date. *Fog index* is equal to 0.4* (average number of words per sentence + percent of complex words). *Average words per sentence* is the number of words in the 10-K divided by a count of sentence terminations. *Percent complex words* is the percentage of 10-K words with more than two syllables. See the Appendix for control variable definitions. All regressions also include an intercept, calendar year dummies, and Fama and French (1997) 48-industry dummies. t-statistics are in parentheses with standard errors clustered by year and industry. All regressions include 66,707 firm-year observations during 1994 to 2011.

	(1)	(2)	(3)	(4)
Readability measures:				
Fog index		0.017		
		(2.04)		
Average words per sentence			0.005	
			(4.02)	
Percent complex words				−0.006
				(−0.77)
Control variables:				
Pre-filing alpha	−0.913	−0.908	−0.908	−0.912
	(−4.12)	(−4.09)	(−4.10)	(−4.11)
Pre-filing RMSE	0.539	0.539	0.539	0.539
	(12.07)	(12.01)	(12.08)	(12.18)
Abs(filing period abnormal return)	5.057	5.052	5.051	5.056
	(17.52)	(17.57)	(17.57)	(17.53)
Log(size in $ millions)	−0.105	−0.105	−0.105	−0.105
	(−5.45)	(−5.45)	(−5.52)	(−5.50)
Log(book-to-market)	−0.133	−0.133	−0.133	−0.133
	(−2.41)	(−2.41)	(−2.41)	(−2.40)
NASDAQ dummy	0.262	0.262	0.263	0.263
	(3.37)	(3.38)	(3.38)	(3.45)
R^2	46.92%	46.93%	46.93%	46.92%

volatility is more inclusive. As noted earlier, our assumption is that more readable 10-Ks should be more informative to investors. The more effectively managers convey relevant valuation information to outsiders, the lower should be subsequent stock market volatility in the month following the 10-K filing (after controlling for past return volatility, firm size, etc.).

A. RMSE and the Fog Index

In Table III, we first consider a regression of a market model RMSE as the dependent variable with the Fog Index as the measure of readability. The control variables are selected because of their ability to explain subsequent stock return volatility. The firm-specific control variables are: 1) *Pre-filing alpha*, the alpha from the market model during the period prior to the filing date;

2) *Pre-filing RMSE*, the RMSE from the prior-period market model regression; 3) *Absolute filing period abnormal return*, the absolute value of the two-day buy-and-hold abnormal return from the filing date (day 0) to day +1; 4) *Log(size in $ millions)*, the log of market capitalization on the day before the file date; 5) *Log(book-to-market)*, the log of the book-to-market ratio taken from data reported prior to the filing date; and 6) *NASDAQ dummy*, a dummy variable equal to one if the firm trades on NASDAQ, and zero otherwise. More detailed variable descriptions are provided in the Appendix.

All regressions also include an intercept, calendar year dummies, and Fama and French (1997) 48-industry dummies. We report *t*-statistics in parentheses with the standard errors clustered by year and industry.

For the results in column (1), where we only consider the control variables, all six of the independent variables are statistically significant in explaining RMSE. The higher are the pre-filing performance and market value, the lower is the subsequent volatility. Firms tilted toward growth (i.e., low book-to-market ratio), firms with higher pre-filing stock return volatility, firms with larger absolute returns on the filing date, and firms listed on the NASDAQ have higher RMSE. The R^2 for the first regression is 46.92%.

The second regression includes the Fog Index as an explanatory variable. The coefficient on *Fog index* is positive and statistically significant (*t*-statistic of 2.04). This is reassuring given prior literature's use of the Fog Index as a readability measure. Recall that the higher is the Fog Index, the less readable is the 10-K text. Thus, the higher is the Fog Index (i.e., the more unreadable the text), the higher is subsequent volatility. Since the Fog Index has only two components (the average number of words per sentence and the percent of complex words), it is interesting to determine which component has a more notable impact on subsequent stock return volatility. Columns (3) and (4) separately consider the components of Fog, that is, the average number of words per sentence and the percent of complex words, respectively.

Column (3) reports that *Average words per sentence* is positively linked with the post-filing date RMSE. The sign of the coefficient on *Average words per sentence* is positive, as expected. The more words per sentence that are in a 10-K, the higher is subsequent volatility after controlling for other variables. The last column of Table III shows that the coefficient on *Percent complex words* is negative (−0.006) and insignificant (*t*-statistic of −0.77). Thus, the percent of complex words has no relation with post-filing RMSE.

Why does the proportion of complex 10-K words have no effect on subsequent volatility? One might think that, as the complexity of 10-K words increases, investors should have more difficulty understanding the firm's operations, which, in turn, should increase volatility. To better understand the role of word complexity in business text, Table IV reports the first quartile of the most frequent complex words (more than two syllables) contributing to the complex word counts for the 10-Ks. Out of more than 45,000 different complex words appearing in 10-Ks during our sample period, only 52 words account for more than a quarter of the total complex word count. As the table shows, the words *financial, company, interest, agreement, including, operations,* and *period*

Table IV

First Quartile of Most Frequently Occurring Complex Words in 10-Ks

Complex words are words containing more than two syllables. The sample contains 66,707 10-K firm-year observations during 1994 to 2011.

Word	% of Total Complex Words	Cumulative%	Word	% of Total Complex Words	Cumulative%
FINANCIAL	1.51%	1.51%	ACCOUNTING	0.38%	16.76%
COMPANY	1.44%	2.95%	INCORPORATED	0.37%	17.13%
INTEREST	0.99%	3.94%	INCLUDED	0.37%	17.49%
AGREEMENT	0.78%	4.73%	COMPENSATION	0.36%	17.85%
INCLUDING	0.77%	5.50%	APPLICABLE	0.36%	18.21%
OPERATIONS	0.71%	6.21%	PRIMARILY	0.35%	18.56%
PERIOD	0.71%	6.92%	ACCORDANCE	0.35%	18.91%
RELATED	0.60%	7.52%	SIGNIFICANT	0.34%	19.26%
MANAGEMENT	0.60%	8.12%	SUBSIDIARIES	0.34%	19.60%
CONSOLIDATED	0.58%	8.70%	CUSTOMERS	0.34%	19.94%
INFORMATION	0.58%	9.28%	RESPECTIVELY	0.34%	20.28%
SERVICES	0.55%	9.83%	REGISTRANT	0.34%	20.62%
PROVIDED	0.55%	10.38%	OBLIGATIONS	0.33%	20.95%
PURSUANT	0.55%	10.93%	PROVISIONS	0.33%	21.28%
FOLLOWING	0.54%	11.47%	LIABILITIES	0.32%	21.60%
SECURITIES	0.54%	12.01%	ADDITION	0.32%	21.92%
APPROXIMATELY	0.52%	12.54%	OTHERWISE	0.32%	22.24%
REFERENCE	0.49%	13.03%	PROPERTY	0.32%	22.56%
OPERATING	0.47%	13.50%	EMPLOYEES	0.32%	22.87%
MATERIAL	0.46%	13.96%	BENEFIT	0.32%	23.19%
CAPITAL	0.43%	14.39%	REPORTING	0.32%	23.51%
EXPENSES	0.42%	14.81%	PRINCIPAL	0.31%	23.82%
CORPORATION	0.40%	15.21%	DEVELOPMENT	0.31%	24.13%
OUTSTANDING	0.40%	15.61%	REVENUE	0.30%	24.43%
ADDITIONAL	0.39%	16.00%	EQUITY	0.30%	24.73%
EFFECTVE	0.38%	16.38%	INSURANCE	0.30%	25.04%

account for almost 7% of all words with more than two syllables. None of the frequently used complex words would cause investors difficulty in determining their meaning. Frequent 10-K usage of words like *management*, *corporation*, *customers*, *revenue*, or *expenses* should also not confound investors' attempts to understand firm valuation, as these are all commonly known words used to describe business operations. When we examine the most frequently used complex words contained in 10-Ks by number of syllables, we find that *telecommunication, telecommunications,* and *confidentiality* account for more than 50% of all seven-syllable words, and *consolidated*, *approximately*, *incorporated*, *subsidiaries*, and *liabilities* account for more than 25% of all five-syllable words. Taken together, the list in Table IV highlights why increased complexity of 10-K words does not affect subsequent volatility: the most frequently used multisyllable words contained in a 10-K are easily understood by investors. Table IV thus underscores the challenge of measuring business document readability

Table V

Correlations of Alternative Readability Measures

Log(file size) is the natural log of the text document file size in megabytes. *Fog index* is equal to 0.4* (average number of words per sentence + percent of complex words). *Average words per sentence* is the number of words in the 10-K divided by a count of sentence terminations. *Percent complex words* is the percentage of 10-K words with more than two syllables. *Common words* is the average across all words in the document of the percent of documents from all 10-Ks in which each word appears. *Financial terminology* is the count of all financial words taken from Campbell Harvey's Hypertextual Finance Glossary. *Vocabulary* is the number of unique words from Loughran–McDonald's (2011) Master Dictionary appearing in a 10-K divided by the total number of words in the Master Dictionary. *Log(# of words)* is the natural logarithm of the word count from the 10-K.

	Log (file size)	*Fog index*	*Average words per sentence*	*Percent complex words*	*Common words*	*Financial terminology*	*Vocabulary*
Fog index	0.367						
Average words per sentence	0.316	0.885					
Percent complex words	−0.015	−0.089	−0.542				
Common words	−0.619	−0.465	−0.572	0.385			
Financial terminology	−0.407	−0.301	−0.372	0.254	0.781		
Vocabulary	0.668	0.497	0.596	−0.377	−0.970	−0.724	
Log(# of words)	0.712	0.560	0.652	−0.384	−0.916	−0.615	0.946

using the Fog Index. Although syllabication is an important discriminator in separating 1st grade from 6th grade text books, it likely does not measure clarity in business writing.

The other component of the Fog Index is based on accurately measuring sentence length. As previously discussed, sentence parsing is very difficult for business documents where things such as abbreviations, section headings, and long lists provide tripwires for automated detection of sentence boundaries. Many extremely long 10-K sentences are merely the result of bullet points not easily identified by a parsing algorithm. This further suggests that the Fog Index does not translate well into a readability measure for business documents.

B. Correlations among Alternative Readability Measures

Due to the above concerns with using the Fog Index as a readability measure, it is natural to ask whether other measures better capture the readability of business text. Table V reports the correlations between 10-K file size, Fog Index, the components of the Fog Index, and a series of other measures of document comprehensibility that we define below. Of all the listed readability measures, file size is by far the simplest to obtain. This variable requires no parsing since it is merely the file size (in megabytes) listed for the "complete submission text file" on EDGAR for the 10-K filing.

To construct *Common words,* one of our alternative measures of readability in Table V, we first determine the relative frequency of all words occurring in all 10-K filings across the full sample period. Words like *and*, *the*, *to*, and *be* occur in virtually all documents, while *recapitalize* occurs in about 0.1% of all 10-K filings. We then calculate *Common words* as the average of this proportion for every word occurring in a given 10-K document. The higher the value of common words, the more ordinary is the 10-K's language and thus the more readable it should be.

Next, we construct *Financial terminology* as the number of unique words appearing in a 10-K that are also contained in Professor Campbell Harvey's finance glossary, divided by the number of unique 10-K words.[11] Note that this measure is not based on a frequency count. The *Financial terminology* variable tabulates only the first occurrence of a term. This procedure avoids giving enormous weights to repeatedly used words like *assets*. As noted by Rennekamp (2012, p. 1345), "the use of 'jargon' is likely to feel more fluent to those with more experience in financial reporting settings, suggesting important experience effects of processing fluency." If financial terminology is considered "jargon" in its usual sense, the impact of such terms could be posited to have a negative impact on readability. However, given the commonality of many of these words, this measure could capture a document's focus on valuation-relevant material (i.e., for individuals with experience in a field, discipline-specific terminology has higher information content). As we will see below, the empirical results support the latter interpretation.

Our third alternative measure of "readability" is *Vocabulary*, the number of unique 10-K words appearing in a filing divided by the total number of possible words in a master dictionary. Thus, the higher is the vocabulary value, the lower is the 10-K's readability, to the extent we assume that an extensive vocabulary makes a document less comprehensible. It is interesting to note that one of the highest *Vocabulary* values belongs to the *New York Times*' 10-K filed on February 26, 2008. Our final readability-related measure, *Log(# of words)*, has been used by several accounting papers as a measure of complexity (see Li (2008), You and Zhang (2009), Miller (2010), and Lawrence (2013)).

In Table V, *Log(file size)* has the expected relations with the alternative readability measures. That is, file size is positively correlated with the number of 10-K words (0.712), the document's vocabulary (0.668), and the Fog Index (0.367) while negatively linked with financial terminology (−0.407) and common word usage (−0.619). File size has almost no correlation with the percentage of complex words in the 10-K (−0.015).

Also, note the surprisingly high negative correlation between *Average words per sentence* and *Percent complex words* (−0.542). One might expect the two components of the Fog Index to be positively correlated with each other. Yet, in business text, as the length of the sentence increases, more short words are needed to link complex words together. Thus, long sentences will have a lower

[11] The words were downloaded from http://people.duke.edu/~charvey/Classes/wpg/glossary.htm. We did not include abbreviations and phrases.

percentage of complex words. Given the highly negative correlation between *Average words per sentence* and *Percent complex words*, it is not likely that both of these components are measuring readability.

Taken together, the evidence in Table V strongly suggests that complex words are adding measurement error and not just noise as a component of the Fog Index. One should expect that as the frequency of complex words increases, the number of common words in the 10-K should decline. Yet, the correlation between complex and common words is positive (0.385). Also, the correlation between complex words and vocabulary (−0.377) has an unexpected sign. As the proportion of complex words rises (i.e., the 10-K contains more multisyllable words), the number of unique words should increase, not decline.

Thus, in the traditional interpretation of the complex word component of the Fog Index, complex words are not simply adding noise, but in some cases appear to contradict the presumed impact of longer words on reading comprehension. This result could be explained by the recent findings of Piantadosi, Tily, and Gibson (2011), who show that information content predicts word length, which implies that more complex words are more informative. Whether complex words are being interpreted as making a document less readable when they are, in fact, adding information, or they are simply identifying very basic and common words as challenging, the results suggest that a key component in the Fog Index is misspecified in its application to financial documents.

There are many alternative ways to potentially gauge readability. However, several of the measures are collinear with each other and none of them provide a dominant alternative. All of the alternative measures require some degree of parsing, which, given the nature of parsing algorithms, can make replication challenging. Given the collinearity of the alternative measures, we could derive one or two principal components as an alternative to Fog, but this would only compound the parsing issue. Thus, we focus on file size as an omnibus measure of readability, a measure that appears to strongly correlate with readability attributes, and one that is easily replicated. Later in the paper, we also show how the alternative measures in Table V compare with file size in the context of our regression analyses.

C. Impact of File Size in Regressions of RMSE

How well does 10-K file size explain subsequent stock return volatility? Table VI reports the coefficient values and significance levels when 10-K file size and Fog Index are regressed against RMSE for days [6, 28] relative to the 10-K file date. In all the columns of Table VI, the six control variables from Table III are included in the regressions and reported in an Internet Appendix.[12] Column (1) of Table VI reports that the coefficient on *Log(file size)* is positive

[12] The Internet Appendix is available in the online version of the article on the *Journal of Finance* website.

Table VI

A Comparison of *Log(file size)*, *Fog Index*, and the Components of *Fog Index* Using Post-Filing Date Market Model RMSE as the Dependent Variable

The dependent variable in each regression is the market model RMSE for trading days [6, 28] relative to the 10-K filing date. *Log(file size)* is the natural log of the text document file size in megabytes. *Fog index* is equal to 0.4* (average number of words per sentence + percent of complex words). *Average words per sentence* is the number of words in the 10-K divided by a count of sentence terminations. *Percent complex words* is the percentage of 10-K words with more than two syllables. Control variables from Table III are included in the regressions and their corresponding estimates are reported in the Internet Appendix. All regressions also include an intercept, calendar year dummies, and Fama and French (1997) 48-industry dummies. *t*-statistics are in parentheses with standard errors clustered by year and industry. All regressions include 66,707 firm-year observations during 1994 to 2011.

	(1)	(2)	(3)
Readability measures:			
Log(file size)	0.073	0.069	0.076
	(4.60)	(4.25)	(3.36)
Fog index		0.006	
		(0.73)	
Average words per sentence			0.003
			(0.75)
Percent complex words			0.010
			(0.67)
R^2	46.96%	46.97%	46.97%

and significant (*t*-statistic of 4.60).[13] This implies that, as the 10-K file size increases, so does subsequent firm volatility.

When *Fog index* is added in column (2), *Log(file size)* remains significant while the coefficient on *Fog index* is statistically insignificant. Column (3) includes *Log(file size)* along with the two components of *Fog index*. In the third regression, the coefficient on *Log(file size)* continues to be positive and significant at the 1% level, while both *Average words per sentence* and *Percent complex words* are not significantly related to subsequent volatility. This suggests that file size might be a better proxy for 10-K readability (as measured by subsequent volatility) than the more commonly used Fog Index. However, the result by itself is not conclusive.

These results are econometrically ambiguous. We have two collinear indicators, measured with error, and at least one might be capturing a correlation with an important omitted variable. Thus, the results in Table VI themselves do not provide an unambiguous case for file size dominating the Fog Index. We believe, however, that these results along with our subsequent findings make a strong argument in favor of file size as a measure of readability in 10-Ks.

[13] In regression (1), the standard errors are clustered by year and industry. If instead the standard errors are clustered at the firm level, the coefficient on *Log(file size)* remains statistically significant (*t*-statistic of 7.61).

It is important to address the economic significance of the results. The standard deviation of *Log(file size)* and subsequent volatility are 1.12 and 2.13, respectively. The regression results imply that a one-standard-deviation increase in file size leads to an increase that is only 4% of subsequent volatility's standard deviation ((0.076*1.12)/ 2.13). This is in contrast to the substantial effect of prior-period volatility. For pre-filing RMSE, a one-standard-deviation increase leads to an increase that is 48% of subsequent volatility's standard deviation ((0.539*1.88)/2.13). Thus, the economic magnitude of file size in explaining subsequent volatility, in the presence of the control variables, is limited, but this is not surprising. We would not argue that readability is a primary determinant of stock price volatility.

IV. Alternative Measures of the Information Environment

Although we believe that RMSE provides the most inclusive surrogate for readability, there are other measures of the information environment that focus more on the ability of analysts to assimilate data into earnings projections. In this section, we consider two alternative dependent variables that focus on the ability of analysts to incorporate information from the 10-K: earnings surprises and analyst dispersion.

A. Earnings Surprise

In this section, we examine SUE, a variable that is different from subsequent volatility but should be linked to 10-K readability as well. Our assumption is, all else being equal, better written 10-Ks should have lower earnings surprises. The SUE measure is defined as (actual earnings − average expected earnings)/stock price. Actual earnings and mean analyst forecasts are obtained from the I/B/E/S unadjusted data files (to avoid the rounding issue). Since we are interested in the magnitude of the earnings surprise (regardless of whether it is a positive or negative surprise to analysts), we use the absolute value of SUE. Because of our need for I/B/E/S data in this analysis, the number of observations drops substantially (from 66,707 to 28,434) when *Abs(Sue)* is the dependent variable compared to the sample using post-filing RMSE.

As before, the various control variables, industry and calendar year dummies, and intercept are included in the regressions. Since the number of analysts following a firm might be linked with earnings surprises and is available to investors prior to the earnings announcement, *# of analysts* is added as an additional control. The first regression in Table VII includes only the control variables with the absolute value of SUE as the dependent variable. Firms with higher absolute earnings surprises tend to be smaller in size, tilted toward value (i.e., high book-to-market), listed on the NYSE/Amex, and subject to worse stock performance and higher volatility in the pre-filing period.

In column (2), when *Log(file size)* is added as an independent variable, its coefficient (0.046) is positive and statistically significant at the 1% level (*t*-statistic of 5.53). Thus, the larger the 10-K's file size, the higher is the absolute

Table VII

Robustness of *Log(file size)* and *Fog Index* to Alternative Measures of 10-K Impact

The dependent variable for the regressions in the first three columns is *Abs(Sue)*, measured as the absolute value of the standardized unexpected earnings. The dependent variable for the regressions in the last three columns is *Analyst dispersion*, defined as the standard deviation of analysts' earnings forecasts prior to the subsequent earnings announcement date scaled by stock price. *Log(file size)* is the natural log of the text document file size in megabytes. *Fog index* is equal to 0.4 × (average number of words per sentence + percent of complex words). See the Appendix for detailed definitions of the variables. All regressions also include an intercept, calendar year dummies, and Fama and French (1997) 48-industry dummies. *t*-statistics are in parentheses with standard errors clustered by year and industry.

	Dependent Variables					
	(1) *Abs(Sue)*	(2) *Abs(Sue)*	(3) *Abs(Sue)*	(4) *Analyst dispersion*	(5) *Analyst dispersion*	(6) *Analyst dispersion*
Readability measures:						
Log(file size)		0.046			0.023	
		(5.53)			(3.51)	
Fog index			−0.003			−0.000
			(−0.82)			(−0.02)
Control variables:						
Pre-filing alpha	−0.365	−0.361	−0.366	−0.270	−0.268	−0.270
	(−4.68)	(−4.68)	(−4.69)	(−4.99)	(−4.99)	(−4.99)
Pre-filing RMSE	0.117	0.115	0.117	0.088	0.087	0.088
	(5.47)	(5.36)	(5.48)	(4.84)	(4.80)	(4.84)
Abs(filing period abnormal return)	0.960	0.958	0.962	0.514	0.510	0.514
	(4.84)	(4.84)	(4.85)	(4.36)	(4.33)	(4.33)
Log(size in $ millions)	−0.054	−0.061	−0.054	−0.022	−0.026	−0.022
	(−6.21)	(−6.71)	(−6.22)	(−4.15)	(−4.78)	(−4.16)
Log(book-to-market)	0.104	0.099	0.104	0.065	0.062	0.065
	(4.85)	(4.64)	(4.85)	(4.09)	(3.96)	(4.10)
NASDAQ dummy	−0.089	−0.086	−0.089	−0.053	−0.051	−0.053
	(−6.56)	(−6.58)	(−6.53)	(−3.81)	(−3.87)	(−3.81)
# of analysts	−0.002	−0.002	−0.002	0.006	0.006	0.006
	(−1.11)	(−1.37)	(−1.11)	(3.35)	(3.34)	(3.35)
Number of observations	28,434	28,434	28,434	17,960	17,960	17,960
R^2	23.30%	23.54%	23.30%	25.34%	25.52%	25.34%

earnings surprise, even after controlling for other variables. If the standard errors are clustered at the firm level, the coefficient on *Log(file size)* remains significant (*t*-statistic of 8.30). When the Fog Index is added as a variable in column (3), its coefficient is insignificant at conventional levels. Again, this evidence is consistent with 10-K file size being a better proxy for readability than the Fog Index. If *Average words per sentence* is an independent variable instead of *Fog index* in the column (3) regression, its coefficient is positive and statistically significant (*t*-statistic of 2.23). Thus, the insignificant relation between *Fog index* and *Abs(Sue)* is being driven by the inclusion of complex words in the tabulation of the Fog Index.

A one-standard-deviation increase in file size leads to an increase in *Abs(Sue)* that is 9% of its standard deviation ((0.046*1.12)/0.577). As before, in terms of economic significance, we do not expect readability to be a primary determinant of SUE outcomes. However, the results indicate that the impact of file size is significant and nontrivial.

B. Analyst Dispersion

Following the work of Lehavy, Li, and Merkley (2011), the last three columns of Table VII use analyst dispersion as the dependent variable. One could imagine that, as the readability of the 10-K declines, analysts would have a more difficult time forecasting earnings. Hence, less readable 10-Ks should be directly linked with firms having higher analyst dispersion. It is important to point out that, if the 10-K text is unclear, analysts obviously have the ability to directly ask management for clarification during conference calls or during a one-on-one interaction (see McCafferty (1997)). For example, Bowen, Davis, and Matsumoto (2002) provide evidence that conference calls help lower analyst dispersion.

For the results in column (4) of Table VII, where we only consider the control variables, all seven of the independent variables are statistically significant in explaining analyst dispersion. The higher are the pre-filing performance and market value, the lower is analyst dispersion. Firms tilted toward value (i.e., high book-to-market ratio), firms with higher pre-filing stock return volatility, firms listed on the NYSE/Amex, and firms followed by more analysts have higher analyst dispersion. The finding that small firms and value firms have significantly higher analyst dispersion is consistent with the evidence reported by Diether, Malloy, and Scherbina (2002).

When *Log(file size)* is added as an explanatory variable in column (5), the variable has a positive and statistically significant coefficient.[14] Thus, a one-standard-deviation increase in file size leads to an increase in analyst dispersion that is 8% of its standard deviation ((0.023*1.12)/0.34). The last regression of Table VII includes the Fog Index as a right-hand-side variable. As with absolute SUE, the coefficient on the Fog Index is statistically insignificant (*t*-statistic of −0.02).[15] In contrast to our results, Lehavy, Li, and Merkley (2011) find a statistically significant positive relation between the Fog Index and analyst dispersion. What accounts for the difference between the papers?

If we restrict our analysis to their 1995 to 2006 time period, controlling for common control variables between the two papers like log(size), absolute filing-period returns, and calendar time and industry fixed effects, we can replicate

[14] If the standard errors are clustered at the firm level, the coefficient on *Log(file size)* remains significant (*t*-statistic of 5.76).

[15] If average words per sentence is included in the regression instead of the Fog Index, the variable has a positive (0.002) and statistically significant (*t*-statistic of 2.34) coefficient.

the results of Lehavy, Li, and Merkley (2011, Table VII). Lehavy, Li, and Merkley (2011) report a coefficient on the Fog Index of 0.002 with analyst dispersion as the dependent variable, while we obtain a value of 0.003. However, once the sample is expanded to the 1994 to 2011 period (six more years of data), the coefficient on the Fog Index becomes insignificant.

Taken together, the results suggest that one component of the Fog Index is not always robust to different information environments. In sum, a simple measure like 10-K file size appears to better capture how effectively managers communicate valuation-relevant information to investors as measured by subsequent volatility, earnings surprises, and analyst dispersion than a traditional readability measure like the Fog Index.

V. Robustness and Alternative Readability Measures

A. Robustness Tests Using Business Segment Data

As noted in the introduction, it is difficult to disentangle 10-K readability from firm complexity. It might be the case that firms with more complexity in the type of business/projects they engage in have greater subsequent volatility, higher earnings surprises, and larger analyst dispersion. Perhaps, once the complexity of the firm is properly controlled for, the link with file size and our three dependent variables will disappear.

Jennings, Stoumbos, and Tanlu (2012) examine the impact of organizational complexity on earnings forecast behavior. We focus on their measure of structural complexity based on a business segment index for each firm. Specifically, we define the business segment index as the sum of the squared business segment proportions as reported for the firm in the COMPUSTAT Segment database. This measure ranges in value from 0.11 to 1.00. Lower values of *Business segment index* imply more firm-specific complexity (i.e., numerous different business segments with substantial sales). As an example, the diverse conglomerate General Electric consistently has very low values for the business segment measure. Due to the availability of segment data, there is a drop in the sample size compared to our earlier analysis.

In Table VIII, we rerun the paper's main results. Each regression has a different dependent variable: column (1) uses post-filing RMSE; column (2) uses the absolute value of SUE, while the last regression uses analyst dispersion as the left-hand-side variable. Each of the three regression models also includes the same control variables as in the prior tables. The complete regression results including the control variables are reported in the Internet Appendix.

In the presence of the business segment measure, *Log(file size)* remains significant in all three regressions in Table VIII. Both the coefficient on *Log(file size)* and its significance level remain similar to our earlier results. For example, in column (1) of Table VI, *Log(file size)* has a coefficient of 0.073 (t-statistic of 4.60) when the dependent variable is post-filing RMSE compared to 0.084 (t-statistic of 4.84) in the first regression of Table VIII. Only in column (2) does the coefficient on *Business segment index* have the expected sign. Lower values

Table VIII

The Effect of Complexity as Measured by a Business Segment Index on the Various Regression Models

The dependent variable in column (1), *Post-filing RMSE*, is the market model RMSE for trading days [6, 28] relative to the 10-K filing date. The dependent variable in column (2) is *Abs(Sue)*, measured as the absolute value of the standardized unexpected earnings. The dependent variable in column (3) is *Analyst dispersion*, defined as the standard deviation of analysts' earnings forecasts prior to the subsequent earnings announcement date scaled by stock price. *Log(file size)* is the natural log of the text document file size in megabytes. *Business segment index* is the sum of the squared proportion in each business segment based on COMPUSTAT Segment data. Each regression model also includes the same control variables as the prior tables with their associated statistics reported in the Internet Appendix. The regressions also include an intercept, calendar year dummies, and Fama and French (1997) 48-industry dummies. *t*-statistics are in parentheses with standard errors clustered by year and industry.

	Dependent Variable		
	(1) *Post-filing RMSE*	(2) *Abs(Sue)*	(3) *Analyst dispersion*
Log(file size)	0.084	0.044	0.022
	(4.84)	(4.93)	(3.49)
Business segment index	0.156	−0.045	−0.009
	(3.80)	(−2.14)	(−0.87)
Number of observations	50,739	22,783	14,516
R^2	47.27%	21.83%	23.98%

for the business segment variable are related to higher absolute SUE. Thus, a common measure of business complexity does not appear to explain the relation we find between file size and the information environment.[16]

One could argue that file size is not so much a proxy for readability as it is for firm complexity. Although file size remains significant even in the presence of a well-accepted measure of complexity, ultimately, it is impossible to totally disentangle the concepts of complexity and readability. That certain aspects of complexity might be inherent in a true measure of readability seems reasonable.

B. Alternative Measures of Readability

This section expands our analysis of the alternative readability measures reported in Table V. Would researchers be better off using more complicated and more technically challenging measures for 10-K readability? In Table IX, we consider eight readability measures in the context of the regressions reported

[16] We also considered, from Jennings, Stoumbos, and Tanlu (2012), geographic segment data and "sophistication and quality of accounting and control systems," defined as the time lag between the fiscal year-end and the 10-K filing date. The geographic segment data further reduced the sample and were not significant in any of the regressions in Table VIII. The time lag variable was significant only in the post-filing RMSE regression. None of these variations affected the significance of 10-K file size.

Table IX

A Comparison of Alternative Readability Measures in 24 Separate Regressions

For each dependent variable, we report the coefficient and t-statistic associated with the different readability measures. Each entry in the table is based on a separate regression (i.e., 24 separate regressions). Control variables from the prior analysis are included in the regressions with their associated statistics reported in the Internet Appendix. For the dependent variable *Post-filing RMSE*, the control variables are those from Table III. For *Abs(Sue)* and *Analyst dispersion*, the control variables are those from Table VII. See the Appendix for the readability measure and control variable definitions. All regressions also include an intercept, calendar year dummies, and Fama and French (1997) 48-industry dummies. t-statistics are in parentheses with standard errors clustered by year and industry.

	Dependent Variable		
	(1)	(2)	(3)
Readability Measure	*Post-filing RMSE*	*Abs(Sue)*	*Analyst dispersion*
Log(file size)	0.073	0.046	0.023
	(4.60)	(5.53)	(3.51)
Fog index	0.017	−0.003	−0.000
	(2.04)	(−0.82)	(−0.02)
Average words per sentence	0.005	0.002	0.002
	(4.02)	(2.23)	(2.34)
Percent complex words	−0.006	−0.014	−0.009
	(−0.77)	(−5.75)	(−4.08)
Common words	−1.295	−0.614	−0.437
	(−4.56)	(−5.49)	(−4.47)
Financial terminology	−8.601	−1.460	−0.906
	(−4.34)	(−2.68)	(−2.51)
Vocabulary	7.826	4.094	2.835
	(4.72)	(6.31)	(5.68)
Log(# of words)	0.086	0.062	0.041
	(4.27)	(6.55)	(4.79)
Number of observations	66,707	28,434	17,960

in Tables VI and VII. That is, for the dependent variables *Post-filing RMSE, Abs(Sue),* and *Analyst Dispersion,* each readability measure is included in a separate regression with the corresponding control variables (a total of 24 separate regressions). Thus, for *Log(file size)*, the first coefficient and t-statistic in column (1), 0.073 and 4.60, are identical to the full regression results in the first cell of Table VI.

In column (1), when the dependent variable is RMSE, all of the various readability measures are statistically significant and have the expected sign except for *Percent complex words*. Firms with higher file size, Fog Index, average words per sentence, vocabulary, or number of words are linked with higher subsequent volatility. As expected, both common words and financial terminology have negative coefficients. Thus, greater usage of common words or financial jargon (i.e., *assets, lease, securities,* and *partnership*) is associated with lower levels of volatility.

Generally, the same pattern exits when *Abs(Sue)* and *Analyst dispersion* are the dependent variables in the last two columns. The exceptions are that *Fog index* now has insignificant coefficients, while the coefficient on *Percent complex words* has the wrong sign with the different dependent variables. The economic effect of the different readability measures is somewhat similar, with file size having a slight edge. For example, the regression results imply that a one-standard-deviation increase in file size, number of words, common words, and vocabulary leads to a respective increase that is 9%, 7%, 6%, and 6% of absolute SUE's standard deviation. When the dependent variable is analyst dispersion, the regression coefficients imply that a one-standard-deviation increase in file size, number of words, vocabulary, and common words leads to a respective increase that is 8%, 8%, 7%, and 6% of analyst dispersion's standard deviation.

In sum, with the exceptions of *Fog index* and *Percent complex words*, all of the alternative readability measures considered in Table IX appear to provide reasonable proxies for measuring readability, where readability is defined as the ability to assimilate valuation-relevant information. Yet, all of these measures require parsing of 10-K filings with the exception of *Log(file size)*. Since file size performs at least as well as the other readability items and is the easiest to calculate, we recommend its use by researchers in measuring readability of financial text.

VI. Conclusions

The Fog Index has become a popular measure of financial disclosure readability in recent accounting and finance research. The SEC has even contemplated the use of the Fog Index to help identify poorly written financial documents. However, the measure has migrated to financial applications without its efficacy in the context of business disclosures having been determined.

We argue that traditional readability measures like the Fog Index are poorly specified in the realm of business writing. The Fog Index is based on two components: sentence length and word complexity. Although sentence length is a reasonable readability measure, it is difficult to accurately measure in financial documents. More importantly, we show that the count of multisyllabic words in 10-K filings is dominated by common business words that should be easily understood. Frequently used "complex" words like *company*, *operations*, and *management* are not going to confuse consumers of SEC filings. Additionally, the correlation of complex words with alternative measures of readability contradicts its traditional interpretation.

We find that the 10-K file size provides a better proxy for readability than traditional measures. As a measure of readability in financial disclosures, where readability is the ability to assimilate valuation-relevant information, we recommend that researchers use the file size of the "complete submission text file" available on the SEC's EDGAR website. The measure does not require any parsing of the document and is readily replicated. Note, however, that we would not expect this measure to translate well to other types of documents like news articles and press releases.

In regression analysis, we report that, after controlling for other variables, larger 10-K file sizes have significantly higher post-filing date abnormal return volatility, higher absolute SUE, and higher analyst dispersion. This relation does not seem to be a simple artifact of firm complexity. The less material investors and analysts must digest to get valuation-relevant information from company managers, the better they are at predicting subsequent value-relevant events.

Our paper has a policy implication for the SEC. If a central purpose of the 10-K is effective communication of valuation-relevant information to investors, then the SEC should focus less on style—which is undifferentiated in 10-Ks—and instead encourage managers to write more concisely. Clearly, an SEC rule simply dictating page limitations (presumably conditional on factors such as firm size and industry) is not a reasonable solution. The SEC, however, should emphasize to filers that the benefit of exhaustive disclosure in the interest of litigation avoidance must be balanced with the costs of information overload and effective communication. Concisely written documents are more likely to be read, and the information from the 10-K is more likely to be effectively incorporated into stock prices and analyst forecasts.

Initial submission: August 31, 2011; Final version received: December 4, 2013
Editor: Campbell Harvey

Appendix: Variable Definitions

Tested Readability Measures	
Fog index	Equal to 0.4*(average number of words per sentence + percent of complex words). High values of the Fog Index imply less readable text.
Average words per sentence	The number of words in the 10-K divided by the total number of sentence termination characters after removing those associated with headings and abbreviations.
Percent complex words	The percentage of 10-K words with more than two syllables.
Log(file size)	The natural logarithm of the file size in megabytes of the SEC EDGAR "complete submission text file" for the 10-K filing.
Alternative Readability Measures	
Common words	We multiply each parsed word times the percent of 10-K documents in which that word appears based on the Loughran–McDonald (2011) Master Dictionary and average this result across all words in the document. For example, the average of *Common words* in the sample is 0.39. Thus, the average word in the average filing appears in about 39% of all 10-K filings.

(*Continued*)

Appendix—*Continued*

Financial terminology	The number of unique words in a 10-K that appear in Professor Campbell Harvey's Hypertextual Finance Glossary (http://people.duke.edu/~charvey/Classes/wpg/glossary.htm) divided by the number of unique 10-K words. We remove abbreviations and phrases from his list.
Vocabulary	The number of unique words appearing in the filing divided by the total number of entries in the Loughran–McDonald (2011) Master Dictionary.
Log(# of words)	The natural logarithm of the word count from the 10-K, based on words appearing in the Loughran–McDonald Master Dictionary.
Dependent Variables	
Post-filing RMSE	The RMSE from a market model estimated using trading days [6, 28] relative to the 10-K file date (approximately one calendar month). There must be a minimum of 10 observations to be included in the sample.
Abs(Sue)	The absolute value of the SUE, which is defined as (actual earnings−average expected earnings)/stock price. This variable is multiplied by 100, winsorized at the 1% level, and requires at least one analyst making a forecast. The actual earnings and mean analyst forecasts are obtained from the I/B/E/S unadjusted data files (to avoid the rounding issue). Only forecasts occurring between the 10-K filing date and the next earnings announcement date are included, and thus stale forecasts are not in the sample. If a given analyst has more than one forecast reported during this time interval, only the forecast closest to the filing date is included in the sample.
Analyst dispersion	The standard deviation of analysts' forecasts appearing in the SUE estimate divided by the stock price from before the 10-K filing date. Firms with less than two analyst forecasts are assigned a missing value.
Control Variables	
Pre-filing alpha	The alpha from a market model using trading days [−252, −6]. At least 60 observations of daily returns must be available to be included in the sample.
Pre-filing RMSE	The RMSE from a market model estimated using trading days [−257, −6], with a minimum of 60 complete observations.
Abs(filing period abnormal return)	The absolute value of the filing date excess return, measured by the buy-and-hold return starting on filing date (day 0) through day +1 minus the buy-and-hold return of the CRSP value-weighted index over the same two-day period.
Log(size in $ millions))	The natural logarithm of the CRSP stock price times shares outstanding on the day prior to the 10-K filing date (in $ millions).

(*Continued*)

Appendix—*Continued*

Log(book-to-market)	The natural log of book-to-market, following Fama and French (2001) using data from both COMPUSTAT (book value from most recent year prior to filing date) and CRSP (market value of equity). After removing firms with negative book value, the variable is winsorized at the 1% level.
NASDAQ dummy	Dummy variable set to one if the firm is listed on NASDAQ at the time of the 10-K filing, else zero.
# of analysts	The number of analysts used in the *Analyst dispersion* calculation.
Business segment index	The sum of the squared business segment proportions reported for the firm in the COMPUSTAT Segment database based on sales data.

REFERENCES

Antweiler, Werner, and Murray Z. Frank, 2004, Is all that talk just noise? The information content of Internet stock message boards, *Journal of Finance* 59, 1259–1293.

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

Bloomfield, Robert, 2008, Discussion of annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics* 45, 248–252.

Bowen, Robert, Angela Davis, and Dawn Matsumoto, 2002, Do conference calls affect analysts' forecasts? *Accounting Review* 77, 285–316.

Dale, Edgar, and Jeanne S. Chall, 1948, A formula for predicting readability, *Education Research Bulletin* 27, 37–54.

Davison, Alice, and Robert N. Kantor, 1982, On the failure of readability formulas to define readable texts: A case study from adaptations, *Reading Research Quarterly* 17, 187–209.

De Franco, Gus, Ole-Kristian Hope, Dushyantkumar Vyas, and Yibin Zhou, 2013, Analyst report readability, *Contemporary Accounting Research*, forthcoming.

Diether, Karl, Christopher Malloy, and Anna Scherbina, 2002, Differences of opinion and the cross section of stock returns, *Journal of Finance* 57, 2113–2141.

Dougal, Casey, Joseph Engelberg, Diego Garcia, and Christopher Parsons, 2012, Journalists and the stock market, *Review of Financial Studies* 25, 639–679.

DuBay, William, 2004, *The Principles of Readability* (Impact Information, Costa Mesa, CA).

DuBay, William, 2007, *Unlocking Language* (BookSurge Publishing, Charleston, South Carolina).

Fama, Eugene F., and Kenneth R. French, 1997, Industry costs of equity, *Journal of Financial Economics* 43, 153–193.

Fama, Eugene F., and Kenneth R. French, 2001, Disappearing dividends: Changing firm characteristics or lower propensity to pay? *Journal of Financial Economics* 60, 3–43.

Feldman, Ronen, Suresh Govindaraj, Joshua Livnat, and Benjamin Segal, 2010, Management's tone change, post-earnings announcement drift and accruals, *Review of Accounting Studies* 15, 915–953.

Gray, William S., and Bernice E. Leary, 1935, *What Makes a Book Readable?* (University of Chicago Press, Chicago, Illinois).

Griffin, Paul A., 2003, Got information? Investor response to form 10-K and form 10-Q EDGAR filings, *Review of Accounting Studies* 8, 433–460.

Gunning, Robert, 1952, *The Technique of Clear Writing* (McGraw-Hill, New York).

Jegadeesh, Narasimhan, and Andrew Di Wu, 2013, Word power: A new approach for content analysis, *Journal of Financial Economics* 110, 712–729.

Jennings, Jared, Robert Stoumbos, and Lloyd Tanlu, 2012, The effect of organizational complexity on earnings forecasting behavior, Working paper, Washington University in St. Louis.

Jones, Michael J., and Paul A. Shoemaker, 1994, Accounting narratives: A review of empirical studies of content and readability, *Journal of Accounting Literature* 13, 142–184.

Klare, George R., 1963, *The Measurement of Readability* (Iowa State University Press, Ames, Iowa).

Lawrence, Alastair, 2013, Individual investors and financial disclosure, *Journal of Accounting & Economics* 56, 130–147.

Lehavy, Reuven, Feng Li, and Kenneth Merkley, 2011, The effect of annual report readability on analyst following and the properties of their earnings forecasts, *Accounting Review* 86, 1087–1115.

Leuz, Christian, and Catherine Schrand, 2009, Disclosure and the cost of capital: Evidence from firms' responses to the Enron shock, Working paper, University of Chicago.

Li, Feng, 2008, Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics* 45, 221–247.

Li, Feng, 2010, Managers' self-serving attribution bias and corporate financial policies, Working paper, University of Michigan.

Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *Journal of Finance* 66, 35–65.

McCafferty, J., 1997, Speaking of earning ... Why managing expectations often doesn't work, *CFO* 13, 38–50.

McLaughlin, G. Harry, 1969, SMOG grading: A new readability formula, *Journal of Reading* 12, 639–646.

Mikheev, Andrei, 2002, Periods, capitalized words, etc., *Computational Linguistics* 28, 289–316.

Miller, Brian, 2010, The effects of reporting complexity on small and large investor trading, *Accounting Review* 85, 2107–2143.

Palmer, David, and Marti A. Hearst, 1994, Adaptive sentence boundary disambiguation, *Proceedings of the Fourth ACL Conference on Applied Natural Language Processing*, 78–83.

Piantadosi, Steven, Harry Tily, and Edward Gibson, 2011, Word lengths are optimized for efficient communication, *Proceedings of the National Academy of Sciences* 108, 3526–3529.

Rennekamp, Kristina, 2012, Processing fluency and investors' reactions to disclosure readability, *Journal of Accounting Research* 50, 1319–1354.

Tekfi, Chaffai, 1987, Readability formulas: An overview, *Journal of Documentation* 43, 257–269.

Tetlock, Paul C., 2007, Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance* 62, 1139–1168.

Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy, 2008, More than words: Quantifying language to measure firms' fundamentals, *Journal of Finance* 63, 1437–1467.

You, Haifeng, and Xiao-jun Zhang, 2009, Financial reporting complexity and investor underreaction to 10-K information, *Review of Accounting Studies* 14, 559–586.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.