Working Paper

Annual Report Readability, Earnings, and Stock Returns

Feng Li
Stephen M. Ross School of Business
at the University of Michigan

Ross School of Business Working Paper Series
Working Paper No. 1028
February 28, 2006

Annual Report Readability, Earnings, and Stock Returns *

Feng Li

Ross School of Business, University of Michigan

701 Tappan St., Ann Arbor, MI 48109

Phone: (734)936-2771

Email: Feng@umich.edu

First Draft: July 2005
This Draft: February 28, 2006

*Comments welcome. I thank Daniel Cohen, Ilia Dichev, Suraj Srinivasan, Franco Wong, and the participants of the Hosmer Lunch Workshop at the University of Michigan for comments.

Abstract

This paper examines the relationship between annual report readability and firm performance. This is motivated by the Securities and Exchange Commission's plain English disclosure regulations that attempt to make corporate disclosures easier to read for ordinary investors. I measure the readability of public company annual reports using the Fog Index and the Kincaid Index from computational linguistics. I find that the annual reports of firms with lower earnings are harder to read. Moreover, the positive earnings of firms with annual reports that are easier to read are more persistent in the next one to four years. This suggests that, consistent with the motivation behind the plain English disclosure regulation of the Securities and Exchange Commission, managers may be opportunistically choosing the readability of annual reports to hide adverse information from investors. However, I do not find any correlation between annual report readability and future stock returns, suggesting that the market does impound the implications of disclosure readability into stock prices.

1 Introduction

Ever since the passage of the Securities Act of 1933, the Securities and Exchange Commission ("SEC") has made consistent efforts to make the disclosure documents of public companies more readable(Firtel (1999)). The most recent of these efforts is the plain English disclosure rules adopted by the SEC on January 22, 1998. The underlying argument for the plain English disclosure regulation is that (1) firms could use vague language and format in disclosure to hide adverse information, and (2) the average investors may not understand complex documents and this could result in capital market inefficiency.

The relevance of the regulation, however, is not straightforward. First, there are other information sources (such as financial analyst reports) for investors. Depending on whether the different information sources are complements or substitutes, annual report readability may or may not be relevant. Second, some critics contend that disclosure should primarily be geared towards sophisticated investors due to the complicated nature of technical and financial information (Firtel (1999)). Finally, to the extent that the marginal investors are sophisticated and understand complex disclosure, stock price may not be distorted even if complicated language and format are used.

This paper attempts to provide the first large sample evidence on the relevance of the plain English disclosure. More specifically, I ask the following two sets of questions. First, is there a relation between a company's annual report readability and its current performance? What is the implication of disclosure readability for future performance? Second, do stock market investors understand the implication (if any) of annual report readability for firm performance?

If disclosure readability is strategically used by managers to hide adverse information, a relationship between firm performance and readability would be expected. This management opportunism story argues that managers have incentives to obfuscate information when the current performance is bad. Given that the annual report contains detailed financial numbers on historical firm performance, however, the marginal benefit of using disclosure readability to hide current poor performance seems small. Hence, I also examine the implication of

disclosure readability for future performance. In particular, I examine whether the positive earnings of firms with more complex annual reports are less persistent and whether the negative earnings of these firms are more persistent in the next several years.

I empirically measure annual report readability using established statistics from computational linguistics based on syntactical textual features (such as syllables per word and words per sentence) in the 10-K filing. The intuition is that, everything else equal, more syllables per word or more words per sentence make a document harder to read. Using a sample with more than 50,000 firm-years, I find that firms with lower earnings tend to file annual reports that are more difficult to read; an increase (decrease) in earnings from the previous year also results in annual reports that are easier (more difficult) to read compared with the previous year's reports. This effect holds after controlling for other firm and industry specific factors. However, although this effect is statistically significant, the economic magnitude is small.

I find that annual report readability is related to earnings persistence. Firms with more complicated annual reports have a lower persistence in earnings when they are profitable. The effect is significant both economically and statistically. An inter-quartile change in readability has a similar impact on positive earnings persistence comparable to the effect of an inter-quartile change in the absolute amount of accruals.

However, there is little evidence that the stock market does not understand the implication of annual report readability for future earnings. I do not find any correlation between annual report readability and future stock returns in my main sample, in a sub-sample of small firms, or in a sub-sample of firms with low institutional ownership. Taken together, the evidence in this paper suggests that, consistent with the motivation behind the plain English disclosure regulation of the Securities and Exchange Commission, managers may be opportunistically choosing the readability of annual reports to hide adverse information from investors. The evidence that the stock market impounds this information into prices suggests that contrary to the SEC's concerns, small investors may not be affected by the lack of readability.

Certain caveats are in order. My empirical measures of annual report readability capture

only part of the many requirements of the SEC plain English rules. Hence, it does not speak to other parts of the regulation such as formatting of the prospectuses. Due to data availability, my sample period includes only the post-1994 years. It is possible that the cross-sectional variation in annual report readability may be smaller during this period, resulting in lower powered tests.

This paper contributes to the literature in three ways. First, there is extensive research on the determinants and consequences of accounting choices (Fields, Lys, and Vincent (2001)) and quality of disclosure (Healy and Palepu (2001)). But there is no large-sample study of the determinants and consequences of annual report readability even though the SEC has been advocating disclosure based on more plain English (SEC (1998)). This paper is the first large-sample study to examine the cross-sectional variation in annual report readability and its implication for future earnings and stock returns. The evidence sheds light on the relevance of the plain English disclosure regulation.

Second, much of the empirical literature on corporate disclosure quality has focused on the determinants and consequences of the amount of disclosure. Most papers have small sample sizes, as the disclosure is in general manually coded. Annual report readability captures the ease of reading, rather than the content of disclosure. This paper therefore provides a new empirical measure of disclosure quality that can be studied in a large sample.

Finally, there is extensive research on earnings quality (see Dechow and Schrand (2004) for a comprehensive review). But prior research in general does not study the association between firm disclosure quality and earnings quality.[1] While many papers explicitly link firm performance with disclosure quality (e.g., Lang and Lundholm (1993)) and other papers use earnings quality as a proxy for disclosure quality (e.g., Francis, LaFond, Olsson, and Schipper (2005)), few examine the implication of disclosure quality for future earnings. This paper extends this literature by showing that the quality of disclosure is correlated with earnings persistence and contains information about earnings quality.

[1] One exception is Francis, Nanda, and Olsson (2005), who examine the relation between voluntary disclosure and accrual quality.

The remainder of the paper proceeds as follows. I discuss prior literature and hypotheses in Section 2 and empirical measures of annual report readability in Section 3. I present the basic empirical findings in Section 4. I explore other tests and the robustness of the findings in Section 5. Section 6 concludes.

2 Literature and Hypotheses

The SEC has continually attempted to make public company prospectus more readable and understandable. In several Securities Act Releases after the 1933 Securities Act, it encouraged greater clarity in the disclosure documents with an emphasis on not compromising full and fair disclosure (Firtel (1999)). In 1967, the SEC constituted an internal study group to examine and make recommendations for improving its disclosure regime. This resulted in the 1969 "*Wheat Report*". Among other findings, *the Wheat Report* noted that the average investor could not readily understand the complicated prospectuses and therefore recommended that companies avoid unnecessarily complex, lengthy or verbose writing.

In October 1998, the SEC adopted new plain English disclosure rules that require the usage of plain English in the drafting and format of all prospectuses in registered public offerings by domestic and foreign issuers. The SEC's Investor Ed Office published and posted on its website "A plain English handbook, how to create clear SEC disclosure documents" which provides practical tips for disclosure documents. For instance, when drafting the front and back cover pages, the summary and risk factors sections, an issuer must comply with the following six basic principles: short sentences; definite, concrete, everyday language; active voice; tabular presentation or bullet lists for complex material, whenever possible; no legal jargon or highly technical business terms; and no double negatives. More recently, the SEC has taken several steps in making the disclosure of mutual funds more readable (Glassman (2005)).

2.1 Literature

Given the importance of the plain English disclosure regulation, surprisingly, there is little large sample empirical evidence on its relevance. Jones and Shoemaker (1994) reviewed 32 studies in the fields of accounting, business communication, and management which study the readability of annual report narratives (26 studies), tax law (3 studies), or accounting textbook (3 studies).

Most studies try to assess the reading ease of the annual report and its components. For instance, Smith and Smith (1971) study the readability of the financial statements footnotes of Fortune 50 companies and conclude that the readability level of the notes is restrictive. Healy (1977) studies the reading ease of the footnotes to the financial statements of 50 New Zealand firms. Lebar (1982) studies the Forms 10-Ks, annual reports, and press release by 10 NYSE firms in 1978 and compares the differences in topics and information between them. The general conclusion from these studies is that corporate annual reports are very difficult to read and may be classified as technical literature which risks "being inaccessible to a large proportion of private lay shareholders" (Jones and Shoemaker (1994)). Some studies also specifically investigate whether annual reports have become more difficult to read over time (e.g., Soper and Dolphin (1964); Barnett and Leoffler (1979)) and the evidence is rather mixed (Jones and Shoemaker (1994)).

Other studies examine the association between readability and other variables, including the identity of the external auditor (Smith and Smith (1971) and Barnett and Leoffler (1979)) and corporate profitability (Courtis (1986); Baker and Kare (1992); Subramanian, Insley, and Blackwell (1993)). The evidence is again mixed and inconclusive. For instance, Courtis (1986) finds no strong correlation between readability and net profits and return on capital. However, Subramanian, Insley, and Blackwell (1993) find annual reports of profitable firms are significantly easier to read than those of poor performers.

The sample sizes of the previous studies, however, are very small. Only two of the thirty-two studies reviewed by Jones and Shoemaker (1994) have a sample size slightly larger than 100. Among the sixteen papers examined in Table I of Clatworthy and Jones (2001), fourteen

have a sample size of 50 or smaller and the largest sample size is 120.

In this paper, I extend this literature by evaluating the relevance of plain English disclosure using a large sample with a particular focus on the association between annual report readability and firm performance, future earnings, and stock returns.

2.2 The Implications of Annual Report Readability

2.2.1 Current Performance

One of the implicit assumptions underlying the plain English rules, like many other disclosure regulations, is that management is willing to be less forthcoming in disclosing information when the firm is performing poorly (Lang and Lundholm (1993) and SEC (1998)). This "management opportunism" argument predicts a positive association between firm performance and annual report readability.

Firms with better performance, on the other hand, are likely to make disclosure documents more readable. In the voluntary disclosure models (e.g., Verrecchia (1983)), adverse selection induces "good" firms to disclose more to reveal their performance to investors, so that they can distinguish themselves from the obfuscators. As a result, we observe cross-sectional variation in disclosure quality as a function of firm performance.

Hence, if annual report readability is a managerial choice variable (i.e., if it is relevant to investors and managers), the management obfuscation hypothesis predicts that annual report readability of firms with bad performance is lower than that of the firms with good performance.

2.2.2 Future Performance

The annual reports contain a lot of financial information about current and historical performance. Hence, the benefit to the managers of making the annual reports harder to read in order to hide adverse information about current performance may be small. I therefore further examine the implication of annual report readability for future performance. In

particular, I focus on the persistence of earnings.

The intuition on the relation between disclosure quality and a firm's current performance can be extended to future performance. Opportunistic managers may have incentives to make the annual report harder to read, if good earnings of this year are not persistent or if poor earnings are very persistent. On the other hand, firms with better future performance may want to disclose information more precisely to distinguish themselves from the "lemons" by making their annual reports easier to read. In other words, to the extent that complicated annual reports can hide the transitory nature of the good news or the permanent nature of the bad news, the management obfuscation hypothesis predicts that the profits (losses) of firms with more complex annual reports are less (more) persistent.

Most prior studies on disclosure either examine the relation between disclosure quality and firm performance (e.g., Lang and Lundholm (1993)) or use earnings quality as a proxy for disclosure quality (e.g, Francis, LaFond, Olsson, and Schipper (2005) and Cohen (2005)). A few papers study the relation between disclosure quality and earnings quality: Francis, Nanda, and Olsson (2005) find a positive relationship between voluntary disclosure quality and the accruals quality; Riedl and Srinivasan (2005) examine the implication for earnings persistence of whether special items are recognized as a line item on the income statement or only disclosed in the footnotes. I extend this literature by examining the implication of disclosure readability for earnings persistence.

2.2.3 Stock Returns

An implicit assumption of the plain English disclosure regulation is that the stock market investors may be misled by vague and incomprehensible disclosure documents. In particular, one of the goals of the SEC plain English regulation is to protect small investors. If annual report readability is positively associated with future earnings as predicted by the management obfuscation hypothesis, it is interesting to test whether the stock market investors understand this implication. A systematic correlation between annual report readability and future stock returns indicates an inefficient stock market and supports the necessity of

mandatory plain English disclosure regulation.

To further increase the power of the tests, I also examine two sub-samples (small firms and firms with low institutional ownership) in addition to the main sample. If there is any potential mispricing related to readability, it is more likely to be detected in the two sub-samples.

3 Data and Empirical Measures of Annual Report Readability

3.1 Sample

I collect my sample as follows: (1) I start with the intersection of CRSP-COMPUSTAT firm-years. (2) I then manually match GVKEY (from COMPUSTAT) and PERMNO (from CRSP) with the Central Index Key (CIK) used by SEC online Edgar system. Firms without matching CIK are dropped. (3) I download from Edgar the 10-K filing for every remaining firm-year. Those firm-years that do not have electronic 10-K filings on Edgar are then excluded.[2] (4) For each 10-K file, all the heading items, paragraphs that have fewer than one line, and tables are deleted and those 10-K filings that have less than 3000 words or 100 lines of remaining texts are dropped. The calculation of the annual report readability is based on the remaining material. Details of these steps are presented in the Appendix. Notice that it is important to delete the tables and financial statements in this step, since the readability indices are designed for text rather than numbers or tables. (5) Finally, firm-years that have operating earnings (scaled by book value assets) greater than 1 or less than -1 are deleted from the sample. This yields a sample of 55,719 firm-years from 1994 to 2004.

[2]SEC has electronic Edgar filing available online from 1994.

3.2 The Readability Measures

I use two readability statistics from computational linguistics to measure the annual report readability. These measures assume that the text is well formed and logical and capture text complexity as a function of syllables per word and words per sentence.[3]

(1) Fog

The Fog index, developed by Robert Gunning, is a well known and simple formula for measuring readability. The index indicates the number of years of formal education a reader of average intelligence would need to read the text once and understand that piece of writing with its word sentence workload. It is calculated in the following way:

$$Fog = (words\_per\_sentence + percent\_of\_complex\_words) * 0.4, \quad (1)$$

where complex words are defined as words with three syllables or more. The relation between Fog and reading ease is as follows: $FOG >=$18 means the text is unreadable; 14-18 (difficult); 12-14 (ideal); 10-12 (acceptable); and 8-10 (childish).

(2) Kincaid

This score is calculated as:

$$Kincaid = (11.8 * syllables\_per\_word) + (0.39 * words\_per\_sentence) - 15.59. \quad (2)$$

The Kincaid Index (also referred to as Flesch-Kincaid formula) rates text on U.S. grade school level. So a score of 8.0 means that the document can be understood by an eighth grader. A score of 7.0 to 8.0 is considered to be optimal. Related to the Kincaid Index is the "Flesch" or "Flesch-Kincaid" Reading Ease which rates text on a 100 point scale[4]. The higher the Flesch Reading Ease, the easier is the text. For ease of interpreting the empirical results, I omit the results based on the Flesch statistics because it is negatively correlated with the Fog and Kincaid Indices. However, the empirical results remain similar if Flesch Index is used.

[3]For more information see http://www.plainlanguage.com/Resources/readability.html.

[4]It is calculated as (206.835 - (1.015 * words per sentence) - (84.6 * syllables per word)).

I use the Lingua::EN:Fathom package of PERL language to analyze the raw 10-K files and calculate the readability indices. The Appendix gives the details of the calculation.[5] This program has been often used in fields including information science and business communication. Examples include Collins-Thompson and Callan (2005) and Muresan, Cole, Smith, Liu, and Belkin (2006).

To check the calculation validity of the PERL program, I first compare the numbers reported in this study with those from other studies. Smith and Smith (1971) manually calculate the Flesch Reading Index of some randomly selected footnotes of the 50 biggest Fortune companies. The mean of the Flesch Index per their calculation is 23.49 (Table II of Smith and Smith (1971)). For my sample, the mean and median of the Flesch Index calculated using PERL program are 24.44 and 24.63, which are pretty close to their manually calculated number.

A second way of checking the validity of the calculation is to compare it with manual calculation or other computer program using the same text. MS WORD can report the Kincaid Index. However, for unexplained reasons, Microsoft's version of the Kincaid Index does not score above grade 12, although the original formula scored up to a graduate school level. Since any grade level above 12 will be reported as grade 12, documents at a graduate school reading level will be reported as grade 12 - a measurement error of about 7 grades.[6] For this reason, I randomly select 3 paragraphs from 10 annual reports and count the number of words per sentence and syllables per word manually. The difference between the results from the manual calculation and the PERL programs is smaller than 5% in most cases, confirming the validity of the program.

Overall, I believe that the programs used in this paper should measure the readability

[5]For more information, see http://search.cpan.org/dist/Lingua-EN-Fathom/lib/Lingua/EN/Fathom.pm.

[6]There is evidence that previous research relying on Word's Flesch-Kincaid formula seriously underestimates a document's grade level and all of the research done on health materials using Word's Flesch-Kincaid is seriously flawed. Several readability researchers have contacted Microsoft about this problem, but the company has neither acknowledged the problem nor fixed it. See the Reader Feedback section of this page: http://www.wats.ca/resources/determiningreadability/1.

reasonably well. There is no reason to believe that any measurement error is systematic and will bias the results.

One concern regarding using syntactical features such as the Fog Index to measure readability is that they may not reflect actual comprehension difficulty. However, this concern is more problematic if researchers want to assess the *absolute* level of readability (Jones and Shoemaker (1994)). The focus here is on the relative readability of the annual reports in a cross-section. Hence, while still a caveat, the concern is less worrisome.

3.3 Summary Statistics

Table 1 Panel A presents the summary statistics of the sample. Overall, the annual reports of public companies are very difficult to read. The mean and median Fog Index are 19.4 and 19.2 respectively, which are "unreadable" according to the usual interpretation of the index. The mean (median) Kincaid index is 15.2 (15.0) and is also on the high end. To provide a benchmark, I check the readability index for the articles from *the Wall Street Journal*. I download all the Editorials from the June 2005 issues of *the Wall Street Journal*. On average, the Fog index of the Editorials is 15.2 and the Kincaid index 12.2, suggesting they are much easier to read than a typical annual report.

The standard deviation and the inter-quartile range of the Fog (Kincaid) index of the 10-K filings in my sample are 1.4 (1.4) and 1.7 (1.6) respectively. This variation seems substantial. For instance, the difference in the Fog index between Reader's Digest and the TIME magazine is about 2.[7]

The variation in year-by-year change in the readability indices is not small either. The standard deviation of change in Fog Index is 1.46 and that of Kincaid is 1.48. The 25th and the 75th percentile of year-to-year change in Fog are -0.59 and 0.65 respectively.

Figure 1 A plots the median level of readability indices for the sample firms over time. Interestingly, there is an obvious drop in the indices in the years immediately after 1999, suggesting that the plain English disclosure regulation of 1998 does make companies take

[7]Source: http://en.wikipedia.org/wiki/Fog_index#Typical_Gunning-Fog_indices_of_selected_magazines.

efforts to make their annual reports more readable. As can be seen from Figure 1 B, this drop is also seen in constant sample, defined as firms with at least 8 years of data between 1994 and 2004. However, this trend reverses dramatically after 2002 and the annual reports filed by public firms seem to become even more difficult to read compared with the pre-1998 years. It would be interesting to get a longer time-series of data to examine whether this is related to the Sarbanes-Oxley Act regulation.

Not surprisingly, there is very high correlation between Fog and Kincaid indices (Pearson correlation of 0.97, Panel B of Table 1). Overall, bigger firms tend to have annual reports that are more difficult to read, as evidenced by the correlation coefficient of 0.007 and 0.060 between Fog and Kincaid and firm size (Panel B of Table 1). Growth firms (firms with higher market-to-book ratio) also have annual reports more difficult to read, with the Pearson correlation coefficients between market-to-book and Fog and Kincaid being 0.014 and 0.034. Firms with annual reports more difficult to read tend to have more negative special items, as evidenced by the negative and significant correlation between special items and Fog and Kincaid.

I next explore the determinants of annual report readability in a multivariate regression setting. In Panel C of Table 1, I show the regression results of Fog and Kincaid Indices on size, market-to-book, special items, and year and 2-digit SIC code industry fixed effects (the coefficients on year and industry dummies are not shown in the table). There is no strong relation between the Fog Index and size and market-to-book after other factors are controlled: The coefficients on both are insignificant (t-statistics of -0.26 and -0.03 after clustering the standard errors at 2-digit SIC code level). On the other hand, size is still positively and significantly related to Kincaid (t-statistic of 4.18). However, the incremental R-squared's of the two variables are only 0.002 and 0.004, suggesting that they are not driving significant variation in the annual report readability. Special items are still negatively and significantly correlated with Fog and Kincaid in the multivariate regressions (t-statistics of -3.36 and -4.51 respectively).

Combined together, size, market-to-book, special items, and the industry and year effects

can explain about 8% of the variation in the Fog and Kincaid Indices. The majority of this explanatory power comes from industry dummies (incremental R-squares of about 4% to 6%). Thus, it is important to control for industry effects in later tests. From Panel D of Table 1, the five 2-digit SIC industries with the highest Fog are Insurance Agents (2-digit SIC code 64), Health Services (80), Insurance Carriers (63), Electric and Gas (49), and Building Construction (15); the five industries with the lowest Fog are Stone, Clay, Glass, and Concrete Products (32), Transportation by Air (45), Leather and Leather Products (31), Apparel and Accessory Stores (56), and Food and Kindred Products(20).

Panel E shows the persistence of annual report readability for firms in the first and fifth quintiles of Fog and Kincaid. Every year, firms are sorted into five quintiles based on Fog or Kincaid. For firms in the first and fifth quintiles, I track their readability level in the next three years. For instance, there are 11,479 (100%) firm-years in the fifth quintile of Fog in year 0. In the next year, 44.60% of these firms still remain in the fifth quintile, 24.57% switch to quintile 4, 14.17% are in quintile 3, 10.00% are in quintile 2 and 6.65% go to quintile 1. Overall, there seems to be some time-series variation in annual report readability. Of the firms in the fifth quintile of Fog in year 0, only about 61% stay in quintiles 4 and 5 and the rest belong to the first three quintiles in year 3.

4 Empirical Results

4.1 Current Earnings and Annual Report Readability

I first check the relation between firm performance and annual report readability. Table 2 shows the results of regressing Fog and Kincaid indices on earnings (scaled by book value of assets). In columns (1) and (2) of Table 2 Panel A, a univariate regression is carried out. The negative coefficients on earnings indicate that firms with higher earnings have annual reports that are easier to read (i.e., lower Fog and Kincaid indices). The t-statistics, which are adjusted for clustering at industry level (about 70 clusters), show that the effects are significant (e.g., t-statistics of -4.72 for column (1)). As can be seen from columns (3) and

(4), the negative coefficients remain similar in magnitude and statistically significant even if size, market-to-book, industry and year fixed effects are controlled.

In columns (5) and (6), I use a profit/loss dummy, which equals one if a company reports a profit and zero otherwise, instead of earnings level. The results show that the annual reports of loss firms are harder to read than those of profit firms.

The R-squared of the regressions, however, is trivial (0.00 in the simple regression and 0.08 with control variables and fixed effects). This suggests that economic performance is not a first-order determinant of annual report readability. To gauge the economic size of the effects, I do the following calculation. On average, increasing a firm's earnings from .00 (25th percentile of the sample) to 0.11 (75th percentile) will lead to a decrease in Fog Index of about 0.06. This is small compared to the variation of Fog in the sample (Table 1). Put differently, the annual reports of firms at 25th percentile of earnings have about 0.15 more syllables per word or about 0.15% more complex words than those of firms at 75th percentile. The Fog (Kincaid) Index of loss firms is higher than that of profit firms by 0.19 (0.24), which is also small.

The results hold in a change specification. Firms that experienced an increase in earnings tend to write their annual report in a more readable way than last year. In Panel B of Table 2, when the control variables and fixed effects are included, year-to-year change in earnings is negatively related to change in Fog and Kincaid Indices (columns (3) and (4)). Columns (5) and (6) show that, on average, the change in Fog (Kincaid) index of firms with an increase in earnings is 0.093 (0.088) lower than those with a decrease in earnings.

To summarize, I find that firms with better performance have annual reports that are harder to read. The effects are statistically significant, but the economic magnitude is small. This is consistent with the marginal benefit of making annual reports more complex to hide poor current performance is small.

4.2 Future Earnings and Annual Report Readability

I further examine the implication of annual report readability for future performance and earnings persistence. Management opportunism suggests that when annual reports are harder to read, good news may be more transitory and bad news may be more persistent.

I find that, indeed, the profits of firms with higher Fog Index and Kincaid Index are less persistent than those with lower indices. Table 3 presents the results. Columns (1) to (4) of Panel A show the regression results of one-year to four-year ahead earnings on this year's earnings, Fog, and their interaction using a sample of all firm-years with positive earnings. The interaction term captures the change in earnings persistence as annual report readability changes. In all cases, the interaction term is negative and statistically significant with the standard errors clustered at industry-level. This means that, as Fog goes up (i.e., annual reports become harder to read), the one-year to four-year earnings persistence becomes smaller for profitable firms.

In addition to size and market-to-book, I also control for the absolute amount of accruals, because it is an important determinant of earnings persistence (Sloan (1996)). The results (Columns (5) to (8)) become slightly stronger if the control variables are added.

To gauge the economic significance, I compare the impact of annual report readability on earnings persistence with that of accruals. Everything else equal, for an inter-quartile increase in Fog (an increase from 18.44 to 20.16), the earnings persistence of profitable firms goes down by 0.06 (calculated as -0.034 * (20.16-18.44), where -0.034 is from Column (5) of Table 3 Panel A.). Untabulated results also indicate that, on average, firms with Fog Index greater than 18 has an earnings persistence lower than those with Fog Index less than 14 by 0.12. An inter-quartile increase in the absolute amount of accruals, on the other hand, will lower the earnings persistence by about 0.05.

This suggests that the Fog Index has economically significant implications for the persistence of earnings of profitable firms. Panel B of Table 3 confirms the results using Kincaid Index.

On the other hand, I find little evidence that the annual report readability affects the

persistence of losses. As can be seen from Table 4, Fog Index and Kincaid Index have no impact on the persistence of losses in the next three years. Both the coefficient size and the t-statistics (columns (1) to (3) in Table 4) are close to zero. However, the interaction of earnings and Fog Index is positive and significant (0.023 with a t-statistics of 2.25, from column (8) of Panel A) when four-year ahead earnings is the dependent variable, suggesting that the losses are more persistent in a four-year horizon for firms with higher Fog Index. Evidence from Kincaid Index (Panel B) is qualitatively similar.

In summary, firms seem to use more complicated language in their annual reports to present good news that are less persistent. On the other hand, I do not find significant evidence that firms make their annual report more difficult to read to hide more persistent bad news.

4.3 Stock Returns and Annual Report Readability

Finally, I examine whether the stock market understand the implication of 10-K report readability for future earnings. I regress the 12-month, 24-month, 36-month, and 48-month stock returns following the 10-K filing date on Fog and Kincaid Indices and their interactions with current year earnings.

Table 5 shows the Fama-MacBeth regression results. The evidence indicates that there is no significant association between annual report readability and future stock returns in the next four years. Not only is none of the coefficient on Fog and Kincaid statistically significant, the economic magnitude is also tiny. For instance, an inter-quartile change in Fog is associated with 1% higher one-year ahead stock returns.

Since the annual report readability indices only affect earnings persistence for profit firms, I also do the return tests using only firm-years that report profits. The evidence (Panel B of Table 5) is essentially the same: There seems to be no association between the Fog and Kincaid Indices and future stock returns. Unreported results show that if size and book-to-market and are controlled, Fog and Kincaid Indices still have no predictive power for future stock returns. This is consistent with the stock market understanding the implication of

annual report readability for future earnings.

I further focus on two sub-samples: small firms (defined as firms with a market value lower than $2 billion) and firms with low institutional ownership (defined as firms with institutional ownership lower than 20%). To the extent that investors of these two set of firms are more likely to ignore the implication of annual report readability for future earnings, they provide more powerful setting to test the hypothesis. Unreported results show that there is some positive relation between Fog and Kincaid Indices and future stock returns for these sub-samples. The coefficients, however, are mostly insignificant and are only marginally significant in a couple of cases.

Overall, I conclude that there is no systematic evidence of stock market investors not understanding the implication of annual report readability for future performance.

5 Further Discussion

In this section, I discuss the findings and rule out some alternative explanations. Although used in many previous studies in other fields, the Fog and Kincaid Indices are still primitive and capture readability and understandability with errors. It is therefore critical to search for and rule out omitted variables extensively.

5.1 Litigation Explanation

The litigation hypothesis (Skinner (1994) and Skinner (1997)) argues that firms with poor performance will disclose more because of litigation risk. Some argue that disclosure based on more plain English is more likely to be involved in litigation than "sophisticated" disclosure with more legal jargon (Bencivenga (1997)). To the extent that this is true, the litigation concern of firms may be an alternative explanation for my findings: Firms with poor current or future performance are more likely to use more sophisticated language in disclosure to avoid potential lawsuits.

This is unlikely to be the main reason for the findings. First, there seems to be no

empirical evidence in support of the litigation consequences of plain English disclosure. In fact, the SEC has emphasized in many occasions that there is no reason to believe simple and clear language leads to higher chances of being involved in litigation. Second, my main empirical finding is that the positive earnings of firms with simpler (more complex) annual reports are more persistent (transitory). That is, my results come from the profitable firms. Untabulated results show that more than 90% of these firms still report a profit in the next year and more than 80% of them remain profitable every year in the next one to four years. It seems unlikely that litigation is a first-order concern for these firms. Finally, to the extent that litigation is industry-related, the fact that industry fixed effects are controlled in the tests further mitigates this concern.

5.2 Alternative Measures of Readability

As a robustness check, I measure the readability of the forward-looking statements in the annual reports. There might be heterogeneity in readability in the different components of annual reports (e.g., sales and marketing description versus the footnotes to the financial statements). Hence, one concern of measuring the readability of the whole annual report is that it is dominated by the variation of some routine statements. To mitigate this concern, I extract all the forward-looking sentences from the annual report, defined as those that contain words such as "may", "will", "expect", "plan", "intend", "anticipate", "believe", "estimate", "predict", "potential", and "continue".[8] These forward-looking statements are likely to be important and not just routine in nature.

Untabulated results show that, compared with the readability of the whole documents, the readability of the forward-looking statements is lower. The mean (median) of the Fog is 23.4 (22.7), significantly higher than that of the whole document. The variation in forward-looking statement Fog is also much bigger, with a standard deviation of 2.96 and an inter-quartile range of about 3.3. As indicated in Panel A of Table 6, there is a negative and significant correlation between the readability and profit persistence, confirming the main

[8] Confusing strings such as "stock option plan" are not counted as part of the forward-looking statements.

findings.

Second, I use the length of the annual report (i.e., number of words in the document) as an alternative measure of readability. Everything else equal, longer documents seem more deterring and more difficult to read. The SEC has consistently suggest companies avoid lengthy sentences and documents (SEC (1998)). Practitioners also use lengthy document as an example of bad and complex disclosure (e.g., Barker (2002)). There are pros and cons of using the length of a document as a measure of readability. The advantage is that it is easy to calculate and understand. Compared with the readability indices, the disadvantage of the document length as a measure of readability is that it is more likely to be correlated with the complexity of the disclosure contents. As a robustness check, the results (Panel B of Table 6) indicate a negative correlation between the length of annual reports and profit persistence, further confirming the findings in previous sections.

5.3 Earnings Volatility

Earnings persistence could be mechanically related to earnings volatility, since firms with more volatile earnings by definition have lower earnings persistence. Therefore, one concern is that the readability measures are somehow correlated with earnings volatility. For instance, to the extent that firms with more volatile income numbers have more complicated annual reports, the results could be driven by the omitted earnings volatility in my empirical tests.

To rule out this explanation, I calculate the earnings volatility for each firm-year using earnings numbers from the last 10 years. This shrinks the sample size by about 40%. I then include earnings volatility and its interaction with earnings in the regressions of future earnings on current earnings. Results (in Panel C of Table 6) remain pretty much unchanged.

5.4 Omitted Firm Characteristics (Firm Fixed Effects)

A more general concern is that some firm characteristics drive both the annual report readability and earnings persistence. Do the measures of annual report readability simply proxy for some omitted firm-specific variables? In the empirical tests, I already control for year

and industry fixed effects and some common firm characteristics (such as size and growth opportunity). This should mitigate the concerns. For further robustness check, I include firm fixed effects in the tests. I construct a panel data by keeping firms with at least 10 years of data. I then redo the tests by adding firm dummies in the regressions. Thus, essentially, only the time-series variation in the variables is explored. The results in Panel D of Table 6 show that there is a significantly negative relation between readability and profits persistence in the next two years. While still negative, this relation becomes statistically insignificant for earnings in years t+3 and t+4. Overall, the evidence suggests that the results in the paper are not likely to be driven simply by some omitted firm characteristics.

5.5 Non-linearity in Earnings Persistence

Do the results on future performance simply proxy for any nonlinearity in the persistence of earnings? Since the readability of annual reports are correlated with current earnings, it is possible that the findings on the relation between annual report readability and earnings persistence simply capture the nonlinearity in earnings persistence. To rule out this possibility, I put earnings-squared as an additional explanatory variable. Results in Panel E of Table 6 show that the results are slightly stronger both statistically and economically.

5.6 Omitted Firm-year Characteristics (Special Items)

Finally, could it be the case that the readability constructs proxy for some unusual events in the firm-year? In particular, Section 3 shows that there is a negative relation between Fog/Kincaid and the amount of special items. This means that, firms with more complex annual reports tend to report more negative special items, suggesting the possibility of unusual events (e.g., assets write-off) associated with higher Fog/Kincaid. Dechow and Ge (2005) find that the low persistence of earnings in low accrual firms is primarily driven by balance sheet adjustments relating to special items. They also show that low accrual firms with special items have performed poorly, are financially distressed, and have declines in investor recognition measured by analysts coverage and institutional holdings.

Hence, in this section, I examine whether special items are driving the empirical findings. Before discussing the additional tests, it is worth noting that ex ante special items seem unlikely to be the factor that drive the results. This is because the main finding is based on the profit firm-years and the effects of special items mainly come from loss firms or low accrual firms.

Nevertheless, I do the following robustness checks. First, I add special items (SI) and its interaction with current earnings in the regressions. This is motivated by (1) the strong negative correlation documented in Section 3 and (2) the effect of special items on earnings persistence documented in Dechow and Ge (2005). Results in Panel F of Table (6) confirm that the empirical findings are not sensitive to the inclusion of SI.

Second, I examine a sub-sample of firm-years which have a zero amount of special items. This requirement shrinks the sample of profit firm-years by more than 50%. Results (Panel G of Table 6) based on this sub-sample are similar to the main results.

6 Conclusions

I empirically study the implications of annual report readability for current performance, earnings persistence, and stock returns. In doing so, this study sheds some light on the relevance of the SEC plain English disclosure regulation. The empirical findings can be summarized as follows. First, annual reports of firms with poor performance are harder to read. The effect is statistically, but not economically, significant. Second, the profits of firms with annual reports that are easier to read are more persistent in the next one to four years. The effect is economically significant: An inter-quartile change in annual readability has about the same impact on profit persistence as accruals. This suggests that managers may be opportunistically choosing the readability of annual reports to hide information from investors. Finally, the lack of correlation between annual report readability and future stock returns suggests that, on average, the stock market understands this implication.

Taken together, the evidence in this paper suggests that, consistent with the motivation

behind the plain English disclosure regulation of the Securities and Exchange Commission, managers may be opportunistically structuring the annual reports to hide adverse information from investors. Contrary to the SEC's concerns, small investors may not be affected by the lack of readability, since the stock market seems to impound the implication of annual report readability into prices.

Appendix: Steps to Calculate the Readability Indices

This Appendix explains the details of calculating the readability indices starting from the raw 10-K filings used in this paper. I first download the 10-K report from Edgar and do the following editing before further analysis. First, the heading information that is contained between <SEC-HEADER> and </SEC-HEADER> is deleted. Second, all the tables that begin with <TABLE> and end with </TABLE> or the paragraphs that contain <S> or <C> are deleted, because <S> and <C> tags used by some firms in presenting tables. Next, all the paragraphs that contain string such as </TEXT>, </DOCUMENT>, <PAGE>, <TYPE> or /PRIVACY-ENHANCED/ are deleted. All the special characters in the format of <...> and <&..>, which are used widely in documents in SEC HTML format are replaced with blanks. Finally, to make sure that all tables, tabulated texts, or financial statements are excluded, all the paragraphs with more than 50% of non-alphabetic characters (e.g., white spaces or numbers) are deleted.

The file after the editing is then analyzed using the Fathom package in Perl. The package can calculate the typical text statistics, including number of characters, number of words, percent of complex words (i.e., words with more than three syllables), number of sentences, number of text lines, number of paragraphs, syllables per word and words per sentence. Based on the statistics, the package also produce the summary readability indices used in the paper.

References

Baker, H.E., and D.D. Kare, 1992, Relationship between annual report readability and corporate financial performance, *Management Research News* 15.

Barker, Robert, 2002, A three-point plan for SEC reform, *Business Week Online*.

Barnett, A., and K. Leoffler, 1979, Readability of accounting and auditing messages, *Journal of Business Communication* 16.

Bencivenga, Dominic, 1997, Short cut for investors: Why read a prospectus when a profile will do?, *New York Law Journal*.

Clatworthy, Mark, and Michael J. Jones, 2001, The effect of thematic structure on the variability of annual report readability, *Accounting, Auditing & Accountability Journal* 14.

Cohen, Daniel, 2005, Financial reporting quality: Determinants and economic consequences, Working Paper, New York University.

Collins-Thompson, Kevyn, and Jamie Callan, 2005, Predicting reading difficulty with statistical language models, *Journal of the American Society for Information Science and Technology* 56.

Courtis, John K., 1986, An investigation into annual report readability and corporate risk return relationships, *Accounting and Business Research*.

Dechow, Patricia, and Weili Ge, 2005, The persistence of earnings and cash flows and the role of special items: Implications for the accrual anomaly, *Review of Accounting Studies,* forthcoming.

Dechow, Patricia M., and Catherine M. Schrand, 2004, *Earnings Quality* (Research Foundation of CFA Institute: Charlottesville, Virginia) first edn.

Fields, Thomas D., Thomas Z. Lys, and Linda Vincent, 2001, Empirical research on accounting choice, *Journal of Accounting and Economics* 31.

Firtel, Kenneth B., 1999, Plain English: A repraisal of the intended audience of disclosure under the securities act of 1933, *Southern California Law Review* 72, 851–897.

Francis, Jennifer, Ryan LaFond, Per Olsson, and Katherine Schipper, 2005, The market pricing of accruals quality, *Journal of Accounting and Economics*.

Francis, Jennifer, Dhananjay Nanda, and Per Olsson, 2005, Voluntary disclosure, information quality, and costs of capital, Working Paper, Duke University.

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

Healy, Paul, 1977, Can you understand the footnotes to financial statements?, *Accountants Journal*.

Healy, Paul M., and Krishna G. Palepu, 2001, Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature, *Journal of Accounting and Economics* 31.

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

Lang, Mark, and Russell Lundholm, 1993, Cross-sectional determinants of analyst ratings of corporate disclosures, *Journal of Accounting Research* Autumn, 246–271.

Lebar, Mary Ann, 1982, A general semantics analysis of selected sections of the 10-k, the annual report to shareholders, and the financial press release, *The Accounting Review* 57.

Muresan, Gheorghe, Michael Cole, Catherine L. Smith, Lu Liu, and Nicholas J. Belkin, 2006, Does familiarity breed content ? taking account of familiarity with a topic in personalizing information retrieval, *Proceedings of the Hawaii International Conference on System Sciences (HICSS-39)*.

Riedl, Edward J., and Suraj Srinivasan, 2005, The strategic reporting of special items, Working Paper, Harvard University and University of Chicago.

SEC, 1998, *A plain English handbook: how to create clear SEC disclosure documents* (U.S. Securities and Exchange Commission: Washington, DC).

Skinner, Douglas J., 1994, Why firms voluntarily disclose bad news, *Journal of Accounting Research* 32.

———, 1997, Earnings disclosures and stockholder lawsuits, *Journal of Accounting and Economics* 23.

Sloan, Richard, 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings?, *Accounting Review* 71, 289–315.

Smith, James E., and Nora P. Smith, 1971, Readability: A measure of the performance of the communication function of financial reporting, *The Accounting Review* 46.

Soper, Fred J., and Robert Dolphin, 1964, Readability and corporate annual reports, *The Accounting Review* 39.

Subramanian, R., R.G. Insley, and R.D. Blackwell, 1993, Performance and readability: A comparison of annual reports of profitable and unprofitable corporations, *Journal of Business Communication* 30.

Verrecchia, Robert E., 1983, Discretionary disclosure, *Journal of Accounting and Economics* 5, 179–194.

Figure 1 A: Median Fog and Kincaid Indices by the Calendar Year of the File Dates (all firms)

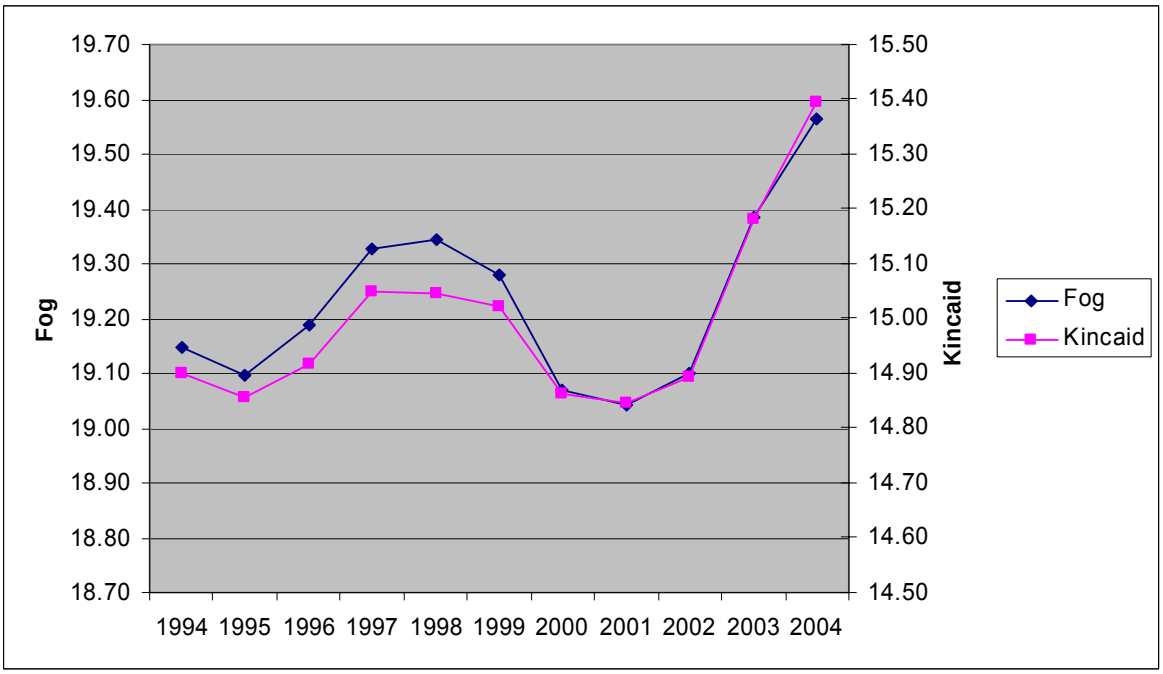

Figure 1 B: Median Fog and Kincaid Indices by the Calendar Year of the File Dates (constant sample)

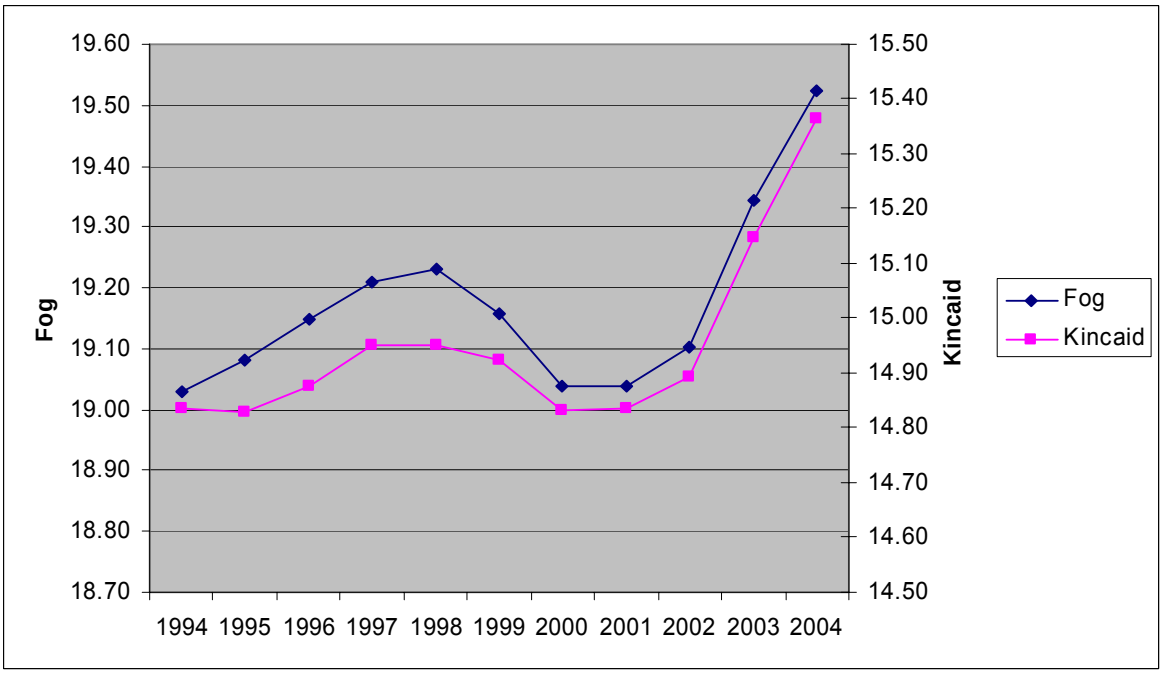

Table 1 Panel A: Summary Statistics

Variable	Mean	Median	Std. Dev.	1st	25th	75th	99th	Obs
Year	-	2000	-	1994	1997	2002	2004	55,719
Fog	19.39	19.24	1.44	16.61	18.44	20.16	23.64	55,719
Kincaid	15.19	15.01	1.42	12.61	14.25	15.90	19.57	55,719
Fog(t)-Fog(t-1)	0.05	0.02	1.46	-3.93	-0.59	0.65	4.34	44,097
Kincaid(t)-Kincaid(t-1)	0.06	0.03	1.48	-3.94	-0.55	0.64	4.36	44,097
Earnings	0.02	0.05	0.19	-0.75	0.00	0.11	0.33	55,719
Market-to-book	2.02	1.30	2.94	0.54	1.03	2.04	11.62	51,297
Market value of equity ($MM)	2022	169	14,209	1	44	731	33,003	51,393
Book value of assets ($MM)	3551	271	24,875	3	67	1,092	57,100	55,719

Note: This Panel shows the summary statistics of some the variables in the paper. Year is the calendar year in which an annual report is filed to the SEC Edgar system. Fog is the Fog Index calculated as (words per sentence + percent of complex words) * 0.4. Kincaid is the Kincaid Index calculated as (11.8 * syllables per word) + (0.39 * words per sentence) - 15.59. Earnings is operating earnings (data178 of Compustat) scaled by assets. Market-to-book is the market value of the firm divided by its book value ((data25*data199+data181)/data6). Market value of equity is calculated as (data25*data199). Size is logarithm of market value of equity calculated as Log(data25*data199). Book value of assets is data6 from Compustat.

Table 1 Panel B: Pearson Correlation Coefficient

	Fog	Kincaid	Market-to-book	Size	Assets	Special items
Fog	1					
Kincaid	**0.973**	1				
Market-to-book	**0.014**	**0.034**	1			
Size	0.007	**0.060**	**0.169**	1		
Assets	**0.017**	**0.027**	**-0.027**	**0.265**	1	
Special items	**-0.028**	**-0.034**	**-0.016**	**0.076**	0.019	1

Note: This Panel shows the Pearson correlation coefficient of Fog and Kincaid with firm characteristics. Fog is the Fog Index calculated as (words per sentence + percent of complex words) * 0.4. Kincaid is the Kincaid Index calculated as (11.8 * syllables per word) + (0.39 * words per sentence) - 15.59. Market-to-book is the market value of the firm divided by its book value ((data25*data199+data181)/data6). Market value of equity is calculated as (data25*data199). Size is logarithm of market value of equity calculated as Log(data25*data199). Book value of assets is data6 from Compustat. Special items is data17 from Compustat, scaled by book value of assets.

The Pearson correlation coefficient in bold is significant at 0.01 level.

Table 1 Panel C: Regression of Fog and Kincaid on Size, Market-to-Book and Year and Industry Fixed Effects

	Fog		Kincaid	
	Coef.	T-stat	Coef.	T-stat
Size	-0.002	-0.26	0.034	4.18
Market-to-book	0.000	-0.03	0.001	0.17
Special items	-0.254	-3.36	-0.344	-4.51
Year fixed effects (omitted)				
2-digit SIC industry fixed effects (omitted)				
R-squared	0.077		0.075	
Incremental R-squared of Size and Market-to-book	0.002		0.004	
Incremental R-squared of Special items	0.002		0.002	
Incremental R-squared of year fixed effects	0.018		0.024	
Incremental R-squared of industry fixed effects	0.057		0.044	

Note:

This Panel shows the regression results of Fog and Kincaid Indices on size, market-to-book, year fixed effects and 2-digit SIC industry fixed effects. The industry fixed effects coefficients are omitted from the table. Fog is the Fog Index calculated as (words per sentence + percent of complex words) * 0.4. Kincaid is the Kincaid Index calculated as (11.8 * syllables per word) + (0.39 * words per sentence) - 15.59. Market-to-book is the market value of the firm divided by its book value ((data25*data199+data181)/data6). Size is calculated as (log(data25*data199)). Special items is data17 from Compustat, scaled by book value of assets.

T-statistics are based on standard errors clustered at two-digit SIC industry code level.

Table 1 Panel D: Five Industries with the Lowest and Highest Fog and Kincaid Index

Five industries with the lowest Fog/Kincaid	SIC	Fog	SIC	Kincaid
Lowest	32	Stone, Clay, Glass, and Concrete Products	32	Stone, Clay, Glass, and Concrete Products
	45	Transportation by Air	45	Transportation by Air
	31	Leather and Leather Products	54	Food Stores
	56	Apparel and Accessory Stores	42	Motor Freight Transportation and Warehousing
	20	Food and Kindred Products	31	Leather and Leather Products
Five industries with the highest Fog/Kincaid	15	Building construction	48	Communications
	49	Electric, Gas, and Sanitary Services	73	Business Services
	63	Insurance Carriers	49	Electric, Gas, and Sanitary Services
	80	Health Services	64	Insurance Agents
Highest	64	Insurance Agents	80	Health Services

Note: This Panel shows the five 2-digit SIC industries that have the highest Fog and Kincaid and the five industries that have the lowest Fog and Kincaid. Industries with fewer than 100 firm-years in the sample are not included. Fog is the Fog Index calculated as (words per sentence + percent of complex words) * 0.4. Kincaid is the Kincaid Index calculated as (11.8 * syllables per word) + (0.39 * words per sentence) - 15.59.

Table 1 Panel E: Persistence of Fog and Kincaid

Fog	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	N
Year 0					100 (%)	11479
Year 1	6.65	10.00	14.17	24.57	44.60	9006
Year 2	7.34	11.49	16.58	24.96	39.63	7303
Year 3	8.54	12.31	17.51	25.12	36.53	5729
Year 0	100 (%)					11479
Year 1	56.87	20.69	9.30	6.21	6.94	9214
Year 2	50.99	22.03	11.29	7.69	8.00	7540
Year 3	47.16	22.91	11.42	9.42	9.09	6016
Kincaid						
Year 0					100 (%)	11479
Year 1	7.40	10.73	14.97	25.00	41.91	9060
Year 2	8.16	12.22	17.22	25.58	36.81	7356
Year 3	8.80	13.51	17.75	26.08	33.86	5759
Year 0	100 (%)					11479
Year 1	57.37	19.43	9.16	6.49	7.55	9151
Year 2	51.97	20.80	10.69	8.14	8.41	7458
Year 3	47.76	21.76	11.64	9.23	9.60	5946

Note: This Panel shows the transition matrix of Fog and Kincaid across quintiles for firms in the 1st and 5th quintiles. Fog is the Fog Index calculated as (words per sentence + percent of complex words) * 0.4. Kincaid is the Kincaid Index calculated as (11.8 * syllables per word) + (0.39 * words per sentence) - 15.59. Each year (year 0), firms are sorted into quintiles based on Fog or Kincaid. In the next three years (year 1 to year 3), the percentages by quintiles for firms that are in the 1st and 5th quintiles in year 0 are calculated and tabulated.

Table 2 Panel A: Firm Performance and Annual Report Readability (level specification)

	(1)	(2)	(3)	(4)	(5)	(6)
	Fog	Kincaid	Fog	Kincaid	Fog	Kincaid
Earnings	-0.525	-0.563	-0.5	-0.62		
	(-4.72)***	(-4.72)***	(-3.72)***	(-5.04)***		
Market-to-book			-0.004	-0.004	-0.003	-0.003
			(-1.19)	(-1.12)	(-0.93)	(-0.90)
Size			0.014	0.053	0.012	0.052
			(1.65)	(6.67)***	(1.47)	(6.47)***
Profit/Loss dummy					-0.188	-0.241
					(-3.52)***	(-4.52)***
Constant	19.395	15.198	19.229	14.791	19.374	14.975
	(318.99)***	(253.54)***	(389.01)***	(299.35)***	(380.29)***	(322.85)***
Year dummy	NO	NO	YES	YES	YES	YES
Industry dummy	NO	NO	YES	YES	YES	YES
Observations	55719	55719	51291	51291	51291	51291
R-squared	0.00	0.01	0.08	0.08	0.08	0.08

Note: This Panel shows the regression results of the annual report readability on firm performance using the level specification. The dependent variables are Fog and Kincaid. Fog is the Fog Index calculated as (words per sentence + percent of complex words) * 0.4. Kincaid is the Kincaid Index calculated as (11.8 * syllables per word) + (0.39 * words per sentence) - 15.59. Earnings is operating earnings (data178 of Compustat) scaled by assets. Profit/Loss dummy is a dummy variable that equals 1 if a company reports a profit and 0 otherwise. Market-to-book is the market value of the firm divided by its book value (data25*data199+data181)/data6). Size is calculated as (log(data25*data199)).

T-statistics in parentheses are based on standard errors clustered at two-digit SIC industry code level.

Table 2 Panel B: Firm Performance and Annual Report Readability (change specification)

	(1) ΔFog	(2) ΔKincaid	(3) ΔFog	(4) ΔKincaid	(5) ΔFog	(6) ΔKincaid
Change in earnings	-0.107	-0.063	-0.169	-0.14		
	(-1.32)	(-0.72)	(-2.46)**	(-1.88)*		
Market-to-book			-0.005	-0.003	-0.004	-0.001
			(-2.05)**	(-1.07)	(-1.75)*	(-0.60)
Size			0.001	0.005	0.002	0.006
			(0.52)	(2.00)**	(0.80)	(2.33)**
Earnings increase/decrease dummy					-0.093	-0.088
					(-4.45)***	(-3.74)***
Constant	0.048	0.065	-0.063	-0.084	0.114	0.083
	(6.33)***	(8.93)***	(-1.30)	(-1.76)*	(2.39)**	(1.73)*
Year dummy	NO	NO	YES	YES	YES	YES
Industry dummy	NO	NO	YES	YES	YES	YES
Observations	43819	43819	40827	40827	41094	41094
R-squared	0.00	0.00	0.01	0.01	0.01	0.01

Note: This Panel shows the regression results of the annual report readability on firm performance using the change specification. The dependent variables are changes in Fog (ΔFog) and Kincaid (ΔKincaid). Fog is the Fog Index calculated as (words per sentence + percent of complex words) * 0.4. Kincaid is the Kincaid Index calculated as (11.8 * syllables per word) + (0.39 * words per sentence) - 15.59. Earnings is operating earnings (data178 of Compustat) scaled by assets. Earnings increase/decrease dummy is a dummy variable that equals 1 if earnings is greater than previous year and 0 otherwise. Market-to-book is the market value of the firm divided by its book value (data25*data199+data181)/data6). Size is calculated as (log(data25*data199)).

T-statistics in parentheses are based on standard errors clustered at two-digit SIC industry code level.

Table 3 Panel A: Earnings Persistence and Fog Index (Profit Firm-years)

	(1) Earnings(t+1)	(2) Earnings(t+2)	(3) Earnings(t+3)	(4) Earnings(t+4)	(5) Earnings(t+1)	(6) Earnings(t+2)	(7) Earnings(t+3)	(8) Earnings(t+4)
Earnings	1.297	1.454	1.121	1.415	1.392	1.469	1.095	1.459
	(7.17)***	(4.54)***	(3.37)***	(3.52)***	(7.32)***	(4.23)***	(2.65)***	(2.71)***
Fog	0.002	0.003	0.002	0.003	0.003	0.004	0.002	0.004
	(2.59)**	(2.01)**	(0.84)	(1.36)	(3.72)***	(2.54)**	(1.07)	(1.66)
Earnings * Fog	-0.026	-0.043	-0.029	-0.048	-0.034	-0.05	-0.034	-0.057
	(-2.77)***	(-2.61)**	(-1.72)*	(-2.33)**	(-3.77)***	(-3.04)***	(-1.78)*	(-2.42)**
Abs(accruals)					-0.012	-0.074	-0.052	-0.02
					(-0.79)	(-2.27)**	(-0.73)	(-0.36)
Abs(accruals) * Earnings					-0.701	-0.818	-0.541	-0.866
					(-6.54)***	(-4.92)***	(-1.59)	(-3.15)***
Size					0.002	0.002	0.004	0.004
					(2.49)**	(1.12)	(2.10)**	(1.41)
Size * Earnings					0.018	0.036	0.029	0.035
					(2.12)**	(2.06)**	(1.34)	(1.28)
Market-to-book					-0.002	-0.004	-0.007	-0.004
					(-1.77)*	(-2.47)**	(-4.63)***	(-3.57)***
Market-to-book * Earnings					0.008	0.007	0.008	0.002
					(2.10)**	(1.91)*	(3.19)***	(0.63)
Constant	-0.018	-0.027	0.010	-0.012	-0.051	-0.054	-0.018	-0.055
	(-1.19)	(-0.88)	(0.27)	(-0.28)	(-3.24)***	(-1.65)	(-0.45)	(-1.07)
Year dummy	YES	YES	YES	YES	YES	YES	YES	YES
Industry dummy	YES	YES	YES	YES	YES	YES	YES	YES
Observations	34588	28808	23397	18370	27964	23466	19229	15233
R-squared	0.42	0.24	0.19	0.16	0.41	0.24	0.19	0.17

Note: This Panel shows the regression results on the effect of annual report readability on earnings persistence. The sample includes all firm-years that report profits. The dependent variables are earnings of year t+1 to year t+4. Fog is the Fog Index calculated as (words per sentence + percent of complex words) * 0.4. Earnings is operating earnings (data178 of Compustat) scaled by assets. Market-to-book is the market value of the firm divided by its book value ((data25*data199+data181)/data6). Size is calculated as (log(data25*data199)). Accruals is calculated as (data178-data308)/data6.

T-statistics in parentheses are based on standard errors clustered at two-digit SIC industry code level.

Table 3 Panel B: Earnings Persistence and Kincaid Index (Profit Firm-years)

	(1) Earnings(t+1)	(2) Earnings(t+2)	(3) Earnings(t+3)	(4) Earnings(t+4)	(5) Earnings(t+1)	(6) Earnings(t+2)	(7) Earnings(t+3)	(8) Earnings(t+4)
Earnings	1.169	1.233	1.033	1.295	1.285	1.315	1.031	1.345
	(8.71)***	(5.30)***	(4.19)***	(4.27)***	(7.96)***	(4.49)***	(3.11)***	(2.98)***
Kincaid	0.002	0.003	0.002	0.003	0.003	0.004	0.002	0.004
	(2.91)***	(2.22)**	(0.97)	(1.81)*	(3.78)***	(2.79)***	(1.17)	(1.98)*
Earnings * Kincaid	-0.024	-0.04	-0.031	-0.054	-0.036	-0.055	(-0.04)	-0.067
	(-2.80)***	(-2.66)***	(-1.99)*	(-2.72)***	(-3.92)***	(-3.29)***	(-2.20)**	(-2.85)***
Abs(accruals)					-0.012	-0.073	-0.05	-0.018
					(-0.78)	(-2.25)**	(-0.71)	(-0.32)
Abs(accruals) * Earnings					-0.702	-0.824	-0.544	-0.874
					(-6.75)***	(-5.02)***	(-1.58)	(-3.30)***
Size					0.002	0.001	0.004	0.004
					(2.39)**	(1.03)	(2.10)**	(1.39)
Size * Earnings					0.02	0.038	0.03	0.037
					(2.33)**	(2.23)**	(1.43)	(1.38)
Market-to-book					-0.002	-0.004	-0.007	-0.004
					(-1.77)*	(-2.47)**	(-4.63)***	(-3.58)***
Market-to-book * Earnings					0.008	0.007	0.008	0.002
					(2.09)**	(1.93)*	(3.19)***	(0.68)
Constant	-0.007	-0.008	0.016	-0.006	-0.037	-0.034	-0.008	-0.041
	(-0.71)	(-0.38)	(0.63)	(-0.19)	(-2.80)***	(-1.33)	(-0.24)	(-1.01)
Year dummy	YES	YES	YES	YES	YES	YES	YES	YES
Industry dummy	YES	YES	YES	YES	YES	YES	YES	YES
Observations	34588	28808	23397	18370	27964	23466	19229	15233
R-squared	0.42	0.24	0.19	0.16	0.41	0.24	0.19	0.17

Note: This Panel shows the regression results on the effect of annual report readability on earnings persistence. The sample includes all firm-years that report profits. The dependent variables are earnings of year t+1 to year t+4. Kincaid is the Kincaid Index calculated as (11.8 * syllables per word) + (0.39 * words per sentence) - 15.59. Earnings is operating earnings (data178 of Compustat) scaled by assets. Market-to-book is the market value of the firm divided by its book value ((data25*data199+data181)/data6). Size is calculated as (log(data25*data199)). Accruals is calculated as (data178-data308)/data6.

T-statistics in parentheses are based on standard errors clustered at two-digit SIC industry code level.

Table 4 Panel A: Earnings Persistence and Fog Index (Loss firm-years)

	(1) Earnings(t+1)	(2) Earnings(t+2)	(3) Earnings(t+3)	(4) Earnings(t+4)	(5) Earnings(t+1)	(6) Earnings(t+2)	(7) Earnings(t+3)	(8) Earnings(t+4)
Earnings	0.664	0.584	0.218	-0.292	0.889	0.746	0.387	-0.017
	(1.99)*	(1.84)*	(0.72)	(-1.00)	(2.39)**	(2.54)**	(1.13)	(-0.06)
Fog	-0.002	-0.003	-0.003	0.001	-0.003	-0.004	-0.003	0.000
	(-0.97)	(-1.07)	(-0.83)	(0.36)	(-1.13)	(-1.45)	(-1.08)	(0.02)
Earnings * Fog	-0.004	-0.008	0.006	0.032	-0.011	-0.013	0.002	0.023
	(-0.22)	(-0.48)	(0.41)	(2.25)**	(-0.56)	(-0.81)	(0.14)	(2.07)**
Abs(accruals)					0.074	0.103	0.008	0.068
					(2.15)**	(1.83)*	(0.19)	(0.97)
Abs(accruals) * Earnings					-0.516	-0.475	-0.636	-0.384
					(-6.16)***	(-3.41)***	(-6.55)***	(-3.56)***
Size					0.003	0.005	0.003	0.007
					(1.46)	(1.61)	(0.80)	(1.82)*
Size * Earnings					0.039	0.039	0.026	0.012
					(5.19)***	(5.25)***	(1.91)*	(0.65)
Market-to-book					-0.005	-0.011	-0.007	-0.007
					(-2.32)**	(-5.89)***	(-2.42)**	(-2.23)**
Market-to-book * Earnings					-0.022	-0.023	-0.013	-0.015
					(-4.03)***	(-3.36)***	(-1.74)*	(-1.98)*
Constant	0.038	0.068	0.056	0.018	0.052	0.082	0.088	0.019
	(073)	(1.11)	(0.76)	(0.23)	(0.95)	(1.3)	(1.24)	(0.25)
Year dummy	YES	YES	YES	YES	YES	YES	YES	YES
Industry dummy	YES	YES	YES	YES	YES	YES	YES	YES
Observations	9372	6996	4955	3410	8960	6691	4735	3256
R-squared	0.34	0.23	0.2	0.21	0.37	0.26	0.22	0.22

Note: This Panel shows the regression results on the effect of annual report readability on earnings persistence. The sample includes all firm-years that report losses. The dependent variables are earnings of year t+1 to year t+4. Fog is the Fog Index calculated as (words per sentence + percent of complex words) * 0.4. Earnings is operating earnings (data178 of Compustat) scaled by assets. Market-to-book is the market value of the firm divided by its book value ((data25*data199+data181)/data6). Size is calculated as (log(data25*data199)). Accruals is calculated as (data178-data308)/data6.

T-statistics in parentheses are based on standard errors clustered at two-digit SIC industry code level.

Table 4 Panel B: Earnings Persistence and Kincaid Index (Loss Firm-years)

	(1) Earnings(t+1)	(2) Earnings(t+2)	(3) Earnings(t+3)	(4) Earnings(t+4)	(5) Earnings(t+1)	(6) Earnings(t+2)	(7) Earnings(t+3)	(8) Earnings(t+4)
Earnings	0.478	0.478	0.173	-0.106	0.795	0.698	0.399	0.17
	(1.74)*	(1.67)	(0.83)	(-0.50)	(2.65)**	(2.66)***	(1.77)*	(0.75)
Kincaid	-0.002	-0.004	-0.004	0.00	-0.003	-0.005	-0.005	-0.002
	(-0.79)	(-1.26)	(-1.33)	(0.15)	(-1.24)	(-1.76)*	(-1.77)*	(-0.60)
Earnings * Kincaid	0.007	-0.004	0.011	0.028	-0.007	-0.014	0.002	0.017
	(0.38)	(-0.18)	(0.82)	(2.21)**	(-0.38)	(-0.73)	(0.15)	(1.66)
Abs(accruals)					0.074	0.103	0.008	0.067
					(2.15)**	(1.84)*	(0.20)	(0.96)
Abs(accruals) * Earnings					-0.515	-0.475	-0.634	-0.385
					(-6.15)***	(-3.47)***	(-6.60)***	(-3.58)***
Size					0.004	0.005	0.003	0.007
					(1.55)	(1.71)*	(0.88)	(1.83)*
Size * Earnings					0.04	0.04	0.026	0.011
					(5.56)***	(5.49)***	(1.97)*	(0.61)
Market-to-book					-0.005	-0.011	-0.007	-0.007
					(-2.31)**	(-5.82)***	(-2.42)**	(-2.23)**
Market-to-book * Earnings					-0.022	-0.023	-0.013	-0.015
					(-4.00)***	(-3.31)***	(-1.73)*	(-1.97)*
Constant	0.026	0.069	0.064	0.048	0.046	0.082	0.097	0.049
	(0.54)	(1.25)	(1.14)	(0.87)	(1.00)	(1.48)	(1.80)*	(0.88)
Year dummy	YES	YES	YES	YES	YES	YES	YES	YES
Industry dummy	YES	YES	YES	YES	YES	YES	YES	YES
Observations	9372	6996	4955	3410	8960	6691	4735	3256
R-squared	0.34	0.23	0.2	0.21	0.37	0.27	0.22	0.22

Note: This Panel shows the regression results on the effect of annual report readability on earnings persistence. The sample includes all firm-years that report losses. The dependent variables are earnings of year t+1 to year t+4. Kincaid is the Kincaid Index calculated as (11.8 * syllables per word) + (0.39 * words per sentence) - 15.59. Earnings is operating earnings (data178 of Compustat) scaled by assets. Market-to-book is the market value of the firm divided by its book value ((data25*data199+data181)/data6). Size is calculated as (log(data25*data199)). Accruals is calculated as (data178-data308)/data6.

T-statistics in parentheses are based on standard errors clustered at two-digit SIC industry code level.

Table 5 Panel A: Fama-MacBeth Regressions of Future Returns on Fog Index

All sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ret (t+1)	Ret (t+2)	Ret (t+3)	Ret (t+4)	Ret (t+1)	Ret (t+2)	Ret (t+3)	Ret (t+4)
Fog	0.007	0.008	0.013	0.012	0.005	0.005	0.011	0.009
	(0.83)	(0.75)	(0.80)	(0.91)	(0.72)	(0.60)	(0.77)	(0.77)
Earnings					0.195	-0.451	0.151	-0.859
					(0.38)	(-0.70)	(0.14)	(-1.15)
Earnings*Fog					-0.020	0.018	-0.017	0.031
					(-0.78)	(0.60)	(-0.30)	(0.89)
Constant	0.071	0.070	-0.004	0.019	0.121	0.143	0.031	0.093
	(0.45)	(0.36)	(-0.01)	(0.09)	(0.93)	(1.12)	(0.12)	(0.48)
Number of years	10	9	8	7	10	9	8	7
Average observations	3897	3340	2973	2578	3890	3334	2968	2574
Average R-squared	0.00	0.00	0.00	0.00	0.02	0.01	0.01	0.01
Profit firm-years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ret (t+1)	Ret (t+2)	Ret (t+3)	Ret (t+4)	Ret (t+1)	Ret (t+2)	Ret (t+3)	Ret (t+4)
Fog	-0.001	0.004	0.008	0.012	-0.002	0.001	0.010	0.012
	(-0.16)	(0.65)	(0.99)	(1.27)	(-0.54)	(0.23)	(1.04)	(0.94)
Earnings					-0.222	-0.562	0.285	0.049
					(-0.48)	(-0.71)	(0.41)	(0.05)
Earnings*Fog					0.014	0.024	-0.017	-0.009
					(0.60)	(0.55)	(-0.46)	(-0.17)
Constant	0.199	0.120	0.048	-0.011	0.219	0.180	0.023	0.001
	(2.73)***	(1.01)	(0.29)	(-0.06)	(3.29)***	(1.40)	(0.12)	(0.01)
Number of years	10	9	8	7	10	9	8	7
Average observations	3024	2654	2438	2174	3022	2652	2436	2172
Average R-squared	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01

Note: The dependent variables are annual returns of year t+1 to year t+4, starting from the month after the annual report filing date. Fog is the Fog Index calculated as (words per sentence + percent of complex words) * 0.4. Kincaid is the Kincaid Index calculated as (11.8 * syllables per word) + (0.39 * words per sentence) - 15.59. Earnings is operating earnings (data178 of Compustat) scaled by assets.
T-statistics in parentheses are based on the coefficients from the annual cross-sectional regressions.

Table 5 Panel B: Fama-MacBeth Regressions of Future Returns on Kincaid Index

All sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ret (t+1)	Ret (t+2)	Ret (t+3)	Ret (t+4)	Ret (t+1)	Ret (t+2)	Ret (t+3)	Ret (t+4)
Kincaid	0.007	0.009	0.015	0.014	0.005	0.005	0.013	0.013
	(0.55)	(0.61)	(0.75)	(0.86)	(0.47)	(0.42)	(0.71)	(0.82)
Earnings					0.243	-0.320	0.004	-0.463
					(0.54)	(-0.62)	(0.00)	(-0.76)
Earnings*Kincaid					-0.029	0.015	-0.012	0.013
					(-1.14)	(0.45)	(-0.20)	(0.34)
Constant	0.109	0.097	0.011	0.038	0.146	0.157	0.048	0.075
	(0.62)	(0.47)	(0.04)	(0.18)	(1.01)	(0.99)	(0.18)	(0.35)
Number of years	10	9	8	7	10	9	8	7
Average observations	3897	3340	2973	2578	3890	3334	2968	2574
Average R-squared	0.00	0.00	0.00	0.00	0.02	0.01	0.01	0.01
Profit firm-years	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ret (t+1)	Ret (t+2)	Ret (t+3)	Ret (t+4)	Ret (t+1)	Ret (t+2)	Ret (t+3)	Ret (t+4)
Kincaid	-0.002	0.004	0.010	0.013	-0.004	0.000	0.011	0.014
	(-0.40)	(0.41)	(0.84)	(1.03)	(-0.63)	(0.04)	(0.80)	(0.83)
Earnings					-0.200	-0.554	0.194	0.132
					(-0.60)	(-0.89)	(0.29)	(0.14)
Earnings*Kincaid					0.016	0.030	-0.016	-0.018
					(0.78)	(0.69)	(-0.36)	(-0.26)
Constant	0.220	0.145	0.063	0.031	0.239	0.202	0.044	0.032
	(2.37)***	(1.07)	(0.35)	(0.17)	(2.40)***	(1.27)	(0.20)	(0.14)
Number of years	10	9	8	7	10	9	8	7
Average observations	3024	2654	2438	2174	3022	2652	2436	2172
Average R-squared	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Note: The dependent variables are annual returns of year t+1 to year t+4, starting from the month after the annual report filing date. Fog is the Fog Index calculated as (words per sentence + percent of complex words) * 0.4. Kincaid is the Kincaid Index calculated as (11.8 * syllables per word) + (0.39 * words per sentence) - 15.59. Earnings is operating earnings (data178 of Compustat) scaled by assets.
T-statistics in parentheses are based on the coefficients from the annual cross-sectional regressions.

Table 6: Additional Tests: Profit Firm-Years

	(1) Earnings(t+1)	(2) Earnings(t+2)	(3) Earnings(t+3)	(4) Earnings(t+4)
Panel A: Forward-looking statement readability				
Earnings	0.963	0.823	0.744	0.737
	(7.30)***	(3.68)***	(2.79)***	(2.10)**
Fog	0.001	0.001	0.001	0.002
	(2.85)***	(2.50)**	(1.89)*	(2.12)**
Earnings * Fog	-0.01	-0.015	-0.014	-0.017
	(-2.33)**	(-2.49)**	(-2.37)**	(-2.16)**
Panel B: Number of words in the annual report as a measure of readability				
Earnings	1.663	1.688	1.664	2.117
	(7.16)***	(4.24)***	(3.70)***	(4.27)***
Log(number of words)	0.006	0.008	0.007	0.014
	(3.14)***	(2.36)**	(1.90)*	(3.38)***
Earnings * Log(number of words)	-0.099	-0.129	-0.134	-0.191
	(-4.47)***	(-3.70)***	(-3.67)***	(-4.68)***
Panel C: Also control for earnings volatility (sub-sample)				
Earnings	1.484	1.527	1.228	1.611
	(9.08)***	(6.32)***	(4.03)***	(4.55)***
Fog	0.003	0.004	0.004	0.006
	(4.22)***	(4.16)***	(1.80)**	(2.50)**
Earnings * Fog	-0.037	-0.052	-0.037	-0.065
	(-4.56)***	(-4.66)***	(-2.45)**	(-3.66)***
Panel D: Also control for firm fixed effects (sub-sample)				
Earnings	1.224	0.905	0.485	0.316
	(5.81)***	(3.11)***	(1.47)	(0.73)
Fog	0.003	0.003	0.000	0.000
	(2.31)**	(1.64)	(0.08)	(0.18)
Earnings * Fog	-0.029	-0.033	-0.010	-0.011
	(-2.71)***	(-2.30)**	(-1.07)	(-0.68)
Panel E: Also control for earnings-squared				
Earnings	1.372	1.363	0.987	1.305
	(8.90)***	(4.88)***	(3.10)***	(3.26)***
Fog	0.003	0.004	0.002	0.004
	(4.07)***	(2.67)***	(1.03)	(1.81)*
Earnings * Fog	-0.033	-0.048	-0.031	-0.052
	(-4.13)***	(-3.29)***	(-1.88)*	(-2.78)***
Panel F: Also control for special items				
Earnings	1.369	1.458	1.035	1.36
	(7.48)***	(4.17)***	(2.46)**	(2.54)**

Fog	0.003	0.004	0.002	0.003
	(3.67)***	(2.37)**	(0.83)	(1.42)
Earnings * Fog	-0.032	-0.049	-0.030	-0.051
	(-3.78)***	(-2.92)***	(-1.56)	(-2.24)**
Panel G: Firms with no special items (sub-sample)				
Earnings	1.442	1.587	1.106	1.321
	(5.88)***	(3.45)***	(1.90)*	(1.86)*
Fog	0.003	0.005	0.003	0.004
	(2.65)***	(2.39)**	(0.92)	(1.44)
Earnings * Fog	-0.035	-0.053	-0.031	-0.047
	(-2.94)***	(-2.54)**	(-1.19)	(-1.65)

Note: All the regressions in this table are based on the sub-sample of profit firm-years. The dependent variables are earnings of year t+1 to year t+4. Fog is the Fog Index calculated as (words per sentence + percent of complex words) * 0.4. Earnings is operating earnings (data178 of Compustat) scaled by assets. Variables that are included as control variables but not reported in the table include Market-to-book (the market value of the firm divided by its book value (data25*data199+data181)/data6), Size, (log(data25*data199)), absolute amount of accruals (Abs(data178-data308)/data6) and their interactions with earnings. Year and two-digit SIC industry fixed effects are also included in all the regressions. Special items is data17 from Compustat, scaled by book value of assets (data6).

T-statistics in parentheses are based on standard errors clustered at two-digit SIC industry code level.