The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts

Reuven Lehavy
E-mail: rlehavy@umich.edu
Phone: (734) 763-1508

Feng Li
E-mail: feng@umich.edu
Phone: (734) 936-2771

Kenneth Merkley
E-mail: kmerkley@bus.umich.edu
Phone: (734) 936-0245

Stephen M. Ross School of Business
University of Michigan
701 Tappan Street
Ann Arbor, MI 48109-1234

First Version: June 2009
Current Version: October 2009

We would like to thank workshop participants from the University of Michigan, University of Minnesota, Erasmus University, and a joint presentation of HEC-Paris and INSEAD for their helpful comments and suggestions.

The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts

ABSTRACT

This paper examines the effect of the complexity of firm written communication on the behavior of sell-side financial analysts. Using a measure of the readability of corporate 10-K filings we document that analyst following, effort (measured as the length of time required for analysts to issue their first forecast revision following the 10-K filing), and the informativeness of their reports are positively related to firm communication complexity. Additionally, we find that communication complexity reduces earnings forecast accuracy and increases forecast dispersion. Overall, our results are consistent with the prediction of an increasing demand for analyst services for firms with more complex communication and a greater collective effort by analysts for firms with less readable disclosures. Our results contribute to our understanding the role of analysts as information intermediaries for investors, as well as to the long-standing debate about the intended audience of financial information and the effect of the complexity of written financial communication on the usefulness of this information.

1. **Introduction**

Over the past two decades changes in financial and reporting regulations (e.g., changes in segment disclosures, employee stock options, and Sarbanes-Oxley disclosures) have significantly increased the amount of required disclosures by firms to external users. In addition, technological advancement and new developments in financial engineering have made it more challenging for firms to communicate information about the underlying fundamentals of their businesses in a clear and informative manner. The increase in the amount of required disclosure accompanied by the challenging task of communicating increasingly complex business transactions to investors has led to concerns about the effectiveness of management communication and the ability of interested users to make informed decisions based on this information. As a result, both the SEC and the popular press have routinely criticized firms for the complexity of the language in their 10-K filings (Schroeder 2003). These groups have expressed concerns about the ability of financial statement users, especially small investors, to understand the complicated writing found in firms' financial documents (see, SEC: *A Plain English Handbook* 1998, Cox 2007, Schroeder 2003).

Given the increasing complexity of firm disclosures and the related concerns about their usability, a natural question arises as to the role of financial analysts in intermediating such information to investors. Specifically, as financial information intermediaries, do financial analysts use their expertise to examine this complex communication and provide useful information to financial statement users, or do they prefer to focus their efforts on firms with less complex communication? In this study we attempt to answer this question by examining how the behavior of financial analysts relates to firms' communication complexity.

Prior literature examines the relation between the properties of firms' financial disclosures and the behavior of security analysts. These properties include the informativeness of disclosures (Lang and Lundholm 1996; Healy, Hutton, and Palepu 1999), the use of segment disclosures (Botosan and

Harris 2000), and the effect of intangible assets (Barth, Kasznik, and McNichols 2001). These studies generally document that firms with better disclosure quality tend to attract greater analyst following. Other studies examine the effect of the complexity of certain financial reporting items on the properties of analyst earnings forecasts. These studies find that analyst forecasts are influenced by firms' accounting choices (Hopkins, Houston, and Peters 2000; Bradshaw, Miller, and Serafeim 2008), changes in the tax law (Plumlee 2003), the clarity of the income effects of specific financial items (Hirst and Hopkins 1998; Hirst, Hopkins and Wahlen 2003), and international diversification (Duru and Reeb 2002).

While each of these studies contributes to our understanding of the effect of firms' financial reporting and disclosure choices on analyst behavior, they generally focus on the effect of a single financial statement item (e.g., interest rates, taxes) or disclosure (e.g., segment reporting).[1] This approach, however, does not explicitly incorporate the potential effect of the properties of other financial statement items or disclosures on analyst following or the properties of their earnings forecasts. More important, this approach does not address the potential impact on analyst behavior of the *overall* complexity of firm financial communication, that is, the cost of processing and interpreting the entirety of firm disclosures. This potential limitation is particularly important given the broad nature of corporate disclosures such as 10-K filings that describe a significant number of interrelated financial items.

In this study we attempt to address this issue by examining the relation between a comprehensive measure of the overall readability (or communication complexity) of corporate 10-K filings and analyst behavior. This measure, known as the Gunning-Fog Index (hereafter Fog Index) incorporates the number of words per sentence and the number of complex words in a document

[1] One notable exception is Lang and Lundholm (1996) who use AIMR scores as a measure of overall disclosure quality. The potential limitations of this measure, however, are that it is based on a subjective survey of analysts, is limited to a subset of large firms, and is no longer available after 1995.

to derive a measure of the readability or communication complexity of firms' 10-K filings (see Li 2008, Courtis 1995, and Jones and Shoemaker 1994).[2] The Fog Index has been widely used in social science research for several decades to examine the relation between the complexity of written information and various decisions or outcomes.[3] This measure offers several important advantages. First, it allows us to study a large and diverse group of firms. Second, it is an objective measure, not based on analyst surveys or opinion, and can be calculated for any narrative disclosure. Finally, it allows us not only to capture the effect of the communication complexity of a variety of financial items, but, more important, to directly examine the overall complexity of firms' written public communication, over and above its specific content. We focus on the readability (or communication complexity) of the 10-K filing for several reasons. First, 10-K filings are required for all publicly traded companies and are frequently cited as an input in the decisions of investors and financial analysts (Previts et al. 1994; Rogers and Grant 1997). Second, both the SEC and popular press have routinely criticized firms for the complexity of their language in these filings and the effect of this complexity on investors (Schroeder 2003). Third, the 10-K filing contains a significant amount of written communication or narrative to use in interpreting its readability. These reasons make the 10-K filing an interesting setting to examine the effects of the communication complexity on the activities of financial analysts.

We test the relation between the readability of 10-K filings and several measures of analyst behavior, including analyst following, analyst forecast revision response time, the information

[2] More specifically, we focus on the so-called "syntactic complexity." Stimulated by Chomsky's (1965) seminal generative grammar, there has been a great deal of theoretical and empirical work on what constitutes syntactic complexity and how it functions as a determinant of both sentence comprehension (Gibson 1998) and production (Thompson and Faroqi-Shah 2002). According to Stone et al. (2005, pages 341-343), sentence complexity is affected by the features of open-class words (nouns and verbs) and their relationships and the number and type of syntactic operations, which is usually reflected in sentence length.

[3]The Fog Index has been used in a variety of applications including medical error, consumer drug use, consumer warranties, mutual fund prospectuses, jury instructions, and academic research prestige. For specific examples see Koo, Krass, and Aslani (2003), Gazmararian et al. (1999), Charrow and Charrow (1979), Armstrong (1980), Johnson (2004), Ott and Hardie (1997), Lee et al. (2006), and Shuptrine and Moore (1980).

content of analysts' reports, and analyst earnings forecast accuracy and dispersion. We begin our analysis by examining the relation between analyst following and 10-K filing readability. We argue that, on the one hand, greater complexity of firm financial disclosures increases the cost of processing the information in these disclosures and therefore will increase the demand for analyst services. If the cost of obtaining analyst reports is less than the information processing costs of firm disclosures, analyst following should be greater for firms with less readable 10-K filings. On the other hand, complex disclosures also may increase the costs of analyst coverage. Analysts may bear greater information processing costs, higher private search costs, and also may produce less accurate forecasts. If these costs are significant, then analysts following should be lower for firms with less readable disclosures.

Consistent with the prediction of an increasing demand for analyst services for firms with more complex 10-K filings, we document a positive and significant association between a firm's 10-K Fog Index and the number of analysts who cover the firm. This result continues to hold even after controlling for other factors related to analyst coverage and business complexity that have been documented in prior studies. We interpret this result as evidence of greater collective effort by analysts for firms with less readable disclosures. We also find evidence of a significant non-linear relationship between 10-K filing readability and analyst following, suggesting that for some levels of communication complexity analysts do provide greater amounts of information to investors. However, at some point the marginal benefits of analyst coverage are outweighed by the marginal costs and analysts' response declines.

Next, we examine whether individual analyst effort depends on the readability of the 10-K filing. We measure individual analyst effort as the length of time required for analysts to issue their first forecast revision following the 10-K filing. We argue that if analysts bear costs in following firms with less readable disclosures, it should take them more time on average to issue reports following less readable disclosures. Using a variety of empirical approaches, we find that analysts who cover firms with less readable 10-K filings take longer time to issue their first report following the 10-K filing. This evidence is consistent with the argument that analysts exert more effort to follow firms with less readable disclosures.

We then turn to an examination of the relation between communication complexity and the information content of analyst reports. Similar to Frankel, Kothari, and Weber (2006), we measure information content as the proportion of a firm's stock returns related to analyst forecast revisions to the total firm's stock return during the time period between the 10-K filing and the subsequent fiscal year-end. We predict that if investors place greater reliance on analyst reports issued for firms with less readable 10-K filings (potentially due to the higher costs associated with processing and interpreting a complex report), the information content of analyst reports, or the proportion of firm information associated with them, should be higher for firms with less readable disclosures. Consistent with this prediction we find evidence of a positive association between the readability of a firm's 10-K report and the proportion of firm information in analysts' reports. This evidence suggests that investors find analysts' reports more informative for firms with more complex corporate disclosures.

Finally, we examine how 10-K filing readability relates to two commonly studied properties of analyst earnings forecasts: accuracy and dispersion. Prior literature finds that the 10-K filing represents a major source of information used by analysts (Previts et al. 1994; Rogers and Grant 1997). However, studies also find that the complexity of items in the 10-K filing affects analysts' use of this information. For example, Plumlee (2003) finds that analyst forecasts are less accurate if they are associated with complex changes in the tax law. Bradshaw et al. (2008) find that differences in accounting choice negatively affect forecast accuracy and increase dispersion. Our tests relate to these findings, since we examine the effect of complexity through the ability of analysts to incorporate less readable financial information into their forecasts. If communication complexity increases the costs of processing and interpreting firm disclosures, it may limit the ability of analysts to correctly incorporate all the pertinent information and also may result in disagreement or ambiguity. We predict and find that analyst forecasts for firms with higher 10-K Fog scores are less accurate and have greater forecast dispersion. These results suggest that analyst forecasts are affected by 10-K readability and provide indirect evidence to support the notion that the information contained in the 10-K filing is used by analysts.

This paper contributes to the literature in the following ways. First, our overall findings that more complex communication is related to greater levels of analyst coverage, effort, and information content, but lower forecast accuracy and higher dispersion complements and contributes to the literature on how analysts respond to firms' disclosure and complexity. While prior studies generally focus on the effect of the complexity of specific attributes of firms' disclosure on analysts behavior, we provide evidence that suggests that analyst behavior is associated with the overall communication complexity.[4] Second, our findings contribute to the legal and regulatory debate about the intended audience of public corporate filings. Since its organization following the passage of the Securities Acts, the SEC has made consistent efforts to encourage firms to make their regulatory filings accessible to the average or "lay" investor (Firtel 1999).[5] The most recent of

---

[4] We note that covering firms with complex communication is likely to result in a trade-off. Investors are more likely to find analyst coverage useful for firms with more complex communication; however, earnings forecasts for these firms will potentially be less accurate. Our findings suggest that the benefits of coverage outweigh the costs, since we document that both greater coverage and less accurate forecasts are associated with more complex communication (see also Li, Rau, and Xu 2009).

[5] Some academic scholars and practitioners have argued that the primary users of this information should be market professionals such as analysts and not the 'lay' investor (Kripke 1970; Schipper 1991). This is because nonprofessionals may not possess the skill or expertise to read and understand the complex financial information contained in disclosure documents and, therefore, any effort to gear disclosure toward the layperson is a waste of time and money (Kripke 1970). Further, critics argue that the SEC's efforts to appeal to the average investor are not only inefficient but are hazardous to the disclosure regime—to appeal to the layperson, the SEC is forced to compel simplified, concise disclosure which often leaves out many issues that are potentially valuable to the professional and leads to potential legal liability (Kripke 1973).

these efforts is the plain English disclosure rules adopted by the SEC on January 22, 1998.[6] Our evidence that more complex corporate disclosures are associated with greater analyst following and more informative analyst reports suggests that analysts serve as an information intermediary for average investors and should alleviate some of the SEC's concerns on the accessibility of these reports.

Section 2 discusses communication complexity and the hypotheses development. Section 3 describes the data and sample, and section 4 presents our empirical evidence. Section 5 concludes.

2. **Hypotheses Development**

Complex communication is more difficult to interpret and process by investors because it requires that investors devote more time and effort to identify and extract relevant information (Bloomfield 2002).[7] In this study we focus on the complexity of firm disclosures, as measured by their readability rather than their content. We interpret disclosure readability to be a measure of the costs incurred by users to process and interpret a firm's written communication after controlling for the operational complexity of the business. Firms with similar operations provide disclosures with varying levels of readability. For example, Berkshire Hathaway and AIG both provide information about their reinsurance businesses in the management's discussion and analysis (MD&A) section of their 10-K filings for fiscal year 2003. While the subject matter and underlying business complexity of this issue are very similar across the two firms, Berkshire Hathaway's

[6] In a recent speech to the Second Annual Corporate Governance Summit, former SEC chairman Christopher Cox specifically identifies readability measures as a tool to examine communication complexity. He states, "Just as the Black-Scholes model is 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" (Cox 2007). Similarly, Core (2001) proposes the use of computational linguistics methods in accounting to measure corporate disclosure quality.

[7] In an analytical setting Indjejikian (1991) finds that under certain conditions investors will decrease their information acquisition activities as the noise in a firm's disclose signal increases. We interpret this noise to be a measure of the complexity of management's communication.

explanation is found to be more readable.[8] This is not surprising since Berkshire Hathaway's CEO, Warren Buffett, is a strong proponent of less complex communication.[9]

Using readability as a measure for communication complexity, we examine the effect of disclosure complexity on financial analysts and their earnings forecasts. We begin by examining the effect of communication complexity on analyst following. We assume that users of financial information have different abilities to process complex communication (Ball 1992; Indjejikian 1991). These differences provide opportunities for information intermediaries, such as financial analysts, to profit from their private analysis of firms by selling their opinions to users with greater information processing costs (Schipper 1991). These profitable opportunities are arguably greater for firms with higher communication complexity because of the greater cost to users of processing the firms' information. As such, if analysts respond to this increased demand for their services, we expect analyst following to be greater for firms with more complex communication.

We note, however, that analysts also face a variety of costs in covering firms with complex communication. First, analysts bear the direct costs of processing the information provided by management. Second, complex communication may require that analysts incur greater private search costs as they obtain additional information to evaluate and interpret management's communication. Third, greater complexity may lead to inaccurate forecasts and recommendations that may affect analysts' careers (e.g., Plumlee 2003; Hong and Kubik 2003).[10] Finally, Li (2008)

[8] Specifically, AIG's Fog Score is 18.51 and Berkshire Hathaway's is 17.23. The difference between these scores is 1.28 and is interpreted as the number of additional years of formal education required to easily understand the text on a first reading. Further details regarding the Fog Index are provided in Section 3.

[9] It is important to note that less complex communication may come at a cost. Legal scholars point out that "it is much harder to simplify than to complicate" (Kimble 1994) and it requires significant skill, work, and time to compose documents in plain language. Additionally, some critics argue that plain English communication increases the risk of litigation (Kripke 1973).

[10] Li et al. (2009) find that once analysts become all-stars they are more willing to cover firms with greater potential earnings management since investors may find their coverage more valuable for these firms. This suggests that coverage decisions can be related to favorable career outcomes; however, it is not clear exactly how analysts assess the trade-offs between the demand for investment information and the cost of potentially less accurate forecasts.

finds that firms attempt to obfuscate bad news by increasing the complexity of their communication, and Lang, Lins, and Miller (2004) find that analysts are less likely to cover firms with incentives to withhold or manipulate information. These potential manipulations along with the costs to analysts of covering firms with complex communication may discourage analyst following. Since there are both potential positive and negative consequences of communication complexity on analyst effort, we test the null hypothesis that communication complexity has no association with analyst following.

Given the difficulty of following firms with more complex communication and meeting the demands of investors, analysts who choose to follow these firms may be required to exert greater effort to do so. One way of measuring the effort that analysts exert in following firms with less readable disclosures is to examine the average time from the firm's 10-K filing to the analyst's first report subsequent to the filing. If disclosure complexity results in higher processing costs to analysts, then firms with more complex communication will have greater average response time from the analysts that cover them as compared with firms with less complex communication. Similarly, analysts who cover a portfolio of firms that have more complex communication should take longer, on average, to issue their reports than other analysts. We test the hypothesis that analysts exert greater individual effort to cover firms with more complex disclosures.

Communication complexity also may affect the properties of analyst earnings forecasts. Specifically, we examine its effect on the information content, accuracy, and dispersion of analyst forecasts. The informativeness of analyst reports depends on how useful investors find the information provided. Analyst reports for firms with more complex communication likely provide information that is more useful to investors due to greater costs in processing public information. Analyst reports for firms with more complex communication also may be more informative if analysts choose to acquire and incorporate more private information in them due to the complexity of firm disclosures. Accordingly, we predict that analyst reports for firms with more complex communication will have greater information content, on average, than firms with less complex communication.

Finally, since analysts who follow firms with more complex communication bear the costs of processing and interpreting such disclosures, communication complexity also may affect the accuracy and dispersion of analyst earnings forecasts. If complex communication makes it more difficult to forecast earnings, then analyst forecast accuracy will be lower for firms with more complex communication. However, this difference in accuracy may be fully or partially offset by greater analyst effort in response to greater complexity. Additionally, complex communication is likely to lead to a more diverse set of interpretations about the firm disclosures, resulting in higher analyst forecast dispersion. Accordingly, we predict that analyst forecasts will be less accurate and have greater dispersion for firms with more complex communication.

3. Sample and Variable Definitions

3.1. Sample Selection

Our initial sample is based on the intersection of firm/years available on the Compustat Fundamental Annual table and the SEC's EDGAR filings database for fiscal years 1995-2006. These databases are joined based on Compustat GVKEY and the SEC's Central Index Key (CIK). Firms without matches are dropped from the sample. For each firm-year observation we download the corresponding 10-K filing. Filings with less than 3,000 words or 100 lines are dropped. This procedure results in 57,417 observations. We obtain stock return data from CRSP, analyst data from I/B/E/S, institutional holdings data from Thomson Reuters (CDA/Spectrum), and information on management earnings guidance from the First Call Company Issued Guidelines database. This procedure yields a sample of 37,734 observations.

3.2 Disclosure Readability

Similar to Li (2008), we measure the readability of 10-K filings using the Fog Index. This index, developed by the computational linguistics literature, captures the complexity of a document as a function of the number of syllables per word and the number of words per sentence. Specifically, we calculate the index as follows:

$$FOG = (\text{average words per sentence} + \text{percent of complex words}) \times 0.4, \quad (1)$$

where a complex word is defined as one with three or more syllables. The index is interpreted as the number of years of formal education required for a person of average intelligence to read the document once and understand it.[11] It is important to note that the Fog Index is a measuring tool, not a rule or formula for good writing, as stated by Gunning (1969), "Nonsense written simply is still nonsense." It predicts the readability of a document, but does not provide information about whether the writing is interesting or informative. Despite these limitations, it is objective and simple to calculate. It allows us to study the disclosure characteristics of a large and diverse group of firms and doesn't depend on analyst surveys or opinions. It also provides us with a comprehensive measure of the overall communication complexity of 10-K filings as opposed to the complexity of individual financial items.[12]

Table 1 provides descriptive statistics on the Fog Index for firms in the intersection of Compustat and EDGAR. As can be seen in Panel A, the overall mean and median of the Fog Index are 19.48 and 19.33, respectively, suggesting that a person of average intelligence would require over 19 years of formal education (on average) to read a 10-K filing once and understand it. This is consistent

[11] We remove all tables, tabulated text, and financial statements from the 10-K before computing the Fog Index. The constant of .4 found in equation (1) was chosen by Robert Gunning based on the scores of a set of literary benchmarks in order that this specific interpretation could be made.

[12] Prior literature examines the readability of the firms' financial reports using the Fog Index and other similarly constructed measures to assess their complexity. These studies examine the readability of the overall annual report (Jones and Shoemaker 1994), management's discussion and analysis (Schroeder and Gibson 1990), and the notes to the financial statements (Smith and Smith 1971; Healy 1977). Recent studies use the Fog Index to examine the readability of annual reports in connection with earnings persistence (Li 2008), timely price adjustment (Callen, Khan, and Lu 2009), and investment efficiency (Biddle, Hilary, and Verdi 2009).

with concerns that financial reports are written in complex language. As a comparison to help validate our measure, it is interesting to note that the average Fog Index for FASB Statement Nos. 1-122 is 22, the CPA Exams is 16, the CMA Exams is 17, *The Wall Street Journal* is 12, and *Readers Digest* is 8 (Phillips, Daily, and Luehlfing 2007; Cox 2007).

While the variation in the mean Fog Index over the sample period is modest, there is large variation within each year. For example, over the total sample years, the inter-quartile range for the Fog Index is from 1.61 to 1.88. There is also significant variation in the Fog Index within industries, despite similarities in the underlying business complexity within each industry. Panel B of Table 1 provides examples of specific industries that have high and low levels of the Fog index.[13] The healthcare, insurance, utilities, telecommunications, and trading (e.g., security brokers, investment offices, etc.) industries make up the group with the highest fog. The fact that these industries have a high level of disclosure complexity is not surprising since these industries are characterized by complex contracts and business models that are difficult to communicate; however, there is also significant variation within each industry. The low Fog group is composed of the precious metals, shipping container, agriculture, food, and defense industries. While the precious metals and shipping container industries have small within industry variation, most industries have an inter-quartile range of Fog above 1.60. The relative rankings of these industries also helps to validate the ability of the FOG Index to measure the readability of annual reports.

3.3 Variable Definitions

Our tests focus on examining the relation between the Fog index and analyst following, their forecast revision response time, the information content of their reports, and the accuracy and dispersion of their forecasts. A description of the construction of these variables as well as our control variables follows.

---

[13] Industries are classified using the Fama and French 48 industry classification.

Analyst following

Similar to prior work (see, for example, O'Brien and Bhushan 1990; Brennan and Subrahmanyam 1995), we define analyst following as the number of analysts (#ANALYSTS) that comprise the first I/B/E/S consensus annual earnings forecast after the filing date of the 10-K report.[14] We follow Bhushan (1989) and interpret this measure as a proxy for the collective effort of the financial analyst community in the analysis of an individual firm.[15] Since some firms are not covered by I/B/E/S, we conduct tests for both the I/B/E/S sample and the full sample of firms, where missing coverage is coded as zero analyst coverage (Barth et al. 2001). Since these results are similar our main results are based on the I/B/E/S sample.

Analyst forecast revision response time

We define the analyst forecast revision response time as the time from the 10-K filing to the first annual or quarterly earnings forecast issued by each *individual* analyst following the firm. To ensure that we include only analysts who actively follow the firm, we require that each analyst issue a forecast in the 90 days prior to the 10-K filing and then another report within 90 days of the 10-K filing.[16] Since earnings announcements may prompt analysts to issue reports, we exclude reports made after any earnings announcement that occurs after the 10-K filing, but before the end of the 90-day window. We then define analyst report duration as the length of time in days between the 10-K filing and the first report following the filing. In addition, we average individual analyst

---

[14] We obtain similar results using the second consensus forecast. In addition, in untabulated results we estimate our regressions using the natural log of the number of analysts following a firm. Our results are largely unchanged in terms of direction and significance. Rock, Sedo, and Willenborg (2001) note that regression using a dependent variable that is count data, e.g., number of analysts, can be sensitive to the estimation procedure used. As a check on our results we perform our tests using negative binomial and Poisson regression techniques and find that our results are largely unchanged in terms of sign and significance.

[15] Similar to Bhushan (1989) we acknowledge that our proxy is not a perfect measure since it assumes homogeneity among analyst effort levels. For example, there are differences in individual analysts' effort levels based on differences in compensation, brokerage houses, etc.

[16] This process eliminates 12,767 observations related to analysts who stopped coverage. In the duration tests, we will account for these potential problems using a duration model that accounts for censored data.

report duration at both the individual firm and analyst level for each year. We interpret this duration variable as a measure of the required amount of effort for an individual analyst to read, understand, and process the information contained in the 10-K filing and to issue an updated earnings forecast. As mentioned above, we expect analyst response time to be longer for more complicated 10-K filings.

Information content of analyst reports

Similar to Frankel et al. (2006), we measure the information content of analyst reports as the proportion of a firm's stock returns related to analyst forecast revisions to the total stock return during the time period between the 10-K filing and the subsequent fiscal year-end. This measure is constructed as the sum of the one day, absolute size-adjusted returns on the analyst forecast revision day divided by the sum of the one day, absolute size-adjusted returns over the entire window.[17] In this calculation, we exclude reports that coincide with earnings announcements. We treat multiple reports issued on the same day as a single report. Also an observation must have a minimum of 90 trading days with available data to be considered in our tests. Similar to Frankel, et al. (2006), we interpret this measure as the percentage of total firm information provided to investors that is related to analyst reports.

Forecast accuracy and dispersion

We define analyst forecast accuracy as the absolute analyst earnings forecast error (ACCURACY) calculated as the absolute difference between I/B/E/S reported earnings and the first analyst consensus annual earnings forecast issued after the 10-K filing for the fiscal period following the

[17]Frankel et al. (2006) divide this measure by the number of forecast revision dates to obtain a measure of the average informativeness of an analyst report date. We omit this final step since our variable of interest is the overall firm information that comes from analysts, not the average information content. Our inferences are robust to the inclusion of the natural logarithm of the number of analyst forecast revisions in the regression results.

10-K filing, scaled by share price 90 days before the consensus forecast date. Analyst forecast dispersion (DISPERSION) is computed as the standard deviation of the individual analyst forecasts scaled by share price 90 days before the consensus forecast date.

Control Variables

Our analysis controls for a variety of variables that have been shown by prior literature to be associated with firms' information environment and business complexity and therefore relate to analyst following and the properties of their forecasts. Prior work finds that firm size is the most important determinant of analyst following (Bhushan 1989; O'Brien and Bhushan 1990; Brennan and Hughes 1991; Lang and Lundholm 1996; Barth et al. 2001). These studies find that larger firms have greater analyst following and suggest that large firms have better information environments, potentially more complex operations and greater demand for investment advice. We use the natural logarithm of market value as of the year ending prior to the 10-K filing (LOGSIZE) as a proxy for size.

Following Barth et al. (2001) we include controls for growth. High-growth firms may attract greater analyst following due to investor interest and the potential for future investment banking deals. Further, analysts may find it more difficult to accurately forecast earnings for firms with high growth, which may lead to greater disagreement among analysts and lower accuracy of their forecasts. We define the variable GROWTH as the compounded average growth rate in sales over the prior three to five fiscal years. Our inferences are robust to the inclusion of the natural logarithm of the book-to-market ratio as a proxy for growth.

Similar to Bradshaw et al. (2008) we include the natural logarithm of the number of business segments reported in the Compustat Segment File as a control for the underlying complexity of the firm. We also include a control for the level of institutional holdings following the evidence in Bhushan (1989), Brennan and Subrahmanyam (1995), and Frankel et al. (2006). These studies find that institutional ownership is positively associated with analyst following and with the information content of their reports.[18] It also can be associated with higher analyst forecast accuracy and lower dispersion since firms with high levels of institutional holding may have better information environments. We define the variable PINST as the percentage of a firm's shares that are held by institutions from the 13f disclosures for the most recent quarter prior to the 10-K filing. Motivated by the evidence in Lang and Lundholm (1996), who document that analyst following increases with the quality of disclosures, we include the number of management earnings forecasts made during the prior year (MFCOUNT) as a proxy for firm discretionary disclosure (see also Nagar, Nanda, and Wysocki 2003). Cotter, Tuna, and Wysocki (2006) find that managers guide analysts' expectations, which also may affect the properties of analyst forecasts.

We include controls for analysts' coverage incentives, as well as additional variables that related to firms' information environment and business complexity, since these may affect the properties of analyst earnings forecasts. Barth et al. (2001) examine the association between analysts' incentives to follow firms and the extent of their intangible assets. Since many intangible assets are generally not recognized and estimates of their fair value are not disclosed, analysts may have increased incentives to follow firms with greater intangibles due to increased demand from investors. They find that analyst following is greater for firms with larger research and development and advertising expenses. They interpret these results as evidence that analysts respond to the demands of investors for more information because of the difficulty in evaluating firms' intangible assets. Other studies examine the effect of intangibles on the properties of analyst forecasts. Barron,

[18] O'Brien and Bhushan (1990) suggest that analyst following and institutional ownership are potentially endogenous. Frankel et al. (2005) assume that the variables are exogenous. Similar to Barth et al. (2001), we examine our results with and without the inclusion of institutional ownership and also with the inclusion of lagged institutional ownership. Our results are largely unchanged in both magnitude and significance under these specifications.

Byard, Kim, and Riedl (2002) find that analyst uncertainty increases with the level of a firm's intangible assets. Gu and Wang (2005) find that analyst forecast errors are increasing in firm intangible intensity. Intangibles also may be important for our tests since they also may be associated with a firm's 10-K readability. For example, it may be more difficult to explain the operations of firms with high levels of research and development costs in a less complex manner. Similar to Barth et al. (2001), we define R&D as the ratio of research and development expense to operating expense and ADV as the ratio of advertising expense to operating expense.[19]

Finally, we include the standard deviation of firm monthly stock returns from the prior year (STD_RET) as a measure of information uncertainty.[20] Bhushan (1989) suggests that private information is more valuable for firms with higher return volatility and thus positively related to the demand for analyst services. However, it also may be the case that analysts bear increased costs for following firms with higher return volatility. Additionally, most of our tests include industry fixed affects (based on the Fama and French classification) and year fixed effects to account for variation in analyst following across specific industries and over time. This approach is used to help control for variation in business complexity that is driven by industry or time.

4. Empirical Results

4.1 Summary Statistics

Table 2 presents the summary statistics for the sample of 37,734 remaining after imposing the availability of data on CRSP and I/B/E/S. Similar to Table 1 and consistent with the concerns raised by the SEC about the disclosure complexity, the mean (median) FOG score is 19.47 (19.32) which is classified as "unreadable." The standard deviation and inter-quartile range are 1.42 and

[19] Our results are robust to the inclusion of the amount of recognized intangibles and depreciation expense which are also examined by Barth et al. (2001).

[20] The inclusion of an alternative measure, earnings volatility, does not significantly affect our results; however, it does impose sample restrictions due to the need for a sufficient time-series of earnings.

1.74, respectively, with an inter-quartile range from 18.52 to 20.26. The mean (median) number of analysts per firm-year observation is 6.77 (5). The mean (median) average analyst forecast revision response time is 17.51 (15.71) days at the firm-year level and 18.45(17.25) days at the analyst-year level. This suggests that, on average, analysts may not respond immediately to 10-K filings due to the significant amount of information that must be processed and interpreted. The mean (median) analyst report information content (AI) is .1477 (.1186). This indicates that, on average, almost 15% of the information reflected in stock returns during the period between the 10-K filing and the end of the fiscal period is derived from analyst reports. Absolute forecast error (ACCURACY) has a mean (median) .0349 (.0103). The mean (median) forecast dispersion is .0075 (.0027).

Table 2 also provides statistics on our control variables. The mean (median) size of our sample firms is $2.5 billion ($447 million), and mean (median) compound averages growth rate of sales (GROWTH) is .18 (.11). The mean (median) number of business segment is 1.85 (1), and mean (median) percent of institutional ownership is .48 (.49). The mean (median) number of management earnings forecasts is 1.75 (0). The mean (median) ratio of research and development expense (R&D) and advertising (ADV) to operating expense are .07 (.00) and .01 (.00), respectively.[21] The mean median standard deviation of returns (STD_RET) is .14 (.12).

Table 3 presents the univariate correlations for the variables used in this study. FOG is significantly correlated with many of the variables; however, the extent of these correlations is relatively small for most variables. This is consistent with the results in Li (2008) which suggest that it is difficult to explain a significant proportion of the variation in FOG using firm characteristics. This is important because in this study we assume that FOG measures the readability of 10-K disclosures rather than their content. Communication complexity (or readability) may be

[21] Note that the medians for the number of management earnings forecasts, research and development expense, and advertising expense are zero. This is primarily because we follow prior literature and code missing items to be zero.

decomposed into innate and discretionary components. While we include a variety of variables to control for innate complexity, analysts' information costs are based on total complexity so the relative magnitude of these components is less relevant. Our inferences are limited by the extent to which FOG measures this construct and is not confounded with other (uncontrolled) firm characteristics. It is also important to note that for a specific variable to influence the FOG measure it must either increase the length of the average sentence in the 10-K document or increase the percentage of complex words. Simply increasing the length of the 10-K will not directly affect the measure. However, we examine the effect of disclosure length on our main results in sensitivity tests since length also may be a form of complexity.

Similar to prior work, the correlation coefficients reported in Table 3 indicate that analyst following is positively correlated with size (.77), institutional ownership (.52), and management guidance (.26). It is also highly correlated with analyst report information content (.82). Analyst following is negatively related to absolute forecast error and dispersion. Also absolute forecast error and dispersion are positively correlated (.53) consistent with uncertainty being related to forecast inaccuracy (Bradshaw et al. 2008). In general, the correlations among the independent variables used in the results that follow are not indicative of a strong multi-collinearity problem.

4.2 Analyst Following and the Readability of 10-K Filings

Our first prediction is that analyst following is affected by the level of plain English or readability of firms' 10-K filings as measured by the Fog index. The univariate correlations found in Table 3 suggest a positive and significant relationship between analyst following and 10-K Fog. To control for other factors that may affect analyst following we estimate the following regression:

$$\#ANALYSTS = \beta_0 + \beta_1 FOG + \beta_2 LOGSIZE + \beta_3 GROWTH + \beta_4 LSEGMENTS + \beta_5 PINST + \beta_6 MFCOUNT + \beta_7 ADV + \beta_8 R\&D + \beta_9 STD\_RET + \varepsilon. \quad (2)$$

The estimation is performed using ordinary least squares regression with industry and time fixed effects. T-statistics, presented in brackets, are based on standards errors that are robust to heteroskedasticity and are clustered at the firm level.

Column 1 of Table 4 (Model 1) reports the results of the linear model regression. The coefficient on FOG is positive and statistically significant, suggesting that analyst following is greater for firms with less readable disclosures. This is consistent with the notion that analysts respond to investors' demand for investment information for firms whose disclosures are more costly to process. The coefficients on the control variables are consistent with the findings of prior literature. Larger firms are associated with greater analyst following as well as firms with higher growth, institutional ownership, and greater discretionary disclosure. Consistent with Lang and Lundholm (1996), we find that disclosure practice, as captured by MFCOUNT, is related to analyst following. We find that analyst following is negatively associated with the number of business segments.[22] Similar to the findings in Barth et al. (2001), we document that analyst following is greater for firms with higher amounts of advertising and research and development expenses. Consistent with prior work, we also find that analyst following is positively associated with firm stock return volatility, suggesting that analysts provide greater support to investors when private information is valuable (Bhushan 1989).

In addition to examining the linear association between FOG and analyst following, we also examine the association of analyst following and FOG using non-linear and semi-parametric specifications (second and third columns of Table 4, respectively). In the non-linear specification (model 2) we add a square term (FOGxFOG) to equation (2) to capture the possibility that the association between readability and analyst following differs based on the level of FOG. In model

[22] This effect becomes negative after controlling for firm size, but is positive if size is omitted. The negative coefficient is consistent with the findings of Bhushan (1989).

3, we examine a semi-parametric specification by including indicator variables for the various quartiles of the FOG Index. The coefficient on FOGxFOG (model 2) is negative and significant. Also the coefficients on the second, third, and fourth FOG quartiles (model 3) are positive and significant but increasing at a decreasing rate across the quartiles. These results suggest that analysts make a trade-off between the benefits of covering firms with less readable disclosures and the costs of following them. While these costs include the processing costs of covering firms with less readable disclosures, analysts also may find it more difficult to accurately forecast the earnings of firms with more complex disclosures. These costs may be important since prior literature has found that forecast accuracy may have important career effects for analysts (Hong and Kubik 2003). The effects of FOG on the properties of analyst earnings forecasts will be examined later.[23]

Panel B of Table 4 compares the effects of an inter-quartile change in the independent variables on analyst following based on the marginal effect estimates in Panel A. Consistent with prior literature, size (LOGSIZE) has the largest association with analyst following. An inter-quartile change in the natural logarithm of size is associated with an increase of 6.8 analysts, while that of all other variables is less than 1. The inter-quartile effect of FOG is .15 in model 1 and .20 in models 2 and 3. While the incremental effect of the complexity of 10-K filings on analyst following is small, it is comparable with that of most of the remaining variables (except firm size). Based on these comparisons, we conclude that disclosure readability, measured by FOG, is an important determinant of analyst following.

Similar to Lang and Lundholm (1996), we also examine the association between *changes* in FOG and *lead changes* in analyst following. If analysts do, in fact, respond to changes in FOG by

---

[23] In a related study, You and Zhang (2009) find that the market under-reacts to firms with more complex annual filings using the returns following 10-K filings. This suggested under-reaction may provide additional motivation for analyst coverage since prior work has found that analysts help improve market efficiency (Gu and Chen 2004, Barron et al. 2002, Brown and Sivakumar 2003).

increasing their coverage, we expect to observe a positive relation between current changes in FOG and future changes in analysts following. We investigate this relationship by estimating multivariate models using the changes in the independent variables from Table 4. We define the current change in the Fog Index and other independent variables as the difference between their current values and those of the prior fiscal year. We define the lead change in analyst following as the difference between the number of analysts following the firm after the next 10-K filing and the number of analysts following the firm after the current report. Table 5 presents our results from these tests. As shown in Panel A, we find that contemporaneous changes in FOG are positively related to lead changes in analyst following. We also find that changes in firm size and institutional ownership are positively related to lead changes in analyst following and that changes in management earning guidance, number of business segments, and return volatility are negatively related.[24] Panel B reports the estimated effect on lead change in analyst following of an inter-quartile change in each independent variable. We find that FOG has the third-highest effect of the variables employed in our analysis. The results from this test are important because they help to alleviate concerns about endogenieity or prior period information shocks relating to our levels results in Table 4.

4.3 Analyst Report Duration

One potential measure of the costs or effort that analysts bear in following firms with less readable disclosures is the amount of time it takes them to issue reports following the 10-K filing. We label this length of time "analyst report duration." In this section, we examine the analyst report duration at both the analyst and firm level. We expect a positive relation between the communication complexity of the 10-K filing and analyst report duration. As noted previously, to ensure that we capture analyst reports issued in response to the filing of the 10K (and not in

[24] While less intuitive, one potential concern is that analyst following might lead FOG. For example, management might provide more complex information in response to a greater analyst following. In untabulated tests we find no relation between changes in FOG and lag analyst following.

response to other corporate events) we include in our analysis only firm-analyst observations where the analyst has issued at least one report during the 90 days prior to the 10-K filing and issues a report in the 90 days subsequent to the 10-K filing.[25] Panel A of Table 6 presents univariate results based on individual analyst reports and the Fog index of the firms they cover. We classify each firm-specific Fog as high (low) if it is greater (less) than the median Fog value for all firms in the sample. We find that analysts covering firms with a high Fog require 1.11 days longer on average to issue their reports. Panels B and C of Table 6 present the average analyst report duration per firm-year and analyst-year, respectively. The value of FOG on the firm level is the individual firm 10-K Fog for each specific year. The FOG score on the analyst level is the average 10-K Fog score of the firms that the analyst covered in each year.[26] The univariate results suggest that the average (median) analyst report duration for firms with FOG scores higher than the median is .84 (1.18) days (Panel B). On the analyst level, we find that analysts who cover a portfolio of firms with an average FOG score higher than the median analyst portfolio average take 2.47 (2.36) days longer on average (median) to issue their first reports. These differences are statistically significant at a 1% level.

We next perform a more rigorous analysis of the analyst report duration by employing a hazard model of the time until an analyst issues her first report following the 10-K filing. Similar to O'Brien, McNichols and Lin (2005), we estimate a hazard model of the following form:

$$\ln h_{ij}(t) = \lambda(t) + \beta \, FOG_{ij} \tag{3}$$

where:

$h_{ij}(t)$ = the instantaneous risk of hazard, or in our case the likelihood that analyst i will issue her first report following the 10-K for firm j at time t given that she has not yet issued her first report;

---

[25] As noted previously, we exclude analyst observations made after earnings announcements that follow the 10-K filing and reports made after 90 days after the 10-K. Our results are similar if we use reports within a year after the 10-K filing; however, these observations are likely unrelated to the FOG score of the 10-K filing.

[26] To be included in our analysis each analyst must cover a minimum of four firms. Similar results are obtained by requiring only a single firm.

$\lambda(t)$ = the baseline hazard;

β = the coefficient corresponding to the effect of the FOG on analyst revision duration.

We allow the model to take into account observations where the analyst stopped coverage after the 10-K filing before issuing her first report.[27]

We estimate this model as a proportional hazard model or Cox model which is commonly used in social science research (Cox 1972, Demaris 2004). We estimate this model using by partial likelihood and handle ties in the ordering of events following Efron (1977). The advantage of this model is that it allows the researcher to examine the relative effects of the variables of interest without defining the specific functional form of the baseline hazard. This method allows us to obtain unbiased and asymptotically normal estimates of the coefficients of interest, but at the cost of full efficiency since we do not know the true form of the baseline hazard. Additionally, the results of this model depend on our assumptions regarding the proportionality of the effects, sample selection, and censoring.

We employ this model to examine the effect of disclosure readability on the time it takes an analyst to issue her first forecast following the 10-K filing both at the firm and at the analyst level. Accordingly, we estimate two separate Cox models by conditioning at the firm and analyst level separately (similar to O'Brien et al. 2005). This conditioning is executed using the fixed-effects partial likelihood (FEPL). The purpose of this technique is to absorb the overall firm or analyst effects into the baseline hazard model, as opposed to observable controls, in an attempt to isolate the effect of FOG. The results of our estimation procedures for the "within firm" and "within analyst" models are presented in Panel D of Table 6. In both cases we find that an increase in

[27] We obtain this information from the I/B/E/S Stopped Estimate file. Stopped observations are considered censored observations in the hazard model estimation. There are 12,377 observations in this group representing about 5.7% of the observations in the hazard model estimation. These observations are not used in the univariate and regression analyses.

FOG is associated with a statistically significant reduction in the relative hazard rate. In the context of our tests, the hazard rate is the probability that the analyst will issue her first report at time t given that she has not done so before time t. We estimate the relative hazard rate, or how the rate of one firm differs relative to the firm's FOG scores. We find that, on average, the probability of an analyst making a report at any given time is between 1%-2.1% lower for each unit change of FOG.[28]

To further examine the analyst report duration, we estimate ordinary least squares regressions on the average analyst report durations at both the firm and analyst level. Panel E of Table 6 presents the results of our regression analysis. Similar to the univariate results, we find that firms with higher fog scores and analysts who cover stocks with higher fog scores are associated with longer analyst report duration. Specifically, controlling for a variety of other factors that might affect firms' information environment, the coefficient on FOG is statistically significant at 0.2931 for the firm-level analysis and 1.4260 for the analyst-level analysis. This suggests that firms' and analysts' portfolios with one unit higher of FOG have analyst response times that are .2931 and 1.4260 days longer as compared with similar firms. These results are consistent with those of our hazard model estimation. In addition, we find that firms with greater institutional ownership, discretionary disclosure, and research and development expense are associated with longer average analyst response times.[29] Overall, the evidence reported in this section is consistent with the notion that analysts exert more effort to cover firms with less readable disclosures.

---

[28] The results of our hazard model analysis are subject to assumptions. We assume that the effect of FOG on analyst report duration is proportional to an arbitrary baseline. We also make assumptions in our sample selection since we only examine firm-analyst observations by analysts that actively follow the firm. Additionally, we allow for the censoring of observations when analysts stop or halt coverage. The model assumes that this censoring is uninformative or unrelated to the complexity of disclosure. However, this may not be the case given that an analyst may drop coverage following unexpected bad news and this news may be related to a firm's 10-K Fog score (McNichols and O'Brien 1997, Li 2008).

[29] In untabulated results, we include a control for the number of reports issued by an analyst for the analyst level results. The coefficient on this variable is not statistically significant and does not quantitatively affect the results as presented.

4.4 Analyst Information Content

Our third hypothesis predicts that the informativeness of analyst reports is positively related to the readability of firms' 10-K reports. To test this hypothesis, we estimate the following ordinary least squares regression with industry and time fixed-effects equation:

$$AI = \beta_0 + \beta_1 FOG + \beta_2 LOGSIZE + \beta_3 GROWTH + \beta_4 LSEGMENTS + \beta_5 PINST + \beta_6 MFCOUNT + \beta_7 ADV + \beta_8 R\&D + \beta_9 STD\_RET + \varepsilon. \quad (3)$$

Similar to the regression estimated in Table 4, we also examine the association of analyst following and FOG using non-linear and semi-parametric specifications. T-statistics, presented in brackets, are based on standard errors that are robust to heteroskedasticity and clustered at the firm level.

The results of the estimation are found in Table 7. Consistent with our hypothesis, the coefficient on FOG is positive and significant, suggesting that the informativeness of analyst reports is increasing in the complexity of the 10-K. This evidence is consistent with the notion that investors find analyst reports for firms with less readable disclosures more useful because of greater processing costs. It also may be the case that analysts' private information searches are more valuable in such cases. In model 2, the square term (FOGxFOG) is negative but insignificant, whereas the coefficients on the individual FOG quartiles in model 3 are significant and monotonically increasing (at a declining rate). This is consistent with our prior results and supports the notion that analyst trade off costs and benefits in their coverage of firms with less readable disclosure. The effects of the control variables on the information content of analyst reports are similar in direction to their effects on analyst following. Specifically, our results suggest that the informativeness of analyst reports is increasing in firm size, growth, institutional ownership, discretionary disclosure, and firm intangibles.[30]

---

[30] As previously mentioned, our inferences are robust to the inclusion of the natural logarithm of the number of analyst forecast revisions in the regression results.

Table 7 also presents the estimated effects of an inter-quartile change in each variable on analyst report information content. An inter-quartile change in FOG increases the information content of analyst reports by about 0.4%, on average, across the three models. An inter-quartile change in the natural logarithm of market value increases the information content by 11% and, similar to our results for analyst following, is the most significant effect. The estimated inter-quartile effect of the percentage of institutional ownership (2%) is also large relative to the other variables. Consistent with our finding for analyst following, we find that the effect of an inter-quartile change in disclosure readability (FOG) is similar to many of the other variables in our analysis.

4.5 Properties of Analyst Earnings Forecasts

Given our findings regarding the association of disclosure readability with analyst following, analyst report duration, and analyst report information content, a reasonable next step is to examine its effect on the properties of analyst forecasts. In particular, we focus our findings on the accuracy and dispersion of analyst forecasts of annual earnings. As noted above, we hypothesize that disclosure readability will decrease analyst forecast accuracy by increasing the absolute forecast error. In addition, we hypothesize that analyst forecast dispersion will be higher for firms with less readable disclosures. For each of these hypotheses we estimate a model of the following form:

$$\text{Forecast Property} = \beta_0 + \beta_1 FOG + \beta_2 LOGSIZE + \beta_3 GROWTH + \beta_4 LSEGMENTS + \beta_5 PINST + \beta_6 MFCOUNT + \beta_7 ADV + \beta_8 R\&D + \beta_9 STD\_RET + \varepsilon, \quad (4)$$

where "Forecast Property" represents the variable of interest. In this case, we measure forecast accuracy with the absolute forecast error (ACCURACY) and forecast dispersion with the standard deviation of the individual analyst forecasts. Each model is estimated using ordinary least squares regression with industry and time fixed effects. Tables 8 and 9 present our empirical results. T-statistics, presented in brackets, are based on standards errors that are robust to heteroskedasticity and clustered at the firm level. Similar to our prior results, we also estimate models that incorporate a squared term (FOGxFOG) and the effects of FOG by quartile.

Table 8 presents our findings regarding forecast accuracy. Under the linear model, the coefficient on FOG is positive and significant indicating that less readable disclosures are associated with less accurate analyst forecasts. In model 2, the square-term (FOGxFOG) is negative, but is not significant. In model 3, only the effects of quartiles three and four are significantly different from zero. This suggests that the effect of FOG on the accuracy of analyst forecasts is based largely on the least readable disclosures. We also find that firm size and institutional ownership are associated with greater accuracy, while the number of segments, discretionary disclosure, advertising expense, and return volatility are associated with less-accurate forecasts. These results are largely consistent with the notion that analyst forecasts are more accurate for firms with better information environments and less accurate for firms with more complex operations.[31]

Table 9 presents our findings for analyst forecast dispersion. This evidence is similar to the results of forecast accuracy. Under the linear model, the coefficient on FOG is positive and significant, indicating that less readable disclosures are associated with more forecast dispersion. In model 2, the square-term (FOGxFOG) is positive, but is insignificant. In model 3, the effect of each FOG quartile on dispersion is positive and significant. In addition, the effect is increasing across the quartiles. This suggests that analyst forecast dispersion is largely affected by the least readable disclosures. These results are not surprising given our other findings regarding the trade-off of analyst coverage costs and benefits. We also find that analyst forecast dispersion is negatively related to firm size, institutional ownership, and discretionary disclosure and positively related to the number of segments, firm intangibles, and return volatility. Similar to our result for analyst forecast accuracy, these results also are consistent with the idea that firms with better information environments and less complex operations have lower forecast dispersion.

[31] One exception to this statement is that we find a positive association between MFCOUNT and ACCURACY. One potential explanation for this finding is that management issues a greater number of forecasts in years when the earnings are more complicated.

4.6 Sensitivity of the Results to 10-K Length

Li (2008) incorporates the length of the 10-K filing as an alternative measure of readability, in addition to the Fog Index, under the assumption that longer documents require higher information processing costs. While length is associated with communication complexity, it may also measure the amount of disclosure provided by the firm (Leuz and Schrand 2009). We examine the sensitivity of our results to our measure of communication complexity by partitioning our sample into 20 groups based on the number of words in each 10-K filing. For each group we estimate the main regression results from each of the prior tables. The coefficients from each group are then averaged and reported in Table 10. T-statistics are computed based on the standard errors of the coefficients across groups.[32] Our results for analyst following, changes in lead analyst following, analyst response time, and analyst information content are robust to considerations of disclosure length. However, our results for analyst earnings forecast accuracy and dispersion appear to be affected by length, although the estimated effects are small.

5. Summary and Conclusion

From the passage of the Securities Acts in 1933 and 1934 to the present, regulators, legal scholars, and various other interested parties have weighed in on the debate about the complexity of financial communication to external users. Currently, the SEC has gone so far as to require that the prospectuses of all registered public offerings meet the requirements of the Plain English Rules and has suggested that similar procedures should be applied to other mandatory filings. Former SEC chairman Christopher Cox has even suggested the use of readability models such as the Fog Index to measure the complexity of financial communication (Cox 2007). However, other users are concerned about the effect that these types of actions may have on the disclosure regime. These

[32] These results are conservative because partitioning the sample into 20 groups limits the potential variation of the Fog Index within each group. In untabulated results, we also estimate each regression for the entire sample by including an indicator variable set to one for 10-K with length greater than the median. Results are similar in both direction and significance. In addition, we also find similar results by including the natural logarithm of the number of words in the 10-K as a control.

parties argue that at a minimum these actions are a waste of time and effort, and may lead to a reduction in disclosure as firms are forced to simplify their public communication (e.g., Firtel 1999, Kripke 1970, 1973). In this study we contribute to this debate by examining the effect of communication complexity on one important financial information intermediary, sell-side financial analysts.

We present evidence consistent with the notion that since more complex firm communication is more costly to process and interpret, investors demand greater amounts of analyst services for firms with more complex communication. We find that analyst following is greater for firms with higher levels of communication complexity as measured by the Fog Index. We also find that analysts who cover firms with more complex communication take longer time on average to issue reports in response to 10-K filings. We interpret this evidence as analysts exerting greater effort to cover these firms. In addition, we find that analyst reports of firms with greater communication complexity are more informative to investors, but that the earnings forecasts of such firms are less accurate and have greater analyst dispersion or disagreement.

Our results suggest that analyst behavior is related to firms' communication complexity. While prior studies have found that analysts are affected by the complexity of individual financial items, we provide evidence that the overall linguistic complexity of firms' communication incrementally influences analyst behavior over and above the effects of the actual content of the document (e.g., taxes, interest rates, etc). Finally, our results that analysts provide greater amounts of information to investors for firms with more complex communication and that investors consider this information informative provide interesting insights for the SEC's debate about the intended audience of financial information and may alleviate the SEC's concerns on the accessibility of these reports. While the SEC moves to reduce the complexity of firms' communication, further research is needed to examine the explicit costs and benefits of such actions.

References

Armstrong, J. S. 1980. Unintelligible management research and academic prestige. Interfaces, 10(2): 80-86.

Ball, R. 1992. The earnings price anomaly. Journal of Accounting & Economics, 15(2-3): 319-345.

Barron, O. E., Byard, D., Kim, O., and Riedl, E. J. 2002. High-Technology Intangibles and Analysts' Forecasts, Journal of Accounting Research40(2): 289-312.

Barth, M. E., Kasznik, R., and McNichols, M. F. 2001. Analyst coverage and intangible assets. Journal of Accounting Research, 39(1): 1-34.

Bhushan, R. 1989. Firm characteristics and analyst following. Journal of Accounting & Economics, 11(2-3): 255-274.

Biddle, G. C., Hilary, G., and Verdi, R. S. 2009. How does financial reporting quality related to investment efficiency? Journal of Accounting & Economics, Forthcoming.

Bloomfied, R. J. 2002. The "incomplete revelation hypothesis" and financial reporting. Accounting Horizons, 16(3): 233-243.

Botosan, C. A. and Harris, M. S. 2000. Motivations for a change in disclosure frequency and its consequences: An examination of voluntary quarterly segment disclosures. Journal of Accounting Research, 38(2): 329-353.

Bradshaw, M. T., Miller, G. S., and Serafeim, G. 2008. Accounting method heterogeneity and analysts' forecasts'. Working Paper, Harvard University and University of Michigan.

Brennan, M. J. and Hughes, P. J. 1991. Stock prices and the supply of information. Journal of Finance, 46(5): 1665-1691.

Brennan, M. J. and Subrahmanyam, A. 1995. Investment analysis and price formation in securities markets. Journal of Financial Economics, 38(3): 361-381.

Brown, L. D. and Sivakumar, K. 2003. Comparing the value relevance of two operating income measures. Review of Accounting Studies, 8(4): 561-572.

Callen, J. L., Khan, M., and Lu, Hai. 2009. Accounting Quality, Stock Price Delay and Future Stock Returns. Working Paper, University of Toronto and MIT.

Charrow, R. P. and Charrow, V. R. 1979. Making legal language understandable - Psycholinguistic study of jury instructions. Columbia Law Review, 79(7): 1306-1374.

Chomsky, N. 1965. Aspects of the Theory of Syntax. MIT Press

Core, J.E. 2001. A review of the empirical disclosure literature: discussion. Journal of Accounting & Economics, 31(2): 441-456.

Cotter, J., Tuna, I., and Wysocki, P. D. 2006 Expectations management and beatable targest: How do analysts react to explicit earnings guidance? Contemporary Accounting Research, 23(3): 593-628.

Courtis, J. K. 1995. Readability of annual reports: Western versus Asian evidence. Accounting, Auditing & Accountability, 8(2): 4-17.

Cox, D. R. 1972. Regression models and life-tables. Journal of the Royal Statistical Society Series B-Statistical Methodology, 34(2): 187-220.

Cox, C. 2007. Speech delivered at the Second Annual Corporate Governance Summit at the USC Marshall School of Business, Los Angeles, CA, March 23. Available at http://www.sec.gov/news/speech/2007/spch032307cc.htm.

DeMaris, A. 2004. Regression with Social Data: Modeling Continuous and Limited Response Variables. Hoboken, NJ: John Wiley & Sons.

Duru, A. and Reeb, D. M. 2002. International diversification and analysts' forecast accuracy and bias. Accounting Review, 77(2): 415-433.

Efron, B. 1977. Efficiency of Cox's likelihood function for censored data. Journal of the American Statistical Association, 72(359): 557-565.

Firtel, K. B. 1999. Plain English: A reappraisal of the intended audience of disclosure under the Securities Act of 1933. Southern California Law Review, 72(2-3): 851-897.

Frankel, R., Kothari, S. P., and Weber, J. 2006. Determinants of the informativeness of analyst research. Journal of Accounting & Economics, 41(1-2): 29-54.

Gazmararian, J. A., Baker, D. W., and Williams, M.V. 1999. Health literacy among medicare enrollees in a managed care organization. The Journal of the American Medical Association, 281(6): 545-551.

Gibson E. (1998), Linguistic complexity: Locality of syntactic dependencies. Cognition, 68, 1-76.

Gu, Z. Y. and Chen, T. 2004. Analysts' treatment of nonrecurring items in street earnings. Journal of Accounting & Economics, 38(1-3): 129-170.

Gu, F. and Wang, W. M. 2005. Intangible assets, information complexity, and analysts' earnings forecasts. Journal of Business Finance & Accounting, 32(9-10): 1673-1702.

Gunning, R. 1969. The Fog Index after twenty years. Journal of Business Communication, 6(2): 3-13.

Healy, P., 1977. Can you understand the footnotes to financial statements? Accountants Journal, 219–222.

Healy, P. M., Hutton, A. P., and Palepu, K. G. 1999. Stock performance and intermediation changes surrounding sustained increases in disclosure. Contemporary Accounting Research, 16(3): 485-520.

Hirst, D. E. and Hopkins, P. E. 1998. Comprehensive income reporting and analysts' valuation judgments. Journal of Accounting Research, 36: 47-75.

Hirst, D. E., Hopkins, P. E., and Wahlen, J. M. 2003. Fair values, income measurement, and bank analysts' risk and valuation judgments. Accounting Review, 79(2): 453-472.

Hong, H. and Kubik, J. D. 2003. Analyzing the analysts: Career concerns and biased earnings forecasts. Journal of Finance, 58(1): 313-351.

Hopkins, P. E., Houston, R. W., and Peters, M. F. 2000. Purchase, pooling, and equity analysts' valuation judgments. Accounting Review, 75(3): 257-281.

Indjejikian, R. J. 1991. The impact of costly information interpretation on firm disclosure decisions. Journal of Accounting Research, 29(2): 277-301.

Johnson, D. T. 2004. Has the Security and Exchange Commission's Rule 421 made mutual fund prospectuses more accessible? Journal of Financial Regulation and Compliance, 12(1): 51-63.

Jones, M. J. and Shoemaker, P. A. 1994. Accounting narratives: A review of empirical studies of content and readability. Journal of Accounting Literature, 13: 142-184.

Kimble, J. 1994. Answering the critics of plain language. The Scribes Journal of Legal Writing, 5: 51-85.

Koo, M. M., Krass, I., and Aslani, P. 2003. Factors influencing consumer use of written drug information. Annals of Pharmacotherapy, 37(2): 259-267.

Kripke, H. 1970. The SEC, the accountants, some myths and some realities. New York University Law Review, 45(6): 1151-1205.

Kripke, H. 1973. The myth of the informed layman. Business Law, 28: 631-638.

Lang, M. H. and Lundholm, R. J. 1996. Corporate disclosure policy and analyst behavior. Accounting Review, 71(4): 467-492.

Leuz, C. and Schrand, C. M. 2009, Disclosure and the cost of capital: Evidence from firms' response to the Enron shock. Working Paper, University of Chicago and University of Pennsylvania.

Lang, M. H., Lins, K. V., and Miller, D. P. 2004. Concentrated control, analyst following, and valuation: Do analysts matter most when investors are protected least? Journal of Accounting Research, 42(3): 589-623.

Lee, S. Y. D., Gazmararian, J. A., and Arozullah, A. M. 2006. Health literacy and social support among elderly Medicare enrollees in a managed care plan. Journal of Applied Gerontology, 25(4): 324-337.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. Journal of Accounting & Economics, 45(2-3): 221-247.

Li, Y., Rau, P. R., and Xu, J. 2009. All-star analyst coverage and corporate transparency. Working Paper, Purdue University.

McNichols, M. and O'Brien, P. C. 1997. Self-selection and analyst coverage. Journal of Accounting Research, 35: 167-199.

Nagar, V., Nanda, D., and Wysocki, P. 2003. Discretionary disclosure and stock-based incentives. Journal of Accounting & Economics, 34(1-3): 283-309.

O'Brien, P. C., McNichols, M. F., and Lin, H. W. 2005. Analyst impartiality and investment banking relationships. Journal of Accounting Research, 43(4): 623-650.

Obrien, P. C. and Bhushan, R. 1990. Analyst following and institutional ownership. Journal of Accounting Research, 28: 55-76.

Ott, B. B. and Hardie, T. L. 1997. Readability of advance directive documents. Image Journal of Nursing Scholarship, 29(1): 53-57.

Phillips, T. J., Daily, C. M., and Luehlfing, M. S. 2007. A note on the readability of professional materials for management accountants. Advances in Management Accounting, 16: 311-318.

Plumlee, M. A. 2003. The effect of information complexity on analysts' use of that information. Accounting Review, 78(1): 275-296.

Previts, G. J., Bricker, R. J., Robinson, T. R., and Young, S. J. 1994. A content analysis of sell-side financial analysts company reports. Accounting Horizons, 8(2): 55-70.

Rock, S., Sedo, S., and Willenborg, M. 2000. Analyst following and count-data econometrics. Journal of Accounting & Economics, 30(3): 351-373.

Rogers, R. K. and Grant, J. 1997. Content analysis of information cited in reports of sell-side financial analysts. Journal of Financial Statement Analysis, 3(1): 17-30.

Roychowdhury, S. and Watts, R. L., 2007. Asymmetric timeliness of earnings, market-to-book and conservatism in financial reporting. Journal of Accounting & Economics 44(1-2): 2-31.

Schipper, K. 1991. Analysts' forecasts. Accounting Horizons, 5(4): 105-121.

Schroeder, M. 2003. SEC Proposes Rules to Improve Disclosure by Public Companies. The Wall Street Journal.

Schroeder, N. and Gibson, C. 1990. Readability of management's discussion and analysis. Accounting Horizons, 4(4): 78-87.

SEC. 1998. A Plain English Handbook: How to Create Clear SEC Disclosure Documents. U.S. Securities and Exchange Commission, Washington, DC.

Shuptrine, F. K. and Moore, E. M. 1980. Even after the Magnuson-Moss Act of 1975, warranties are not easy to understand. Journal of Consumer Affairs, 14(2): 394-404.

Smith, J.E., Smith, N.P., 1971. Readability: a measure of the performance of the communication function of financial reporting. The Accounting Review 46, 552–561.

Stone, C.A., Silliman, E.R., Ehren, B.J., Apel, K, 2005. Handbook of Language and Literacy: Development and Disorders, Guilford Press.

Thompson, C. K. and Faroqi-Shah, Y. 2002. Models of sentence production. In Hillis, A.E. (ed.) Handbook of Adult Language Disorders: Integrating Cognitive Neuropsychology, Neurology, and Rehabilitation (pp. 311-330). Philadelphia: Psychology Press, Taylor and Francis.

You, H. and Zhang, X.-J. 2009. Financial reporting complexity and investor under-reaction to 10-K information. Review of Accounting Studies, 14(4), 559-586.

TABLE 1

Descriptive Statistics for Disclosure Readability Using the Fog Index of the 10-K Filing

Panel A reports descriptive statistics on our measure of disclosure readability, the Fog Index, for observations available on both the SEC's EDGAR and COMPUSTAT databases. The Fog Index is computed as: (words per sentence + percent of complex words) x 0.4, using the text of the 10-K filings for fiscal years 1995-2006. Panel B provides descriptive statistics for the five industries with the highest and lowest mean Fog Index. Industries are based on the Fama and French 48 industry classification.

Panel A. Fog by Fiscal Year (All Compustat Firms)

Year	N	Mean	Std Dev	Q1	Median	Q3	IQ
1995	3,505	19.25	1.36	18.36	19.15	20.05	1.70
1996	5,816	19.42	1.33	18.50	19.34	20.21	1.71
1997	5,857	19.47	1.33	18.56	19.37	20.27	1.72
1998	5,689	19.40	1.27	18.53	19.32	20.18	1.65
1999	5,612	19.24	1.31	18.38	19.10	20.00	1.63
2000	5,518	19.18	1.35	18.30	19.03	19.91	1.61
2001	4,942	19.26	1.43	18.33	19.10	19.96	1.63
2002	4,790	19.57	1.58	18.55	19.35	20.34	1.78
2003	4,556	19.87	1.79	18.74	19.59	20.61	1.87
2004	4,242	19.77	1.91	18.76	19.60	20.64	1.88
2005	3,928	19.82	1.49	18.83	19.64	20.56	1.72
2006	2,962	19.85	1.48	18.86	19.66	20.53	1.68
All Years	57,417	19.48	1.49	18.53	19.33	20.25	1.72

Panel B. Fog by Industry (All Compustat Firms)

High Fog Industries	N	Mean	Std Dev	Q1	Median	Q3	IQ
Healthcare	944	20.22	1.46	19.26	20.07	20.99	1.73
Insurance	2013	20.09	1.28	19.26	20.01	20.81	1.55
Utilities	1638	19.87	1.68	18.91	19.65	20.60	1.69
Telecommunications	1710	19.85	1.47	18.92	19.62	20.52	1.60
Trading	2997	19.84	1.64	18.76	19.69	20.69	1.93

Low Fog Industries	N	Mean	Std Dev	Q1	Median	Q3	IQ
Precious Metals	138	18.30	0.93	17.70	18.22	18.79	1.09
Shipping Containers	143	18.65	1.26	17.87	18.40	19.20	1.33
Agriculture	148	18.82	1.68	17.80	18.78	19.56	1.76
Food Products	738	18.84	1.56	17.93	18.66	19.45	1.53
Defense	93	18.92	1.27	17.83	19.06	19.70	1.87

TABLE 2
Sample Descriptive Statistics

This table reports descriptive statistics for observations that are available on the SEC's EDGAR, COMPUSTAT, CRSP, and I/B/E/S databases. The variables are pooled across fiscal years 1995-2006. Variable defintions are provided below.

Variable	N	Mean	Std. Dev.	Q1	Median	Q3
FOG	37,734	19.47	1.42	18.52	19.32	20.26
#ANALYSTS	37,734	6.77	6.32	2	5	9
RESPONSE - Firm	20,989	17.51	10.58	9.5	15.71	24
RESPONSE - Analyst	8,249	18.45	8.06	12.67	17.25	22.95
AI	35,490	0.1477	0.119	0.0528	0.1186	0.2138
ACCURACY	33,859	0.0349	0.0738	0.0034	0.0103	0.0311
DISPERSION	31,040	0.0075	0.0144	0.0012	0.0027	0.007
SIZE	37,734	2463.46	6322.46	139.98	446.61	1547.37
GROWTH	37,734	0.18	0.3	0.04	0.11	0.22
SEGMENTS	37,734	1.85	1.33	1	1	3
PINST	37,734	48.13	26.15	26.08	48.5	69.09
MFCOUNT	37,734	1.75	2.94	0	0	2
ADV	37,734	0.01	0.03	0	0	0.01
R&D	37,734	0.07	0.16	0	0	0.07
STD_RET	37,734	0.14	0.09	0.08	0.12	0.18

FOG	= The Fog Index of the 10-K filing calculated as .4 x (words per sentence + percent of complex words).
#ANALYSTS	= The number of analysts in the consensus earnings forecast for fiscal year following the 10-K filing.
RESPONSE-Firm	= The average number of days that it takes a firm's analysts to issue their first report following the 10-K filing.
RESPONSE-Analyst	= The average number of days that it takes an analyst to issue her first report following the 10-K filing.
AI	= The sum of the one-day, absolute size-adjusted returns from analyst reports between the 10-K filing and the next fiscal year-end divided by the sum of the one-day, absolute size-adjusted returns over the entire window.
ACCURACY	= The absolute value of the difference between reported earnings and the analyst consensus scaled by price.
DISPERSION	= The standard deviation of the individual analyst forecasts scaled by price.
SIZE	= The number of shares outstanding (Compustat item CSHO) times the share prices at the most recent fiscal year-end (Compustat item PRCC_F).
GROWTH	= The compound average growth rate of firm sales (Compustat item SALE) over the prior 3-5 years $(sales_{t-1}/sales_{t-(1+i)})^{1/i}$.
SEGMENTS	= The number of reported business segments in the Compustat segment file for prior fiscal year.
PINST	= The percentage of institutional ownership from the quarter prior to the 10-K filing.
MFCOUNT	= The number of management earnings forecasts issued in the prior year.
ADV	= Advertising expense (Compustat item XAD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.
R&D	= Research and development expense (Compustat item XRD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.
STD_RET	= The standard deviation of the firm's monthly stock returns from the previous fiscal year.

TABLE 3
Pearson (Spearman) Correlations in Upper (Lower) Triangle

This table reports the correlation of variables for observations that are available on the SEC's EDGAR, COMPUSTAT, CRSP, and I/B/E/S databases. The variables are pooled across fiscal years 1995-2006. Variable definitions are provided below.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) FOG	1	**0.0398**	**0.0521**	**0.0526**	0.0101	**0.0285**	**0.0385**	**0.0490**	**0.0256**	**0.0645**	**0.0412**	**-0.0298**	**0.0799**	**0.0163**
(2) #ANALYSTS	**0.0353**	1	**0.0226**	**0.8201**	**-0.2007**	**-0.1503**	**0.7602**	**0.0147**	**0.1121**	**0.4167**	**0.2254**	**0.0805**	**-0.0321**	**-0.1656**
(3) RESPONSE-Firm	**0.0497**	**0.0642**	1	0.0080	**-0.0208**	**-0.0347**	**0.0838**	**-0.0696**	**0.0610**	**0.1244**	**0.0976**	0.0130	**-0.0068**	**-0.1017**
(4) AI	**0.0460**	**0.8242**	**0.0525**	1	**-0.1733**	**-0.1325**	**0.7268**	**0.0290**	**0.1208**	**0.4809**	**0.2810**	**0.0755**	**-0.0467**	**-0.1212**
(5) ACCURACY	**0.0237**	**-0.3161**	**-0.0230**	**-0.2196**	1	**0.5460**	**-0.3458**	**0.0451**	**-0.0263**	**-0.1961**	**-0.0860**	-0.0103	**0.0312**	**0.2400**
(6) DISPERSION	**0.0407**	**-0.1612**	**-0.0309**	**-0.1292**	**0.5266**	1	**-0.3143**	**0.1104**	**-0.0566**	**-0.1912**	**-0.1269**	**-0.0187**	**0.1563**	**0.2830**
(7) LOGSIZE	**0.0208**	**0.7670**	**0.1100**	**0.7380**	**-0.4173**	**-0.3238**	1	**-0.0533**	**0.2407**	**0.5370**	**0.2857**	**0.0815**	**-0.0887**	**-0.2998**
(8) GROWTH	**0.0539**	**0.0883**	**-0.0783**	**0.0780**	**0.0129**	-0.0097	**-0.0370**	1	**-0.1169**	**-0.1027**	**-0.0462**	**0.0577**	**0.1442**	**0.3138**
(9) LSEGMENTS	**0.0186**	**0.1079**	**0.0592**	**0.1239**	**-0.0128**	**-0.0178**	**0.2283**	**-0.1359**	1	**0.1951**	**0.1590**	**-0.0356**	**-0.1610**	**-0.1056**
(10) PINST	**0.0511**	**0.5178**	**0.1358**	**0.5417**	**-0.1984**	**-0.1469**	**0.5705**	**-0.0766**	**0.1951**	1	**0.3437**	**0.0167**	**-0.0548**	**-0.1916**
(11) MFCOUNT	**0.0171**	**0.2623**	**0.0784**	**0.2953**	**-0.0844**	**-0.1414**	**0.2580**	-0.0070	**0.1618**	**0.3325**	1	**0.0956**	**-0.0718**	**-0.0697**
(12) ADV	**-0.0420**	**0.0552**	**0.0310**	**0.0572**	**-0.0663**	**-0.1004**	**0.0408**	**-0.0147**	**-0.1002**	0.0068	**0.1035**	1	**-0.0548**	**0.0123**
(13) R&D	**0.0581**	-0.0012	**-0.0168**	-0.0067	**0.0773**	**0.1102**	**-0.0610**	**0.0623**	**-0.0482**	**0.0114**	**0.0542**	**-0.0346**	1	**0.3521**
(14) STD_RET	**0.0431**	**-0.1639**	**-0.1226**	**-0.1174**	**0.3336**	**0.3033**	**-0.3422**	**0.2702**	**-0.1028**	**-0.1558**	**0.0108**	**-0.0298**	**0.3842**	1

Correlations which are significantly different from zero at the .05 level are bolded.

FOG	= The Fog Index of the 10-K filing calculated as .4 x (words per sentence + percent of complex words).
#ANALYSTS	= The number of analysts in the consensus earnings forecast for fiscal year following the 10-K filing.
RESPONSE-Firm	= The average number of days that it takes a firm's analysts to issue their first report following the 10-K filing.
AI	= The sum of the one-day, absolute size-adjusted returns from analyst reports between the 10-K filing and the next fiscal year-end divided by the sum of the one-day, absolute size-adjusted returns over the entire window.
ACCURACY	= The absolute value of the difference between reported earnings and the analyst consensus scaled by price.
DISPERSION	= The standard deviation of the individual analyst forecasts scaled by price.
LOGSIZE	= The natural logarithim of the number of shares outstanding (Compustat data item CSHO) times the share prices at the most recent fiscal year-end (Compustat item PRCC_F).
GROWTH	= The compound average growth rate of firm sales (Compustat item SALE) over the prior 3-5 years $(sales_{t-1}/sales_{t-(1+i)})^{1/i}$.
LSEGMENTS	= The natural logarithim of the number of reported business segments in the Compustat segment file for prior fiscal year.
PINST	= The percentage of institutional ownership from the quarter prior to the 10-K filing.
MFCOUNT	= The number of management earnings forecasts issued in the prior year.
ADV	= Advertising expense (Compustat item XAD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.
R&D	= Research and development expense (Compustat item XRD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.
INTANG	= The percentage of firm assets that are recognized intangibles from the prior fiscal year (Compustat item INTAN divided by Compustat item AT).
STD_RET	= The standard deviation of the firm's monthly stock returns from the previous fiscal year.

TABLE 4

The Association Between Annual Report Readability and Analyst Following

Panel A reports coefficient estimates and t-statistics from the regression of analyst following on the 10-K Fog Index and control variables. Analyst following (#ANALYSTS) is defined as the number of analysts contained in the most recent I/B/E/S consensus forecast following the annual report filing. Industry fixed effects are based on the Fama and French 48 industry classification. T-statistics are robust to heteroskedasticity and clustered at the firm level. Panel B reports the estimated effect of an interquartile change in the variable of interest.

Panel A. OLS Regression

Variable	Model 1	Model 2	Model 3
Intercept	-14.3678***	-23.9265***	-12.8946***
	[-17.68]	[-6.11]	[-19.41]
FOG	0.0876***	1.0503***	
	[3.68]	[2.75]	
FOGxFOG		-0.0241**	
		[-2.54]	
FOG Q2			0.1715**
			[2.37]
FOG Q3			0.3536***
			[4.25]
FOG Q4			0.3671***
			[4.02]
LOGSIZE	2.8256***	2.8271***	2.8265***
	[65.76]	[65.81]	[65.78]
GROWTH	0.4496***	0.4429***	0.4423***
	[4.18]	[4.12]	[4.12]
LSEGMENTS	-0.4996***	-0.5003***	-0.5007***
	[-6.25]	[-6.26]	[-6.27]
PINST	0.0080***	0.0080***	0.0081***
	[3.70]	[3.71]	[3.73]
MFCOUNT	0.1093***	0.1093***	0.1094***
	[7.26]	[7.25]	[7.26]
ADV	5.9639***	5.9924***	5.9951***
	[3.45]	[3.47]	[3.47]
R&D	1.7418***	1.7166***	1.7128***
	[5.20]	[5.13]	[5.11]
STD_RET	2.0225***	2.0084***	2.0179***
	[5.19]	[5.16]	[5.18]
Time Fixed Effects	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
N	37,734	37,734	37,734
Adj R^2	0.6403	0.6404	0.6404

*** p<0.01, ** p<0.05, * p<0.1, respectively, two-tailed t-tests.

TABLE 4 - Continued

Panel B. Comparative Interquartile Effects

	Model 1	Model 2	Model 3
FOG	0.1521	0.2010	0.1956
LOGSIZE	6.7894	6.7930	6.7918
GROWTH	0.0825	0.0812	0.0805
LSEGMENTS	-0.5489	-0.5496	-0.5519
PINST	0.3441	0.3441	0.3441
MFCOUNT	0.2186	0.2186	0.2180
ADV	0.0459	0.0461	0.0461
RD	0.1176	0.1159	0.1164
STD_RET	0.2030	0.2015	0.2026

#ANALYSTS	= The number of analysts in the consensus earnings forecast for fiscal year following the 10-K filing.
FOG	= The Fog Index of the 10-K filing calculated as .4 x (words per sentence + percent of complex words).
FOG QX	= Indicator variable set to 1 if the firm's 10-K Fog Index is in quartile X and zero otherwise.
LOGSIZE	= The natural logarithim of the number of shares outstanding (Compustat data item CSHO) times the share prices at the most recent fiscal year-end (Compustat item PRCC_F).
GROWTH	= The compound average growth rate of firm sales (Compustat item SALE) over the prior 3-5 years $(sales_{t-1}/sales_{t-(1+i)})^{1/i}$.
LSEGMENTS	= The natural logarithim number of reported business segments in the Compustat segment file for prior fiscal year.
PINST	= The percentage of institutional ownership from the quarter prior to the 10-K filing.
MFCOUNT	= The number of management earnings forecasts issued in the prior year.
ADV	= Advertising expense (Compustat item XAD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.
R&D	= Research and development expense (Compustat item XRD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.
STD_RET	= The standard deviation of the firm's monthly stock returns from the previous fiscal year.

TABLE 5

Changes in Annual Report Readability and Lead Changes in Analyst Following

Panel A reports coefficient estimates and t-statistics from the regression of lead changes in analyst following on changes in the 10-K Fog Index and control variables. Lead change in analyst following is the difference between the number of analysts following the firm after the next annual report and the number of analysts following the firm after the current report. T-statistics are computed under standard OLS assumptions. For comparision, Panel B reports the estimated effect of an interquartile change in the variable of interest.

Panel A. OLS Regression

Variable	Model 1	Model 2	Model 3
Intercept	-0.0509***	-0.0648***	-0.0694***
	[-4.53]	[-5.32]	[-3.23]
ΔFOG	0.0146**	0.0136*	
	[1.97]	[1.84]	
ΔFOGxΔFOG		0.0064***	
		[2.96]	
ΔFOG Q2			-0.0355
			[-1.19]
ΔFOG Q3			0.0345
			[1.16]
ΔFOG Q4			0.0785***
			[2.63]
ΔLOGSIZE	0.5021***	0.5012***	0.5024***
	[26.13]	[26.07]	[26.15]
ΔGROWTH	0.0540	0.0514	0.0560
	[0.72]	[0.67]	[0.75]
ΔLSEGMENT	-0.1117***	-0.1111***	-0.1115***
	[-3.02]	[-3.00]	[-3.01]
ΔPINST	0.0099***	0.0099***	0.0099***
	[8.87]	[8.83]	[8.82]
ΔMFCOUNT	-0.0150**	-0.0151**	-0.0150**
	[-2.55]	[-2.56]	[-2.55]
ΔADV	0.2425	0.2078	0.2116
	[0.16]	[0.13]	[0.14]
ΔR&D	0.0899	0.0845	0.0905
	[0.28]	[0.27]	[0.29]
ΔSTD_RET	-0.6255***	-0.6224***	-0.6186***
	[-4.82]	[-4.80]	[-4.77]
N	35,644	35,644	35,644
Adj R^2	0.0299	0.0310	0.0320

*** $p<0.01$, ** $p<0.05$, * $p<0.1$, respectively, two-tailed t-tests.

TABLE 5 - Continued

Panel B. Comparative Interquartile Effects

	Model 1	Model 2	Model 3
ΔFOG	0.0186	0.0180	0.1140
ΔLOGSIZE	0.2957	0.2952	0.2959
ΔGROWTH	0.0040	0.0038	0.0042
ΔLSEGMENT	0.0000	0.0000	0.0000
ΔPINST	0.0882	0.0882	0.0882
ΔMFCOUNT	0.0000	0.0000	0.0000
ΔADV	0.0000	0.0000	0.0000
ΔRD	0.0000	0.0000	0.0000
ΔSTD_RET	-0.0415	-0.0413	-0.0410

Lead Δ#ANALYSTS	= The lead change in the number of analysts in the consensus earnings forecast for fiscal year following the 10-K filing.
ΔFOG	= The change in the Fog Index of the 10-K filing calculated as .4 x (words per sentence + percent of complex words).
ΔFOG QX	= Indicator variable set to 1 if the change in a firm's 10-K Fog Index is in quartile X and zero otherwise.
ΔLOGSIZE	= The change in the natural logarithim of the number of shares outstanding (Compustat data item CSHO) times the share prices at the most recent fiscal year-end (Compustat item PRCC_F).
ΔGROWTH	= The change in the compound average growth rate of firm sales (Compustat item SALE) over the prior 3-5 years $(sales_{t-1}/sales_{t-(1+i)})^{1/i}$.
ΔLSEGMENTS	= The change in the natural logarithim number of reported business segments in the Compustat segment file.
ΔPINST	= The change in the percentage of institutional ownership from the quarter prior to the 10-K filing.
ΔMFCOUNT	= The change in the number of management earnings forecasts issued in the prior year.
ΔADV	= The change in advertising expense (Compustat item XAD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.
ΔR&D	= The change in research and development expense (Compustat item XRD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.
ΔSTD_RET	= The change in the standard deviation of the firm's monthly stock returns from the previous fiscal year.

TABLE 6

The Association Between Annual Report Readability and Analyst Report Duration

Analyst report duration is defined as the length of time between the 10-K filing and the first report for each analyst. Panels A, B, and C report descriptive statistics of the differences in the average analyst report duration at the individual analyst report, firm-year, and analyst-year levels. The firm and analyst levels are computed by averaging variables over the unit of observation. The High and Low groups are based on the median of the sample in each group. Stastical differences of means (medians) are computed based on t-tests (Wilcoxon tests). Panel D reports estimates of the relative hazard ratios based on the level of the Fog Index. A proportional Cox model is estimated using fixed effects partial likelihood to condition on firm and analyst effects, separately. Panel E reports coefficient estimates based on ordinary least square regressions at both the firm and analyst levels. On the analyst level, the variables are averaged over the porfortfolio of each analyst-year observation. T-statistics are robust to heteroskedasticity and clustered at the firm and analyst level, respectively.

Panel A. Individual Analyst Reports

Group	N	Mean	Median
High Fog	39,862	18.90	16.00
Low Fog	39,896	17.79	15.00
Difference		1.11***	1.00***

Panel B. Average analyst report duration per firm

Group	N	Mean	Median
High Fog	10,492	17.96	16.18
Low Fog	10,487	17.12	15.00
Difference		0.84***	1.18***

Panel C. Average analyst report duration per analyst

Group	N	Mean	Median
High Fog	4,125	19.69	18.60
Low Fog	4,124	17.22	16.24
Difference		2.47***	2.36***

Panel D. Proportional Cox Hazard Estimates (N=92,525)

	Within Firm			**Within Analyst**		
	Coefficient	**P-value**	**Hazard Ratio**	**Coefficient**	**P-value**	**Hazard Ratio**
FOG	-0.0211	<.0001	0.9790	-0.0100	0.0005	0.9900

*** p<0.01, ** p<0.05, * p<0.1, respectively, two-tailed t-tests for differences of means and Wilcoxon for medians.

TABLE 6 - Continued

Panel E. Regression Analysis

	Firm Avg. Duration	Analyst Avg. Duration
Intercept	10.0223***	-15.4339***
	[8.84]	[-5.66]
FOG	0.2931***	1.4260***
	[5.35]	[10.54]
LOGSIZE	0.0471	0.1195
	[0.78]	[1.31]
GROWTH	-0.9373***	1.2099**
	[-3.17]	[1.98]
LSEGMENTS	0.4886***	0.3929
	[3.40]	[1.53]
PINST	0.0358***	0.1006***
	[9.13]	[12.71]
MFCOUNT	0.1850***	0.0995**
	[7.31]	[2.40]
ADV	3.4866	10.0472*
	[1.06]	[1.69]
R&D	2.4739***	1.8338**
	[4.14]	[2.06]
STD_RET	-10.0091***	-19.4592***
	[-9.27]	[-11.10]
N	21,640	8,249
Adj R^2	0.0291	0.1001

*** p<0.01, ** p<0.05, * p<0.1, respectively, two-tailed t-tests.

FOG	= The Fog Index of the 10-K filing calculated as .4 x (words per sentence + percent of complex words).
LOGSIZE	= The natural logarithim of the number of shares outstanding (Compustat data item CSHO) times the share prices at the most recent fiscal year-end (Compustat item PRCC_F).
GROWTH	= The compound average growth rate of firm sales (Compustat item SALE) over the prior 3-5 years $(sales_{t-1}/sales_{t-(1+i)})^{1/i}$.
LSEGMENTS	= The natural logarithim number of reported business segments in the Compustat segment file for prior fiscal year.
PINST	= The percentage of institutional ownership from the quarter prior to the 10-K filing.
MFCOUNT	= The number of management earnings forecasts issued in the prior year.
GROWTH	= The compound average growth rate of firm sales (Compustat item SALE) over the prior 3-5 years $(sales_{t-1}/sales_{t-(1+i)})^{1/i}$.
ADV	= Advertising expense (Compustat item XAD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.
R&D	= Research and development expense (Compustat item XRD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.
STD_RET	= The standard deviation of the firm's monthly stock returns from the previous fiscal year.

TABLE 7

The Association Between Annual Report Readability and the Information Content of Analysts' Reports

Panel A reports coefficient estimates and t-statistics from the regression of the information content of analyst reports (AI) on the 10-K Fog Index and control variables. Analysts information content is defined as the proportion of firm stock returns related to analyst forecast revisions to the total firm stock return during the time period between the 10-K filing and the subsequent fiscal year-end. Industry fixed effects are based on the Fama and French 48 industry classification. T-statistics are robust to heteroskedasticity and clustered at the firm level. For comparision, Panel B reports the estimated effect of an interquartile change in the variable of interest.

Panel A. OLS Regression

Variable	Model 1	Model 2	Model 3
Intercept	-0.2438***	-0.3317***	-0.2191***
	[-17.42]	[-4.71]	[-19.26]
FOG	0.0015***	0.0103	
	[3.42]	[1.49]	
FOGxFOG		-0.0002	
		[-1.29]	
FOG Q2			0.0025*
			[1.82]
FOG Q3			0.0059***
			[3.84]
FOG Q4			0.0062***
			[3.77]
LOGSIZE	0.0468***	0.0469***	0.0469***
	[70.64]	[70.65]	[70.65]
GROWTH	0.0190***	0.0189***	0.0189***
	[9.43]	[9.40]	[9.37]
LSEGMENTS	-0.0101***	-0.0101***	-0.0102***
	[-6.92]	[-6.92]	[-6.93]
PINST	0.0005***	0.0005***	0.0005***
	[13.69]	[13.70]	[13.72]
MFCOUNT	0.0027***	0.0027***	0.0027***
	[10.22]	[10.21]	[10.22]
ADV	0.0679**	0.0682**	0.0685**
	[2.25]	[2.26]	[2.27]
R&D	0.0049	0.0046	0.0044
	[0.88]	[0.84]	[0.79]
STD_RET	0.0783***	0.0781***	0.0782***
	[10.15]	[10.13]	[10.15]
Time Fixed Effects	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
N	35,490	35,490	35,490
Adj R^2	0.6234	0.6234	0.6235

*** p<0.01, ** p<0.05, * p<0.1, respectively, two-tailed t-tests.

TABLE 7 - Continued

Panel B. Comparative Interquartile Effects

	Model 1	**Model 2**	**Model 3**
FOG	0.0026	0.0044	0.0037
LOGSIZE	0.1125	0.1127	0.1127
GROWTH	0.0035	0.0035	0.0035
LSEGMENTS	-0.0111	-0.0111	-0.0112
PINST	0.0215	0.0215	0.0215
MFCOUNT	0.0054	0.0054	0.0054
ADV	0.0005	0.0005	0.0005
RD	0.0003	0.0003	0.0003
STD_RET	0.0078	0.0078	0.0078

AI	= The sum of the one day, absolute size-adjusted returns from analysts reports between the 10-K filing and the next fiscal year end divided by the sum of the one day, absolute size-adjusted returns over the entire window.
FOG	= The Fog Index of the 10-K filing calculated as .4 x (words per sentence + percent of complex words).
FOG QX	= Indicator variable set to 1 if the firm's 10-K Fog Index is in quartile X and zero otherwise.
LOGSIZE	= The natural logarithim of the number of shares outstanding (Compustat data item CSHO) times the share prices at the most recent fiscal year-end (Compustat item PRCC_F).
GROWTH	= The compound average growth rate of firm sales (Compustat item SALE) over the prior 3-5 years $(sales_{t-1}/sales_{t-(1+i)})^{1/i}$.
LSEGMENTS	= The natural logarithim number of reported business segments in the Compustat segment file for prior fiscal
PINST	= The percentage of institutional ownership from the quarter prior to the 10-K filing.
MFCOUNT	= The number of management earnings forecasts issued in the prior year.
ADV	= Advertising expense (Compustat item XAD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.
R&D	= Research and development expense (Compustat item XRD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.
STD_RET	= The standard deviation of the firm's monthly stock returns from the previous fiscal year.

TABLE 8

The Association Between Annual Report Readability and Analyst Earnings Forecast Accuracy

Panel A reports coefficient estimates and t-statistics from the regression of analyst earnings forecast accuracy on the 10-K Fog Index and control variables. Analyst earnings forecast accuracy (ACCURACY) is defined as the absolute value of the difference between the I/B/E/S actual reported earnings and the most recent analyst consensus earnings forecast following the annual report filing scaled by price. Industry fixed effects are based on the Fama and French 48 industry classification. T-statistics are robust to heteroskedasticity and clustered at the firm level. For comparision, Panel B reports the estimated effect of an interquartile change in the variable of interest.

Panel A. OLS Regression

Variable	Model 1	Model 2	Model 3
Intercept	0.1094***	0.0818	0.1344***
	[5.10]	[1.57]	[6.45]
FOG	0.0014***	0.0042	
	[4.44]	[0.85]	
FOGxFOG		-0.0001	
		[-0.57]	
FOG Q2			-0.0003
			[-0.24]
FOG Q3			0.0034***
			[2.82]
FOG Q4			0.0048***
			[3.77]
LOGSIZE	-0.0131***	-0.0131***	-0.0131***
	[-30.07]	[-30.05]	[-30.07]
GROWTH	-0.0053**	-0.0054**	-0.0054**
	[-2.51]	[-2.52]	[-2.54]
LSEGMENTS	0.0037***	0.0037***	0.0037***
	[3.92]	[3.92]	[3.90]
PINST	-0.0002***	-0.0002***	-0.0002***
	[-7.81]	[-7.80]	[-7.77]
MFCOUNT	0.0003**	0.0003**	0.0003**
	[2.27]	[2.26]	[2.28]
ADV	0.0596***	0.0596***	0.0595***
	[2.89]	[2.89]	[2.89]
R&D	-0.0088*	-0.0089**	-0.0091**
	[-1.96]	[-1.97]	[-2.01]
STD_RET	0.1237***	0.1237***	0.1238***
	[14.06]	[14.05]	[14.07]
Time Fixed Effects	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
N	33,859	33,859	33,859
Adj R^2	0.1776	0.1776	0.1777

*** p<0.01, ** p<0.05, * p<0.1, respectively, two-tailed t-tests.

TABLE 8 - Continued

Panel B. Comparative Interquartile Effects

	Model 1	**Model 2**	**Model 3**
FOG	0.0024	0.0006	0.0051
LOGSIZE	-0.0315	-0.0315	-0.0315
GROWTH	-0.0010	-0.0010	-0.0010
LSEGMENTS	0.0041	0.0041	0.0041
PINST	-0.0086	-0.0086	-0.0086
MFCOUNT	0.0006	0.0006	0.0006
ADV	0.0005	0.0005	0.0005
RD	-0.0006	-0.0006	-0.0006
STD_RET	0.0124	0.0124	0.0124

ACCURACY = The absolute value of the difference between reported earnings and the analyst consensus scaled by price.

FOG = The Fog Index of the 10-K filing calculated as .4 x (words per sentence + percent of complex words).

FOG QX = Indicator variable set to 1 if the firm's 10-K Fog Index is in quartile X and zero otherwise.

LOGSIZE = The natural logarithim of the number of shares outstanding (Compustat data item CSHO) times the share prices at the most recent fiscal year-end (Compustat item PRCC_F).

GROWTH = The compound average growth rate of firm sales (Compustat item SALE) over the prior 3-5 years $(sales_{t-1}/sales_{t-(1+i)})^{1/i}$.

LSEGMENTS = The natural logarithim number of reported business segments in the Compustat segment file for prior fiscal year.

PINST = The percentage of institutional ownership from the quarter prior to the 10-K filing.

MFCOUNT = The number of management earnings forecasts issued in the prior year.

GROWTH = The compound average growth rate of firm sales (Compustat item SALE) over the prior 3-5 years $(sales_{t-1}/sales_{t-(1+i)})^{1/i}$.

ADV = Advertising expense (Compustat item XAD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.

R&D = Research and development expense (Compustat item XRD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.

STD_RET = The standard deviation of the firm's monthly stock returns from the previous fiscal year.

TABLE 9

The Association Between Annual Report Readability and Analyst Earnings Forecast Dispersion

Panel A reports coefficient estimates and t-statistics from the regression of analyst forecast dispersion on the 10-K Fog Index and control variables. Analyst earnings forecast dispersion is defined as the standard deviation of the individual analyst earnings forecasts in the most recent consensus earnings forecast following the annual report scaled by price. Industry fixed effects are based on the Fama and French 48 industry classification. T-statistics are robust to heteroskedasticity and clustered at the firm level. For comparision, Panel B reports the estimated effect of an interquartile change in the variable of interest.

Panel A. OLS Regression

Variable	Model 1	Model 2	Model 3
Intercept	0.0161***	0.0250**	0.0219***
	[5.17]	[2.40]	[7.59]
FOG	0.0003***	-0.0006	
	[4.97]	[-0.56]	
FOGxFOG		0.0000	
		[0.89]	
FOG Q2			0.0004*
			[1.86]
FOG Q3			0.0006***
			[2.63]
FOG Q4			0.0012***
			[4.66]
LOGSIZE	-0.0022***	-0.0022***	-0.0022***
	[-25.75]	[-25.77]	[-25.73]
GROWTH	0.0000	0.0000	0.0000
	[0.05]	[0.06]	[0.04]
LSEGMENTS	0.0005***	0.0005***	0.0005***
	[2.60]	[2.61]	[2.62]
PINST	-0.0001***	-0.0001***	-0.0001***
	[-10.80]	[-10.81]	[-10.77]
MFCOUNT	-0.0001***	-0.0001***	-0.0001***
	[-3.92]	[-3.92]	[-3.89]
ADV	0.0116***	0.0116***	0.0117***
	[2.77]	[2.76]	[2.77]
R&D	0.0092***	0.0092***	0.0092***
	[7.77]	[7.78]	[7.74]
STD_RET	0.0291***	0.0291***	0.0291***
	[14.33]	[14.33]	[14.32]
Time Fixed Effects	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
N	31,040	31,040	31,040
Adj R^2	0.2166	0.2166	0.2164

*** p<0.01, ** p<0.05, * p<0.1, respectively, two-tailed t-tests.

TABLE 9 - Continued

Panel B. Comparative Interquartile Effects

	Model 1	Model 2	Model 3
FOG	0.0005	0.0005	0.0008
LOGSIZE	-0.0053	-0.0053	-0.0053
GROWTH	0.0000	0.0000	0.0000
LSEGMENTS	0.0005	0.0005	0.0005
PINST	-0.0043	-0.0043	-0.0043
MFCOUNT	-0.0002	-0.0002	-0.0002
ADV	0.0001	0.0001	0.0001
RD	0.0006	0.0006	0.0006
STD_RET	0.0029	0.0029	0.0029

DISPERSION	= The standard deviation of the individual analyst earnings forecasts scaled by price.
FOG	= The Fog Index of the 10-K filing calculated as .4 x (words per sentence + percent of complex words).
FOG QX	= Indicator variable set to 1 if the firm's 10-K Fog Index is in quartile X and zero otherwise.
LOGSIZE	= The natural logarithim of the number of shares outstanding (Compustat data item CSHO) times the share prices at the most recent fiscal year-end (Compustat item PRCC_F).
GROWTH	= The compound average growth rate of firm sales (Compustat item SALE) over the prior 3-5 years $(sales_{t-1}/sales_{t-(1+i)})^{1/i}$.
LSEGMENTS	= The natural logarithim number of reported business segments in the Compustat segment file for prior fiscal year.
PINST	= The percentage of institutional ownership from the quarter prior to the 10-K filing.
MFCOUNT	= The number of management earnings forecasts issued in the prior year.
ADV	= Advertising expense (Compustat item XAD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.
R&D	= Research and development expense (Compustat item XRD) as a percentage of operating expense (Compustat item XOPR) from the prior fiscal year.
STD_RET	= The standard deviation of the firm's monthly stock returns from the previous fiscal year.

TABLE 10

Sensitivity of Primary Results to 10-K Length

This table reports estimates of the effect of 10-K Fog and other control variables on analyst following, lead changes in analyst following, analyst revision response time, information content of analyst reports, and analyst earnings forecast accuracy and dispersion after controlling for the length of the 10-K filing. For each regression, the sample is partitioned into 20 groups based on the number of words in the 10-K filing. For the change specification these groups are based on the current change in the number of words. The linear regression model (Model 1) is estimated for each group for each dependent variable. The reported coefficients and t-statistics are based on the mean and standard error of the coefficients across groups.

	Levels Regression						Changes Regression	
Variable	**#ANALYSTS**	**RESPONSE Firm**	**RESPONSE Analyst**	**AI**	**ACCURACY**	**DISPERSION**	**Variable**	**Lead Δ#ANALYSTS**
Intercept	-14.0604***	8.7133***	-10.4883***	-0.2309***	0.1592***	0.0267***	Intercept	-0.1741***
	[-23.16]	[7.41]	[-2.74]	[-20.78]	[10.37]	[18.55]		[-9.71]
FOG	0.0840***	0.3728***	1.2051***	0.0009**	-0.0005***	-0.0001***	ΔFOG	0.0459**
	[5.10]	[6.01]	[6.80]	[2.27]	[-2.92]	[-3.19]		[2.34]
LOGSIZE	2.7947***	0.0284	0.0569	0.0460***	-0.0142***	-0.0024***	ΔLOGSIZE	0.6994***
	[42.67]	[0.48]	[0.45]	[40.47]	[-36.04]	[-28.19]		[23.98]
GROWTH	0.5008***	-1.0563***	1.5075**	0.0210***	-0.0083***	-0.0011	ΔGROWTH	0.1577**
	[5.59]	[-3.17]	[2.36]	[7.08]	[-3.17]	[-1.64]		[2.11]
LSEGMENTS	-0.5299***	0.4591**	0.3874	-0.0108***	0.0025***	0.0002	ΔLSEGMENTS	-0.1330***
	[-19.44]	[2.45]	[1.40]	[-14.33]	[3.51]	[1.41]		[-2.83]
PINST	0.0090***	0.0338***	0.0958***	0.0005***	-0.0002***	-0.0001***	ΔPINST	0.0108***
	[6.22]	[7.03]	[12.05]	[17.89]	[-8.66]	[-11.98]		[7.70]
MFCOUNT	0.1072***	0.1899***	0.1353**	0.0028***	0.0003**	-0.0001**	ΔMFCOUNT	-0.0077
	[9.65]	[8.09]	[2.37]	[8.79]	[2.05]	[-2.49]		[-1.03]
ADV	5.9407***	4.3115	10.4187*	0.0707***	0.0599***	0.0106***	ΔADV	1.2676
	[5.43]	[1.41]	[1.74]	[3.48]	[3.44]	[2.77]		[0.59]
R&D	1.5510***	2.5585**	2.1023***	0.0023	-0.0074	0.0096***	ΔR&D	0.4651
	[7.25]	[2.55]	[3.32]	[0.76]	[-1.47]	[7.50]		[0.91]
STD_RET	1.9570***	-9.5634***	-21.0294***	0.0745***	0.1122***	0.0249***	ΔSTD_RET	-0.8230***
	[5.46]	[-8.56]	[-10.13]	[9.52]	[10.16]	[7.51]		[-5.28]

*** p<0.01, ** p<0.05, * p<0.1, respectively, two-tailed t-tests.