Year and Industry-level Accounting Narrative Analysis: Readability and Tone Variation

Dr Ekaete Efretuei
Newcastle University
ekaete.efretuei@newcastle.ac.uk

Running head: Accounting Narrative Analysis: Readability and Tone Variation

Year and Industry-level Accounting Narrative Analysis: Readability and Tone Variation

ABSTRACT

In this study, I examine variations in the textual complexity of annual report narrative disclosures using the Fog Readability Index and Fin-Neg word list Tone Index given year and industry effects. I analyse accounting narrative Readability and Tone based on firm years, associations between the two narrative measures, and industry data. Tests of the relationship between Readability and Tone show that negative narratives have higher Readability scores, supporting the obfuscation hypothesis that bad news tends to be more difficult to read. A year analysis shows that the negative relationship between Readability and Tone increases in significance over time (2006–2011). An industry analysis shows that the observed obfuscation tends to persist in basic materials; consumer services; financial; technology; and utilities industries. This study shows that considering the effect of variations between industry and firm years can inform annual report textual complexity research and associated empirical analyses.

Keywords: Industry analysis; Accounting narratives; Readability; Tone

JEL Classifications: M41, M49

Data Availability: Data are available from the public sources cited in the text.

I. INTRODUCTION

Accounting research on narrative disclosures uses measures of textual characteristics, such as the Fog Index, to measure disclosure Readability (e.g., Lo, Ramos, and Rogo 2017) and the financial dictionary to measure disclosure Tone (e.g., Huang, Teoh, and Zhang 2014). The use of these measures shows the effects of narrative complexity on accounting information users as measured by Readability (Hwang and Kim 2017), Tone (Huang et al. 2014) and both Tone and Readability (Tan, Wang, and Zhou 2014). Perception of narrative complexity is a function of characteristics of both the text and users' resources (Zakaluk and Samuels 1988). Using measures of textual characteristics offers an objective assessment of one component of narrative complexity, which is independent of users' competencies/resources.

This study analyses annual reports' textual characteristics by using the Fog Index to measure Readability and the Fin-Neg word list (Loughran and McDonald 2011) to measure Tone. The Fog Index measures textual complexity and the Fin-Neg word list measures textual sentiment. I assess the textual characteristics of annual reports and test whether a relationship exists between the Readability and Tone of annual reports. I also assess whether changes in Readability and Tone are year- or industry-specific and examine regulatory and economic implications associated with the observed relationship between the Readability and Tone of annual reports.

The results add to the literature on disclosure Readability and disclosure Tone and to broader research on how economic and business environments affect narrative complexity. The next section outlines my motivations for studying the Readability and Tone of annual reports and offers some background and hypotheses. I then discuss the study's measurement and interpretation of the textual indices. This is followed by a description of the sample selection process, the results and a discussion. Finally, I offer several concluding insights.

II. BACKGROUND AND HYPOTHESIS DEVELOPMENT

Readability

This study views Readability from a broad perspective as the ease of narrative information accessibility. The broader literature studies Readability as elements of a piece of work that affect readers' successful understanding of it (DuBay 2004). The accounting literature defines Readability as the 'effective communication of value relevant information' (Loughran and McDonald 2014). For decades, annual reports have struggled to increase Readability (Pashalian and Crissy 1950), and accounting regulators have acknowledged the need for clarity by making regulations designed to achieve this objective: the United Kingdom (UK) Accounting Standard

Board (ASB) Operating and Financial Review recommended that reports 'be written in a clear style and as succinctly as possible' (Rutherford 2003); the United States 1983 Financial Executive Research Foundation (FERF) recommended that companies produce summary reports to make annual reports more readable (SEC 1995, 35604-35633).

Despite these regulatory requirements, information provided in annual report narratives still appears to be difficult to read with increasingly thicker annual reports and complex narratives (FFSA 2007; Dyer, Lang, and Stice-Lawrence 2017). Consensus on target audiences for annual report communication further exacerbates the problem of narrative complexity. Contention over the choice of target audiences occurs between sophisticated technical investors and average readers. An early attempt to resolve this question using analyses of litigation cases found that information provided in reports should be made clear even to the unsophisticated, ordinary, and uninformed investor (Worthington 1978). More recently, the International Accounting Standard Board (IASB) has focused on investors and creditors as the primary users of reports.[1]

This study investigates report Readability based on the IASB's premise that annual report narratives should be accessible to investors and potential investors (not qualified as sophisticated). Report narratives are intended to effectively communicate firm value by explaining the financial statements and implications of reported figures in accounts. It is widely accepted that effective communication involves writing in a readable style (Subramanian, Insley, and Blackwell 1993; Xu, Zhang, and Chen 2018). Therefore, investigating the syntactical complexity of firms' narrative disclosures can provide insights into that difficulties investors face in reading them. These have been widely documented to be a challenge: You and Zhang (2009) find more investor under-reaction when narratives are complex. Kim, Wang,

[1] The IASB's conceptual framework defines 'primary users' and 'users' as 'existing and potential investors, lenders and other creditors who must rely on general purpose financial reports for much of the financial information they need'.

and Zhang (2019) find that narrative complexity hides adverse information, which eventually leads to a tipping point (i.e., a stock price crash).

Tone

Loughran and McDonald (2011) developed a negative word list, the Fin-Neg list, to evaluate the Tone of disclosure. The authors created a list of 2,337 words with negative implications in the financial context culled from terms appearing in financial reports. The aim was to provide a word list suitable for assessing the Tone of financial text. They find that most of the words identified as negative in the frequently used Harvard IV dictionary should not have negative implications in financial contexts because words such as 'mine' proxy for industry effects while words such as 'taxes' are not negative in the financial context (Loughran and McDonald 2011). Further, the word list captured useful information in annual reports through its significant relationship to firm stock returns (Loughran and McDonald 2011).

Existing evidence further shows that the Fin-Neg word list is relevant beyond the scope of annual reports (e.g. media (Dougal, Engelberg, Garcia, and Parsons 2012) and earnings press releases (Huang et. al. 2014)). Furthermore, using the Fin-Neg word list, associates report Tone and reports of material weaknesses in firms' internal control (Loughran and McDonald 2011). Indicating that firms are reporting material weaknesses, which are more likely to spur negative discussions, shows a positive relationship to the negative word list. The Fin-Neg word list is best suited for the present study due to its focus on the negative Tone of financial disclosure. Negative disclosures are more likely to create opportunities for narrative obfuscation and complexity (Li, 2008). Furthermore, several studies have applied this word list and its variants to examine disclosure Tone (Loughran and McDonald 2015; Henry and Leone 2016; Liu and Moffitt 2016).

Using a negative word list also overcomes the limitations of a positive word list. Positive words can often introduce noise into an analysis because companies using positive words in narrative discussions are more likely to use negation (Loughran and McDonald 2011). For example, firms are more likely to write 'not profitable' than 'did not fail' when reporting in narratives. The limitations of positive word lists and the significance of negative words in business texts have been documented in the accounting literature (Tetlock, Saar-Tsechansky, and Macskassy 2008). This study uses this list of negative words to estimate Tone defined as the positive slant of annual reports, thus providing a robust estimation of firms more likely to include a positive slant in their reports.

While negative words are known to produce a negative reaction from readers, positive words and their negation are more likely to produce a positive or mixed reaction when considered along with the associated negation. Limited attention will cause investors to decipher annual report information inefficiently (Hirshleifer and Teoh 2003). The psychology literature suggests that negative words have a greater impact and are more likely to be fully processed (Tetlock et al. 2008). Other studies have used Loughran and McDonald's (2011) dictionary by employing a combination of positive and negative words (e.g. Chen, Nagar, and Schoenfeld 2018). However, this study focuses on negative words, it aims to investigate reports' obfuscation of pessimism rather than their general disclosure Tone. This study analyses whether negative words tend to be more complex when disclosed in annual reports. It measures Tone to test the associations between disclosure Readability and negative words.

Hypotheses

It is predicted that reports with a negative Tone should be harder to read (obfuscation). This study involves year and industry analyses of the association between the Fog Index and the Tone of corporate annual reports. A number of studies have shown that report Readability is

poor when bad news is reported due to a conscious attempt to obfuscate (Subramanian et al. 1993; Jones and Shoemaker 1994; Li 2008). Other studies have shown that low Readability may not be associated with intentional obfuscation (Bloomfield 2008; Bushee, Gow, and Taylor 2018) and that disclosure complexity can be associated with additional disclosures (Guay, Samuels, and Taylor 2016).

The literature also suggests that bad news may be associated with low Readability because the environment associated with bad news is inherently more complex. Bloomfield (2008) suggests that negative discussions require managers to provide additional explanations, which may require dictionary categories that are more complex. Bushee et al. (2018) suggest that disclosure complexity is likely driven by an attempt to reduce information asymmetry, as complex year/industry data require complex explanations to be informative. If this rationale holds, negative discussions are expected to drive complexity for both years and industries with additional complexities relative to other years and industries, respectively. Industry- and year-specific complex accounting may also drive increases in the Readability of negative discussions such as changes in International Financial Reporting Standards (IFRS) that negatively affect reported income. This study designs an experiment to test whether the observed association between negative disclosures and low Readability varies with time (year analysis) and business environment (industry analysis).

Low Readability can affect language sentiment in two ways: by dampening the effect of disclosure or magnifying it. Tan et al. (2014) found that these two effects are moderated by investor sophistication, as the effect of language sentiment on investors is conditional on Readability. Low Readability can obfuscate the effect of language sentiment either by obfuscating narrative inconsistent with accounting numbers or obfuscating negative discussions in narrative disclosures. Asay, Libby, and Rennekamp (2018) found that bad news

disclosures are less readable than good news disclosures, due to attempts to make good news easier to read. The consensus in the literature is that negative disclosures are more difficult to read, possibly due to attempts to obfuscate and/or because additional disclosures are provided to reduce information asymmetry, which is moderated by the complexity of economic and business environments associated with disclosed bad news. However, the literature does not show how this is affected by specific years and industries. Most studies design experiments that hold this specific complexity constant and do not test the effect of this factor. This study contributes to the literature by testing how year and industry variations influence Readability and Tone. I hypothesise as follows:

H1a: The association between negative disclosures and reading difficulty will vary with year effects.
H1b: The association between negative disclosures and reading difficulty will vary with industry effects.

III. MEASUREMENT/INTERPRETATION

Table 1 depicts summary background information of the Readability and Tone Indexes. Appendix A2 discusses the measurement process used for Readability and Tone. An extract of a cross-section of the Fog Index score for primary education-level and academic research material is shown in Table 2. Primary education-level material is defined as material suitable for those aged seven to 12 years while academic research material is defined as material taken from peer-reviewed journal articles.

<INSERT TABLE 1 HERE>

<INSERT TABLE 2 HERE>

The primary school-level material has an average Fog Index of 8 while the journal articles have an average Fog Index of 15.6. A reader comfortable with primary education material would thus need 7.6 more years of formal education to read a journal article with ease at first reading. The length shows the natural log of the number of words in each document. Word complexity is the percentage of words with more than three syllables weighted by the total number of words, demonstrating the difficulty of words included in a document based on their length. Sentence complexity is the average number of words included in each sentence, which measures the average sentence length in a document.

In Table 3, the interpretation for the Tone Index is based on news articles from the *Financial Times* to illustrate how the Tone disclosure measure is interpreted. The first article titled, '*RIM shares hit by analysts' doubts over new CEO and board changes'*, is classified as an article with negative news content (Article A). The second news article titled, *'Chesapeake leads advance for exploration companies'*, is classified as having positive news content (Article B). Article A, in which nine of 278 words are negative, is a more negative article than Article B, in which five words of 334 are negative. This means that article B has a more positive slant than article A. At the same time, Article C, which also includes nine negative words, is less negative than Article A because Article C includes 520 words while Article A includes 278. The Tone Index measures how negative words impact a reader exposed to all words in a document. It is based on the relative effects of negative words contained in a document in the context of all words in the document. Thus, the score evaluates how much negative words in corporate financial reports will affect readers.

<INSERT TABLE 3 HERE>

IV. SAMPLE SELECTION PROCESS

All of the sampled companies are constituents of the FTSE All-Share (FTALSH) Index, which includes premium listed companies; a company must exhibit more than 0.15 percent of FTSE SmallCap total market capitalisation to be eligible (FTSE 2011)[2]. FTALSH companies are large premium listed companies. The Index's disclosure rules are relevant to this study's research design because the constituent listed companies are obliged to communicate business information to their owners and potential owners in a way that avoids 'the creation or continuation of a false market' in its listed shares' (FSA Instrument 2010).[3] Studies have shown that narrative obfuscation leads to false market creation through investors' underreaction (Huang et al. 2014), earnings management (Lo et al. 2017), or stock price crashes (Kim et al. 2019). This study uses this sample to assess the effect of its narrative communication based on the disclosure environment of FTALSH Index companies, thus assessing the role of narrative complexity in the information communication process.

The FTALSH Index covers approximately 98 percent of the UK's market capitalisation (FTSE 2019). These are firms in the London stock exchange that adhere to strict listing rules, creating a suitable setting for investigating the role of narratives in disclosure communication. Potential stock market reactions to accounting numbers can increase pressures on management to explain the numbers via narrative disclosures. In addition, the broad temporal scope of the FTALSH Index, which dates to 1962 when it was known as the 'FT actuaries all share Index', and of the FTSE 100 and FTSE 250, which date to 1984 and 1992, respectively, makes them suitable for a longitudinal firm-year analysis.

[2] The FTSE All-Share index consists of FTSE 100 (the 100 largest companies), FTSE 250 (the 101st to 350th largest companies), and FTSE SmallCap (the 351st to 620th largest companies). In the regular June annual review, a company with more than 0.15 percent of the full market capitalisation of the FTSE SmallCap Index is added to the FTSE small cap index and current constituents with less than 0.10 percent of FTSE SmallCap full market capitalisation are removed from the FTSE All-share index. See https://www.ftserussell.com/products/indices/uk

[3] Listing rules: Premium listing principle 4 in the FSA handbook.

The first step of the sampling process involves retrieving the list of FTALSH constituents from the Thomson database,[4] which produces a sample of 622 firms. The next step of the data collection process involves retrieving the annual reports of each sampled company for 2000 to 2011 from the Thomson database or from the respective company's website[5] when a report is not included in the database. Using annual reports released between 2000 and 2011 allows for a time series that shows how narrative qualities changed due to external factors that caused historical changes in financial reporting. This approach also provides a sample that shows trends in narrative reporting over time and how these trends have affected the objectives of financial reporting. Using a time series, Li (2008) finds that annual reports have become difficult to read; moreover, a trend analysis of effects of policies over the years showed a drop in reading difficulty after the establishment of the SEC's Plain English rule in 1999.

Data Filtering

Once annual reports are retrieved, a sample of 4,347 firm years is obtained. This step excludes all firms and firm years without annual reports; some companies do not keep historical copies of reports, and some reports are not available for between 2000 and 2011. After all available FTALSH constituents' annual reports are obtained, the next step is to retrieve the reports' text. The text-extraction process excludes all reports with less than 2,000 words as in prior research[6] (Li 2008; Loughran and McDonald 2011). This excludes content such as exhibits and reports with only tabulated information. This filtering process also excludes shorter reports such as management announcements, reports titled 'annual reports' by firms though incomplete, and empty text files. This process reduces the sample from 4,347 to 4,231.

[4] Specifically, data are obtained from the Thomson One Database.
[5] Company websites are accessed through London Stock Exchange individual company pages.
[6] Most PDFs that did not successfully convert to text were locked for editing.

Table 4 illustrates the data elimination process, which arrived at a final sample of 4,226 firm years. Starting with 4,347 firm years of financial reports, the sample decreases to 4,231 after short reports and reports with few words are excluded. Next, after narrative scores obtained for each report are screened, reports with narrative scores with extreme values are excluded, leaving 4,226 firm years.

<INSERT TABLE 4 HERE>

V. ANALYSIS AND RESULTS

Firm Year Descriptive Analysis

Table 5 shows the yearly frequency distribution of the final sample of 4,226 firm years for 2000 to 2011; 2009 includes the most annual reports in the sample, and 2000 includes the least due to the reduced availability of historical reports. As the sample period progresses, the number of reports increases due to company variations in the retention of historical reports.

<INSERT TABLE 5 HERE>

Readability Analysis

The Readability data provide an informative overview of the narrative disclosures of the FTALSH companies for the studied period. Figure 1 Panel A shows how reading difficulty varies over the years. The graph plots the average Fog Index score per firm year. The rising trend shows that the reading difficulty of annual reports increased over the years. Annual reports are easier to read in 2000, which has the lowest average Fog Index. Reading difficulty peaks in 2009 and 2010, when the highest average Fog Index scores appear.

<INSERT FIGURE 1 HERE>

Table 6 Panel A shows descriptive statistics for the Fog Index score for each year presented in Figure 1 Panel A, and t-test values indicating whether the means significantly differ from zero. All firm years have p-values of less than 0.0001 (two-tailed), and unreported values indicate confidence mean intervals of approximately 20 to 23 for all years. This indicates that a reader must complete more than *zero* and between 20 and 23 years of formal education to read report text comfortably. Miller's (2010) two-tailed mean test indicates that the average Fog Index score differs from *zero* and is higher in later years with an average Fog Index of 20.338 in 2006. Other studies have reported similar averages: Bai, Dong, and Hu (2019) reports an average Fog Index of 21.805 for annual reports filed between 2006 and 2011. Lang and Stice-Lawrence (2015) study annual reports across 42 countries including FTALSH firms and report a Fog Index average of 19.520 sampling firms for between 1998 and 2011. Dyer et al. (2017) report a Fog Index average of 21.34 for annual reports filed from 1996 to 2013 while Lee (2012) reports an average Fog Index of 20.558 for 10-Q filings for 2001 to 2007. Table 6 Panel B shows the t-test for mean differences between firm years. The general trend shows that mean levels of reading difficulty for earlier years are significantly different from those of later years. This supports the graph above showing that the reading difficulty of corporate annual reports has increased over the years. Specifically, the difference between mean Readability scores for 2000 and 2010 is 0.8484, significant at the 0.01 level.

<INSERT TABLE 6 HERE>

To illustrate the implications of Readability scores, the Readability variable is divided into its two sub-components: word and sentence complexity. Word complexity measures the percentage of complex words in text. Sentence complexity measures the average length of sentences in a document. For instance, if reports *x* and *y* have word complexity scores of 18.0 and 18.1 percent, respectively, report *y* has 0.1 percent more complex words than report *x*. Conversely, sentence complexity scores of 20.0 and 20.5 for reports *x* and *y*, respectively,

would mean that report *y* includes 0.5 more words in each sentence than report *x*. These figures give report *y* a Fog Index score of *[(18 + 20) *0.4] = 15.20* and report *x* a Fog Index score of *[(18.1 + 20.5) *0.4] = 15.44*, producing a difference of 0.20 between the Fog Index scores of reports *x* and *y*.

To place this in context, from the significant difference in Fog Index scores shown between 2000 and 2010 of 0.8484^{***} in Table 6 Panel B, there are corresponding differences of 0. 7701 percent and 1.3508 in word and sentence complexity, respectively. Regarding word complexity, on average, annual reports filed in 2010 have 0.77 percent more complex words than reports filed in 2000. Regarding sentence complexity, 2010 reports include on average 1.35 more words in each sentence than 2001 reports. It is important to note that the relationship between the Fog Index and each one of its two decomposed variables may not be linear because the Fog Index aggregates the two variables, and the movement of the Fog Index will thus depend on the overall effect of the movement of both variables rather than on the effect of the movement of any one separately.

Figure 1 Panel B shows variations in word and sentence complexity for 2000 to 2011. Both variables appear to show similar trends except in 2003, when sentence complexity increases by roughly 24 decimal points while word complexity decreases by approximately 2 decimal points. Fig. 2 shows that word complexity peaks in 2010 while sentence complexity gently maintains its slope around this period. The sentence complexity plot line shows an increase in sentence length from 2000 to 2011; stability is observed in 2004, however. Following this stability, a sharp increase occurs from 2005 to 2007. The application of the International Financial Reporting Standards (IFRS) began in 2005. The plot line for word complexity shows a slight drop in 2003 and a sharp increase from 2004 to 2006. The trends of both variables support evidence of an increase in the reading difficulty of annual reports from 2000 to 2011.

Disclosure Tone Analysis

Figure 2 illustrates variations in annual report Tone for firms in the FTALSH Index. Higher figures indicate a more positive slant. A rising line shows increased optimism in report language while a drop in the graph line shows increased pessimism. Annual report Tone drops twice in the sample period, in 2002 and 2009, showing that annual reports are more negative in these years with 2009 recording the most negative score. This is expected given the economic climate amid the aftermath of the financial crisis, which filled annual reports with negative discussion.

<INSERT FIGURE 2 HERE>

Table 7 Panel A shows the Tone mean scores for each year and their p values. Table 7 Panel B reports the results of a one-way ANOVA test of the difference between the means of each firm year. The results show that the means of 2002 and 2006 are significantly different, recording a difference of 0.132. For 2009, which has mostly negative reports, the mean is significantly different from 2006 with a difference of 0.169. Annual reports are significantly more negative in 2002 and 2009 than in 2006 and 2000 (significant at < 0.01). Annual reports published in 2008 and 2009 are significantly more negative than reports published in 2007. These results show a significant difference between the consecutive years of 2007 and 2008, affirming that companies produced reports that were more negative during the 2008 financial crisis. This reveals that the Tone measure is indicative, showing that management narrative discussions are informative of the firm environment.

<INSERT TABLE 7 HERE>

Association between Readability and Tone

To assess the association between Readability and Tone, I check whether annual reports with a less positive slant (i.e., more negative words) are more difficult to read. Firms with bad news

may produce annual reports of a greater reading difficulty for two reasons. The first is based on the management obfuscation hypothesis, which argues that management is more likely to obfuscate information when performance is poor by making annual reports more difficult to read (Li 2008). Second, bad news may be more difficult to articulate, and poorly performing firms are thus likely to have more complex discussions, which increases reading difficulty (Bloomfield 2008).

Firm quintile groups are created for each year to check for an association between reading difficulty and bad news. Figure 3 shows the plot of Readability scores for firms in the highest and lowest quintile groups. Quintile q1 includes firms that released annual reports with the least positive slant (more negative). Quintile q5 includes firms releasing reports with the most positive slant (more positive). The figure shows that the annual reports of group q1 are consistently more difficult to read than those of group q5. From 2000 to 2004, a large gap appears between the reading difficulty of annual reports of these two groups. The closest gap is seen in 2008 during the financial crisis, when most annual reports included negative discussions. However, the positive slant of annual reports of the two groups shows a persistent difference. Table 8 shows that the difference in the means of these two groups is significant at the 0.01 level, indicating that the annual reports of firms with a less positive slant are more likely to be more difficult to read.

<INSERT FIGURE 3 HERE>

<INSERT TABLE 8 HERE>

Industry-level Descriptive Analysis

Pashalian and Crissy (1950) analyse Readability scores by industry and show that industries with larger markets (e.g., the railroad, food, and automobile sectors) appear to have lower reading ease scores than those with more concentrated markets (e.g., the machinery and

supplies, metal, and chemical sectors). However, their conclusions are not accompanied by statistical evidence due to the small and variable sample used. Courtis (1995) tests the Readability of Hong Kong annual reports and finds Readability levels similar to those found in Western studies. The study conducts several tests that compare Readability over time between industries, firm sizes, and profitability levels, but the results are inconclusive.[7] A few studies have also investigated differences in the Readability and Tone of narrative disclosures between industries. Investigations on Readability and company life cycles (Bakarich, Hossain, and Weintrop 2019) and brands (Davis, Horvath, and Gretry 2019) show that a firm's environment is important to the Readability of its disclosures.

Due to high variability in industry sample frequencies, I perform a Kruskal–Wallis test (useful for comparing the medians of more than one sample) to determine if the medians of the industries are equal. This test produces a mean rank for each industry based on the Index analysed. Industry classification is based on the Industry Classification Benchmark (ICB). Figure 4 charts the industry plotted according to the mean rank as produced by the Kruskal–Wallis test. To analyse the industries further, Figure 5 plots the mean rank of super sectors; the plot includes sentence and word complexity rankings for each super sector together with the Fog Index ranking. Super sectors are the first divisions of the industry class in the ICB. The horizontal axis in both plots denotes the mean rank for each industry. It is obtained by ranking the medians of all individual observations in the sample; the ranks are then placed into their industry groups. The mean rank is the mean of the rank of all observations for an industry group.

The basic materials industry, consisting of the chemical and basic resource super sectors (see Table 9), has a mean rank of 2339.31. The industry appears to have one of the highest

[7] Small sample sizes were used.

Readability Index scores (i.e., difficult-to-read reports). The ranking of super sectors shown in Figure 5 reveals that high rankings appear to be largely driven by the chemicals industry with the Fog Index rankings showing the highest mean rank in the industry sample. Sub-sectors of the chemical super sector are sub-sectors 1353-Commodity chemicals and 1357-Specialty chemicals. The consumer goods industry has a lower Fog Index ranking. Of its three super sectors, personal and household goods has the lowest Fog Index ranking.

The consumer services industry has the lowest Fog Index mean rank of the industries. Firms in the consumer services industry produce, on average, annual reports that are easier to read than those in other industries. When the consumer services industry is decomposed into its super sector components, the industry consists of the retail, media, and travel and leisure super sectors. The super sector plot shows that the retail super sector has the lowest Fog Index ranking, which is quite low at 1642 compared to the other super sectors, which all have mean ranks starting from 1900 (except for travel and leisure) and mostly exceed 2000. The low reading ease of these consumer services is likely due to the kinds of services provided. These services either have narrative discussions that are less complex or less prone to feature management discussions that use obfuscation.

The financial industry includes the bank, insurance, real estate, and financial services super sectors. The super sector plot shows that super sectors with the highest mean rank in the industry are the bank and insurance super sectors. The insurance super sector includes the bank, non-life, and life insurance sub-sectors. One of the super sectors with the lowest Fog Index mean ranking – real estate – is in the financial industry. The health care industry has a mean that is significantly higher than the sample mean. The super sector plot shows that the highest means are in the biotechnology and medical supplies sub-sectors. The industrial sector has a mean that is significantly lower than the sample mean; its lowest super sector Fog mean rank

is in industrial goods and services, which includes the waste and disposal services and marine transportation sub-sectors.

The Fog mean rank of annual reports for the technology industry is significantly higher than the sample mean. Sub-sectors contributing to this include 9533 – Computer services and 9537 – Software. This high rank could be due to language used in this industry. The telecommunications industry has an average ranking while the rank for utilities is significantly higher than the sample mean of both the industry and super sector analysis. Word complexity in the financial services and oil and gas industries is significantly higher than that in the other industries. This shows that these industries use more complex words in their annual report narrative communications.

<INSERT FIGURE 4 HERE>

<INSERT TABLE 9 HERE>

<INSERT FIGURE 5 HERE>

The plot presented in Figure 6 on disclosure Tone shows more variability between the industries than shown in the plot on disclosure Readability. The super sector analysis on disclosure Tone shown in Figure 7 indicates a high positive mean rank of 2536.76 for the construction and materials super sector and the lowest positive mean rank of 601.57 for the banks super sector; Readability ranges from 1600 to 2500. The plot shows that annual reports of the bank super sector include more negative discussions. The insurance super sector follows the bank sector closely, consisting of sub-sectors such as property and casualty insurance, which are defined as including ‘companies engaged in accident, fire, automotive, marine, malpractice and other classes of nonlife insurance’. This definition is indicative, as firms in this industry are more likely to include high proportions of negative words in their reports.

The construction & materials and retail super sectors have higher mean ranks of above 1900. These higher mean ranks are indicative of a more positive slant. A sub-sector analysis indicates that the retail industry displays the lowest mean rank; this pertains to the specialised consumer services sub-sector, which is defined as including 'providers of consumer services such as auction houses, day-care centres, dry cleaners, schools, consumer rental companies, veterinary clinics, and hair salons'. These firms are more likely to produce annual reports with a small proportion of negative words. Their business lines are less affected by adverse market conditions because they mostly focus on daily necessities.

<INSERT FIGURE 6 HERE>

<INSERT FIGURE 7 HERE>

Hypothesis Testing: Association between Readability and Tone

In this section, I perform a multivariate test of the association between Readability and Tone, firm year, and industry analysis. This section addresses my research question: whether negative words included in annual reports are difficult to read and whether this persists across years and industries. Two multivariate analyses are conducted to address this research question as shown in equations (1) and (2). In both equations, the dependent variable is the Fog Index of the annual report, and the main independent variable is the Tone. In equation (1), the main independent variable is moderated by years included in the sample to allow for an analysis of the yearly persistence of the observed low Readability of negative narratives. A negative relationship between Readability and the interaction variable will show that difficult-to-read reports are determined by negative discussions included in annual reports for years with significant coefficients. In equation (2), the main independent variable is moderated by industry identifiers to identify the industry persistence of the observed low Readability of negative discussions. The control variables are the typical controls included in Readability studies: firm size, firm

age, firm performance, volatility (earnings and price), and complexity (business and geographical segments). See appendix A1 for variable definitions.

$$Equation\ 1: bFog\_Index = bTONE * YEAR + controls^{8} \quad (1)$$

$$Equation\ 2: bFog\_Index = bTone * IND + controls \quad (2)$$

<INSERT FIGURE 8 HERE>

The year analysis regression illustrated by a graphical representation of coefficients and confidence intervals (see Figure 8) shows that from 2003 to 2011, the positive slant is significantly negatively associated with Readability. This shows that during these years, low Readability is determined by negative discussions in annual reports. The F-test shows that the coefficients differ significantly by 5.27, falling below the 0.000 level. This result is consistent with the hypothesis that the association between negative disclosures and reading difficulty varies with year effects (H1a). Thus, bad news is associated with complex disclosures from 2003 to 2011. This may be driven by complex explanations, inherently complex negative discussions (Bloomfield 2008), informativeness (Bushee et al. 2018), or discussions of complex accounting changes with negative impacts on earnings.

Asay et al. (2018) find that the observed lower Readability of bad news disclosures is likely driven by managers' attempts to make good news easier to read. For example, the insignificant relationship between Readability and Tone noted for 2001 and 2002. In Table 6 Panel A 2001 and 2002 have a comparatively low Fog Index, both falling within the q5 positive Tone quintile reported in Table 8. This supports the view that firms with a more positive Tone are less likely to manage Readability and to manage disclosures to improve Readability, which may affect the average level of disclosure complexity.

[8] Equations (1) and (2) include year and firm fixed effects.

In addition, there was less regulatory scrutiny of annual reports in these earlier years, reducing the motivation to obfuscate negative disclosures. Therefore, 2002 has a more negative Tone Index (see Table 7 Panel A) and a lower Fog Index than 2003 while the Fog Index for 2003 is marginally still within the q5 Tone quintile; 2003 shows a statistically significant negative relationship between Readability and Tone (see Figure 8). The main regulatory change occurring around 2003 involves the introduction of statutory UK operating and financial review[9] and the subsequent application of the IFRS by FTALSH firms in 2005. Applying accounting standards involves considerable judgement with the incentive and scope to make annual reports complex (Daske and Gebhardt 2006; Ball, Robin, and Wu 2003; Gordon, Henry, Jorgensen, and Linthicum, 2017). Lang and Stice-Lawrence (2015) find that Fog is higher under more oversight, which could also be driven by more complex disclosures. Regulatory changes such as the IFRS increase requirements of complex disclosures through, for example, disclosures of financial instruments, fair value, and risk factor disclosures (Dyer et al. 2017)[10]. The research design of the present regression includes firm fixed effects to determine whether the within-firm variation in complexity explains within-firm variations in Tone.

The multivariate results for the industry interaction can help identify industries associated with difficult-to-read negative disclosures. Table 10 Panel A further confirms industries with higher Fog Index scores as reported in Figure 4: Basic materials, Technology and Utilities. Similarly, I find that the Basic materials, Consumer services, Financials, Technology, and Utilities industries tend to produce negative disclosures that are more difficult to read. This result shows

[9] The Statutory Operating and Financial Review (OFR) was introduced by the Accounting Standard Board to enhance narrative reports for listed firms (Rowbottom and Schroeder 2014). Studies have found that the production processes of these OFR narratives can affect report Readability (Rutherford, 2003). The Operating and financial Review has now been formally withdrawn as a statutory requirement for quoted companies (FRC 2008).
10 Studies have found it difficult to predict the expected relationship between the fog index and regulatory scrutiny. It can be argued that increased regulatory scrutiny should enhance the clarity of disclosures given that they are geared towards improving disclosures (e.g. the SEC Plain English Initiative). On the other hand, regulatory oversight can increase the complexity of disclosures through additional requirements (e.g. the IFRS), exacerbating incentives and opportunities to obfuscate.

that these specific business environments tend to produce negative disclosures that are more difficult to read. Thus, the null hypotheses is rejected, as the relationship between Readability and Tone tends to vary across industries and is significant only for the specific industries described above. These results are consistent with hypotheses that associations between Readability and Tone vary across years/industries and that significance varies across year/industry effects and is not statistically significant in some cases. This contribution allows future studies to reflect on assumptions on year and industry effects in the research design stage.

<INSERT TABLE 10 HERE>

The health care and consumer goods industries show insignificant associations. The annual reports of firms in the health care industry have comparatively high proportions of negative words.[11] The Tone of these reports is unlikely to determine their Readability given the inherent nature of the disclosures in these reports. For example, complex words such as 'accident' and 'injuries' are used in strategy discussions on the production of medicines in annual reports of health care firms. In addition, health care firms are likely to disclose more negative discussions given the nature of their business, and this may not be associated with complexity. For example, the word 'claim' appeared 1,268 times in the annual reports of health care firms; however, it is not complex. It is unclear why the association between Readability and Tone is insignificant for consumer goods firms; however, given the nature of its products, the consumer goods industry tends to rely less on annual reports to shape views on performance than on the shop floor. As Table 9 shows, this industry includes dealers in personal goods as well as food and automobile firms.

[11] Unreported results show that the average tone of health care firms reports is 1.1509, reporting the highest level of negative tone among the studied industries.

The above results support the obfuscation thesis positing that reports with a negative Tone tend to be more difficult to read, to degrees that vary across years and industries. The assumption being that the positive slant of disclosures is indicative of the direction of firm performance. In the additional analysis (Figure 10), I test whether the Tone of textual disclosures (positive_slant) is related to actual performance. I perform an additional regression analysis by year and industry wherein positive_slant is the dependent variable, and interactions between firm performance and years or industries are the main independent variables. Firm performance is measured as a firm's operating income after depreciation scaled by total firm assets denoted as 'EARNINGS' in Figure 9. The results reported in Figure 9 Panel A show that the relationship between Tone (positive_slant) and performance differs by year but remains mostly positive, indicating that the reduced disclosure of negative words in annual reports is associated with better performance. The same trend appears in Figure 9 Panel B, where the relationship between Tone (positive_slant) and industry is positive though with considerable variation across industries.

<INSERT FIGURE 9 HERE>

These findings are much in the spirit of Bloomfield's (2008) statement that 'Happy families are alike; ***every unhappy family is unhappy in its own way***' (emphasis added). I find that not all industries tend to have difficult-to-read negative discussions and find a difference in the coefficients of industries with significant associations. Utilities tend to have higher coefficients at -1.551, showing that for firms in this industry, a 1.5 percent decrease in positive slant increases the Readability Fog Index by one. In the earlier industry analysis, it had one of the higher Readability scores. On the other hand, the telecommunications industry has a low Fog Index and, on average, a higher positive slant. As Table 10 shows, a more positive slant lowers the Fog Index. While most studies acknowledge complexity in research design by holding it constant, the present findings provide practical implications for academics, practitioners, and

policy makers by showing which industries tend to have more complex negative discussions. Moreover, the industry-specific terminology may drive greater complexity of negative words. This question could be further investigated to assess whether certain industry-specific words in the word list drive complexity.

VI. CONCLUSION

This study explains methods used to produce narrative measures of annual reports. The study illustrates the motivations for narrative measures and explains the measurement of the Fog Index and the positive slant of narrative disclosures. Based on measures discussed, and the measurement process used, it discusses the validity of the measurement process and provides a framework for interpreting Readability and Tone Index results. The study provides a preliminary sample description of changes in narratives' reading difficulty and Tone across firm years. It also provides evidence of an association between the reading difficulty and Tone of annual report narratives and conducts an industry-level analysis revealing significant differences between industries' annual report narratives. The hypothesis tests show that the significant negative relationship between Readability and Tone observed in several studies has existed from 2003. The industry analysis shows that the relationship varies across industries and is significant in five industries: basic materials, consumer services, finance, technology, and utilities.

REFERENCES

Asay, H. S., R. Libby, and K. Rennekamp. 2018. Firm performance, reporting goals, and language choices in narrative disclosures. *Journal of Accounting and Economics* 65 (2-3): 380-398.

Bai, X., Y. Dong, N. Hu, 2019. Financial report readability and stock return synchronicity. Applied Economics, 51(4): 346–363.

Bakarich, K. M., M. Hossain, and J. Weintrop. 2019. Different time, different tone: Company life cycle. *Journal of Contemporary Accounting and Economics* 15 (1): 69-86.

Ball, R., A. Robin, J. S. Wu. 2003. Incentives versus standards: properties of accounting income in four East Asian countries. *Journal of Accounting and Economics* 36 (1–3): 235-270.

Bloomfield, R. 2008. Discussion of 'Annual Report Readability, Current Earnings, and Earnings Persistence'. *Journal of Accounting and Economics* 45 (2-3): 248-252.

Bushee, B. J., I. D. Gow, and D. J. Taylor. 2018. Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research* 56 (1): 85-121.

Chen, J. V., V. Nagar, and J. Schoenfeld. 2018. Manager-analyst conversations in earnings conference calls. *Review of Accounting Studies* 23 (4): 1315-1354.

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

Daske, H. and G. Gebhardt. 2006. International financial reporting standards and experts' perceptions of disclosure quality. *Abacus*, 42 (3-4): 461-498.

Davis, A. K., and I. Tama-Sweet. 2012. Managers' use of language across alternative disclosure outlets: Earnings press releases versus MD&A. *Contemporary Accounting Research* 29 (3): 804-837.

Davis, S. W., C. Horváth, A. Gretry, and N. Belei. 2019. Say what? How the interplay of tweet readability and brand hedonism affects consumer engagement. *Journal of Business Research* 100: 150-164.

Dougal, C., J. Engelberg, D. Garcia; and C. A. Parsons. 2012 "Journalists and the Stock Market." *Review of Financial Studies* 25 (3): 639–79.

DuBay, W. 2004. Principles of readability. *Impact Information, Costa Mesa, California.*

Dyer, T., M. Lang, and L. Stice-Lawrence. 2017. The evolution of 10-K textual disclosure: Evidence from latent Dirichlet Allocation. *Journal of Accounting and Economics* 64 (2): 221-245.

FFSA 2007. Collection of essays: Investor perspectives on IFRS implementation Fédération FranÇaise des Sociétés d'Assurances.

FRC. 2008. Operating and financial review: Financial Reporting Council, England.

FSA. Instrument 2010. Listing rules sourcebook (Amendment No 4). In *LR 7: Listing Principles: Premium Listing*. FSA, UK: 15.

FTSE. 2011. Ground rules for the management of The UK series of the FTSE Actuaries Share Indices. In *Version 11.1, FTSE The Index Company*.

FTSE. 2019. FactSheet. *FTSE ALL-SHARE INDEX*. https://www.ftserussell.com/products/indices/uk. Access date: 31/12/2019.

Guay, W., D. Samuels, and D. Taylor. 2016. Guiding through the Fog: Financial statement complexity and voluntary disclosure. *Journal of Accounting and Economics* 62 (2–3):234-269.

Gordon, E. A., E. Henry, B. N. Jorgensen, and C. L. Linthicum 2017. Flexibility in cash-flow classification under IFRS: determinants and consequences. *Review of Accounting Studies* 22 (1): 839–872.

Gurun, U. G., and A. W. Butler. 2012. Don't believe the hype: Local media slant, local advertising, and firm value. *The Journal of Finance* 67 (2): 561-598.

Henry, E., and A. J. Leone. 2016. Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review* 91 (1): 153-178.

Hirshleifer, D., and S. H. Teoh. 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* 36 (1-3): 337-386.

Huang, X., S. H. Teoh, and Y. Zhang. 2014. Tone management. *Accounting Review* 89 (3): 1083-1113.

Hwang, B. H., and H. H. Kim. 2017. It pays to write well. *Journal of Financial Economics* 124 (2): 373-394.

Jones, M. J., and P. A. Shoemaker. 1994. Accounting narratives: A review of empirical studies of content and readability. *Journal of Accounting Literature* 13: 142-184.
Kim, C. F., K. Wang, and L. Zhang. 2019. Readability of 10-K reports and stock price crash risk. *Contemporary Accounting Research*. 36 (2): 1184-1216.
Lang, M., L. Stice-Lawrence 2015. Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economic*s 60 (2–3): 110–13.
Lee, Y. 2012. The effect of quarterly report readability on information efficiency of stock prices. *Contemporary Accounting Research* 29(4): 1137–1170.
Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2–3): 221-247.
Liu, Y., and K. C. Moffitt. 2016. Text mining to uncover the intensity of SEC comment letters and its association with the probability of 10-K restatement. *Journal of Emerging Technologies in Accounting* 13 (1): 85-94.
Lo, K., F. Ramos, and R. Rogo. 2017. Earnings management and annual report readability. *Journal of Accounting and Economics* 63 (1): 1-25.
Loughran, T., and B. McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance* 66 (1): 35-65.
———. 2014. Measuring readability in financial disclosures. *The Journal of Finance* 69 (4): 1643-1671.
———. 2015. The use of word lists in textual analysis. *Journal of Behavioral Finance* 16 (1): 1-11.
Miller, B. 2010. The effects of reporting complexity on small and large investor trading. *The Accounting Review* 85 (6): 2107-2143.
Pashalian, S., and W. J. E. Crissy. 1950. How readable are corporate annual reports? *Journal of Applied Psychology* 34 (4): 244-248.
Rowbottom, N. and Schroeder, M.A.S. (2014), The rise and fall of the UK operating and financial review. *Accounting, Auditing & Accountability Journal* 27 (4): 655-685.
Rutherford, B.A. (2002), The production of narrative accounting statements: An exploratory study of the operating and financial review. *Journal of Applied Accounting Research*, 6 No. 3: 25-56.
Rutherford, B. A. 2003. Obfuscation, textual complexity and the role of regulated narrative accounting disclosure in corporate governance. *Journal of Management and Governance* 7 (2): 187-210.
Security and Exchange Commission (SEC). 1995. Use of Abbreviated Financial Statements in Documents Delivered to Investors Pursuant to the Securities Act of 1933 and Securities Exchange Act of 1934. *Federal register* 60 (131): 35604-35633. Available at: https://www.federalregister.gov/documents/1995/07/10/95-16389/use-of-abbreviated-financial-statements-in-documents-delivered-to-investors-pursuant-to-the Accessed date: (last accessed June, 30, 2020).
Subramanian, R., R. G. Insley, and R. D. Blackwell. 1993. Performance and readability: A comparison of annual reports of profitable and unprofitable corporations. *Journal of Business Communication* 30 (1): 49-61.
Tan, H.-T., E. Ying Wang, and B. O. Zhou. 2014. When the use of positive language backfires: The joint effect of tone, readability, and investor sophistication on earnings judgments. *Journal of Accounting Research* 52 (1): 273-302.
Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance* 63 (3): 1437-1467.
Worthington, J. S. 1978. Footnotes: Readability or liability. *The CPA Journal (pre-1986)* 48 (000005): 27.
Xu, Y., L. Zhang, and H. Chen. 2018. Board age and corporate financial fraud: An interactionist view. *Long Range Planning* 51 (6): 815-830.
You, H., and X.-j. Zhang. 2009. Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting Studies* 14 (4): 559-586.
Zakaluk, B. L., and S. J. Samuels. 1988. Readability–Its Past, Present, and Future. Newark, Delaware: International Reading Association.

TABLES

Table 1

Summary of Background Information - Indexes

	Fog Index	***Fin-Neg List Tone Index***
Index Name	The Gunning Fog Index	Financial Dictionary – Fin-Neg Word List
Developer of Index	Robert Gunning	Tim Loughran and Bill McDonald
Basis for categorization	Word and sentence length	Negative words in financial context
Motivation for the Index	To assess readability of English text	To assess tone in financial text
Formula for assigning a score to a report	Fog Index = (Words per sentence + Percent complex words) *0.4.	The word list counts weighted by the total number of words (proportional weights)
Relevance of the Index in accounting research	Assesses the complexity of words and sentences in financial reports	Assesses tone of disclosures in financial report narratives

Table 2

Fog Index of Primary Education-level Material (PR) and Academic Research-level Material (JR)

Name	***Words***	***Fog_Index***	***Word_Comp***	***Sent_Comp***	***Length***	***Average Fog_Index***
PR	1365	9.265201	3.663004	19.50000	7.21891	
PR	758	8.569062	8.575198	12.84746	6.630683	
PR	13930	4.999841	4.544149	7.955454	9.5418	
PR	23842	9.308544	6.350138	16.92122	10.0792	
PR (Average)						***8.035662***
JR	15320	15.70907	28.54439	10.72829	9.636914	
JR	9911	17.12315	30.30976	12.49811	9.201401	
JR	16837	15.74597	28.1701	11.19481	9.731334	
JR	14201	13.77206	24.02648	10.40366	9.561068	
JR (Average)						***15.58756***

Words is the total number of words in the document. *Length* is natural logarithm of number of words in the document. All other variables are as defined in Appendix A1: Variable Definitions.

Table 3

Tone of Sample Articles in *Financial Times*

Article Type.	***Tone***	***Length***	***Neg_Words***	***Words***
Article A:				
'*RIM shares hit by analysts' doubts over new CEO and board changes.*	-3.23741	5.627621	9	278
Article B:				
Chesapeake leads advance for exploration companies	-1.497006	5.811141	5	334
Others				
Article C	-1.730769	6.253829	9	520
Article D	-2.29682	7.437206	39	1698
Article E	-1.351351	5.913503	5	370

Article A is titled '*RIM shares hit by analysts' doubts over new CEO and board changes. Article B is titled Chesapeake leads advance for exploration companies. Length* is natural logarithm of number of words in the document. *Neg_Words* is the raw count of the number of words in the document. *Words* is the total number of words in the document. All other variables are defined in Appendix A1: Variable Definitions.

Table 4

Data Elimination Process - Narrative Sample

Event	Firm Years after Event
Initial annual report collections/downloads	4,347
Eliminate interim report data presented as annual	4,268
Eliminate observations with fewer than 2000 words	4,231
Eliminate reports with extreme values	4,226
Final narrative firm year observations =	4,226

Table showing the data-elimination process; 'Event' denotes the relevant action that produces the number of firm years; 'Firm Years after Event' denotes the number of firm years arrived at after the relevant event.

Table 5

Firm Year Analysis

YEAR	Frequency	% Freq	Cum_Freq	Cum_%
2000	146	3.45	146	3.45
2001	200	4.73	346	8.19
2002	238	5.63	584	13.82
2003	288	6.81	872	20.63
2004	331	7.83	1203	28.47
2005	378	8.94	1581	37.41
2006	432	10.22	2013	47.63
2007	489	11.57	2502	59.2
2008	508	12.02	3010	71.23
2009	524	12.4	3534	83.63
2010	510	12.07	4044	95.69
2011	182	4.31	4226	100

YEAR is the year of analysis. *Frequency* is the occurrence per year. *%_Freq* is frequency in percentage. *Cum_Freq* is cumulative frequency. *Cum_%* is cumulative frequency in percentage.

Table 6

Descriptive Statistics of Disclosure Readability

Panel A: Firm Year Descriptive Statistics – Readability

YEAR	N	Mean	p50	sd	p1	p25	p75	p99	t Value	Pr > |t|
2000	146	21.3794	22.5613	1.6304	17.3400	20.3706	22.5613	24.7694	158.45	<.0001
2001	200	21.4446	22.5061	1.5595	17.3848	20.3827	22.5061	25.2524	194.46	<.0001
2002	238	21.6164	22.6324	1.5165	17.3912	20.7438	22.6324	25.1767	219.89	<.0001
2003	288	21.7086	22.6311	1.4783	17.0257	20.8980	22.6311	25.6466	249.21	<.0001
2004	331	21.7899	22.7318	1.5059	17.5693	21.0600	22.7318	25.6424	263.26	<.0001
2005	378	21.9128	22.7346	1.3350	18.2773	21.1509	22.7346	25.0862	319.13	<.0001
2006	432	22.0958	22.8424	1.2504	18.3596	21.4324	22.8424	24.8906	367.29	<.0001
2007	489	22.1973	22.9441	1.1765	18.7873	21.5215	22.9441	25.3233	417.22	<.0001
2008	508	22.2143	22.9221	1.2297	19.0269	21.5283	22.9221	25.2420	407.16	<.0001
2009	524	22.2284	22.9515	1.1808	18.8920	21.4887	22.9515	25.3459	430.92	<.0001
2010	510	22.2278	22.9380	1.2102	18.8736	21.5881	22.9380	24.9988	414.77	<.0001
2011	182	22.1427	22.8512	1.2236	19.0068	21.2596	22.8512	26.3584	244.14	<.0001

A one-sample t-test was run to determine whether the Readability mean scores by year differ from zero. The statistical significance values 't Value' and 'Pr>|t|' are presented as two-tailed with no prediction of Readability scores. YEAR is the year of analysis. Readability is measured using the *Fog_Index* variable defined in Appendix A1: Variable Definitions.

Panel B: Test of Mean Difference between Firm Year Readability Scores

I \ J	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
2001	.0651										
2002	.2369	.1718									
2003	.3291	.2640	.0922								
2004	.4104	.3453	.1736	.0813							
2005	**.5334*****	**.4682*****	.2964	.2042	.						
2006	**.7163*****	**.6512*****	**.4794*****	**.3872*****	**.3059*****	.1829					
2007	**.8179*****	**.7527*****	**.5809*****	**.4887*****	**.4074*****	**.2844*****	.				
2008	**.8348*****	**.7697*****	**.5979*****	**.5057*****	**.4244*****	**.3015*****	.1185	.0170			
2009	**.8489*****	**.7838*****	**.6120*****	**.5198*****	**.4384*****	**.3155*****	.1326	.0310	.0140		
2010	**.8484*****	**.7832*****	**.6115*****	**.5192*****	**.4379*****	**.3150*****	.1320	.0305	.0135	.0005	
2011	**.7632*****	**.6981*****	**.5264*****	**.4341*****	.3528	.2299	.0469	-.0545	-.0716	-.0857	-.0851

This table presents the difference in the yearly mean reading difficulty of annual reports. Firm years run from 2000 to 2011. Mean difference is '**I – J'.** ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively. Readability is measured using the *Fog_Index* variable defined in Appendix A1: Variable Definitions.

Table 7

Descriptive Statistics of Disclosure Tone

Panel A: Firm Year Descriptive Statistics - *TONE*

YEAR	N	Mean	p50	sd	p1	p25	p75	p99	t Value	Pr > |t|
2000	146	-0.8814	-0.8765	0.2266	-1.5725	-0.9988	-0.7369	-0.3918	-47	<.0001
2001	200	-0.9806	-0.9385	0.2855	-2.0369	-1.1240	-0.7798	-0.5225	-48.57	<.0001
2002	238	-1.0684	-1.0497	0.2728	-1.7541	-1.2515	-0.8836	-0.5549	-60.43	<.0001
2003	288	-1.0166	-0.9701	0.2605	-1.8278	-1.1692	-0.8294	-0.5613	-66.24	<.0001
2004	331	-0.9844	-0.9542	0.2552	-1.7553	-1.1315	-0.8144	-0.5187	-70.17	<.0001
2005	378	-0.9512	-0.9128	0.2478	-1.6751	-1.0962	-0.7815	-0.4706	-74.63	<.0001
2006	432	-0.9365	-0.9048	0.2386	-1.6878	-1.0654	-0.7717	-0.4986	-81.59	<.0001
2007	489	-0.9552	-0.9124	0.2437	-1.7702	-1.0715	-0.7965	-0.5406	-86.67	<.0001
2008	508	-1.0652	-1.0222	0.2356	-1.8414	-1.1799	-0.9078	-0.6685	-101.89	<.0001
2009	524	-1.1060	-1.0716	0.2407	-1.8959	-1.2174	-0.9508	-0.7256	-105.16	<.0001
2010	510	-1.0391	-1.0098	0.2318	-1.7988	-1.1262	-0.8977	-0.6405	-101.25	<.0001
2011	182	-0.9846	-0.9442	0.1854	-1.5985	-1.0747	-0.8590	-0.6451	-71.66	<.0001

A one-sample t-test was run to determine whether the Tone mean scores by year differ from zero. The statistical significance values 't Value' and 'Pr>|t|' are presented as two-tailed with no prediction of Tone scores. YEAR is the year of analysis. Tone is measured using the TONE variable defined in Appendix A1: Variable Definitions.

Panel B: Test of the Mean Difference between Firm Year *TONE* Scores

I \ J	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
2001	**-.099****										
2002	**.187*****	**.088****									
2003	**-.135*****	-.036	.052								
2004	**-.103*****	-.003	**.084*****	-.032							
2005	-.069	.029	**.117*****	**.065****	.033						
2006	-.055	.0441	**.132*****	**.080*****	.048	.015					
2007	**-.074*****	.025	**.113*****	**.061****	.029	-.004	-.019				
2008	**-.184***	**-.085*****	.003	-.049	**-.081*****	**-.114***	**-.129*****	**-.110*****			
2009	**-.224*****	**-.125*****	-.0375	**-.089*****	**-.122*****	**-.155*****	**-.169*****	**-.151*****	-.040		
2010	**-.158*****	-.058	.029	-.022	**-.055***	**-.088*****	**.103*****	**.084*****	.026	**-.067*****	
2011	**-.103*****	-.004	**.083*****	.032	-.002	-.033	-.048	-.029	**.080*****	**.121*****	.054

This table presents the difference in the yearly mean tone of annual reports. Firm years run from 2000 to 2011. Mean difference is 'I –J'. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively. Variable descriptions are defined in Appendix A1: Variable Definitions.

Table 8

Test of Mean Difference of Readability Scores of Tone Quintile Groups

Quintile Rank for Variable TONE	**Mean**
q1 most negative	22.28971
q2	22.11605
q3	21.99548
q4	21.90742
q5 most positive	21.79406
Change in Tone Quintile group annual report readability q1 less q5	**0.49565*** **
t Value	(7.464)

Table 9

Key for Industry Classification

IND	**ICB_INDUSTRY**	**ICB_SUPER SECTOR**
BAM	Basic Materials	Basic Resources Chemicals
CMG	Consumer Goods	Personal & Household Goods Food & Beverage Automobiles & Parts
CMS	Consumer Services	Retail Travel & Leisure Media
FIN	Financials	Other Financial Services Financial Services Banks Insurance
HTC	Health Care	Health Care
INS	Industrials	Industrial Goods & Services Construction & Materials
ONG	Oil and Gas	Oil & Gas
TEC	Technology	Technology
TEL	Telecommunications	Telecommunications
UTL	Utilities	Utilities

Table 10

Industry Analysis on *Relationship between Readability and Tone*

Panel A: Summary for variable: *Fog_Index* by categories of *IND* (ICB_INDUSTRY)

IND	Mean	p50	sd	p1	p25	p75	p99	N
Basic Materials	22.3075	22.3281	1.2044	19.3003	21.6896	23.0509	25.0720	247
Consumer Services	21.8019	21.8790	1.3528	18.3460	21.0632	22.6370	25.3018	702
Consumer Goods	21.9459	22.0025	1.6082	17.4516	21.1700	22.9240	25.2168	303
Financials	22.0745	22.1168	1.4057	17.9969	21.3040	22.9135	25.7380	1290
Health Care	22.1757	22.1425	0.8609	19.6554	21.6210	22.7279	24.3210	98
Industrial	21.8427	21.9809	1.2934	18.3027	21.1994	22.6780	24.4511	988
Oil and Gas	22.1872	22.0975	1.2480	19.5896	21.4099	22.8002	26.2373	170
Technology	22.2621	22.3059	1.1736	18.9042	21.4638	23.0756	24.9145	212
Telecommunications	21.9984	21.8865	0.9702	18.9844	21.2806	22.7091	24.1087	60
Utilities	22.2748	22.4089	0.9415	18.3886	21.8144	22.9726	23.8737	80

Panel B: Regression results by categories of *IND* (ICB_INDUSTRY)

Variable	BAM	CMG	CMS	FIN	HTC	INS	ONG	TEC	TEL	UTL
TONE	-0.948 (-2.56)	-1.079 (-2.69)	-0.463 (-0.93)	-0.709 (-2.6)	1.056 (0.98)	-0.617 (-1.6)	-0.937 (-1.72)	-1.169 (-2.25)	-1.768 (-1.99)	-1.551 (-3.17)
controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
year and firm fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
robust	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	*135*	*461*	*226*	*439*	*72*	*749*	*105*	*150*	*40*	*68*
r2	*0.1709*	*0.0488*	*0.0652*	*0.0415*	*0.371*	*0.0817*	*0.1145*	*0.1624*	*0.3655*	*0.4412*

Industry Classifications are defined in Table 9: Key for Industry Classification.
Variable descriptions are defined in Appendix A1: Variable Definitions.

FIGURES

Figure 1

Panel A: Firm Year Analysis – Disclosure Readability of Annual Reports

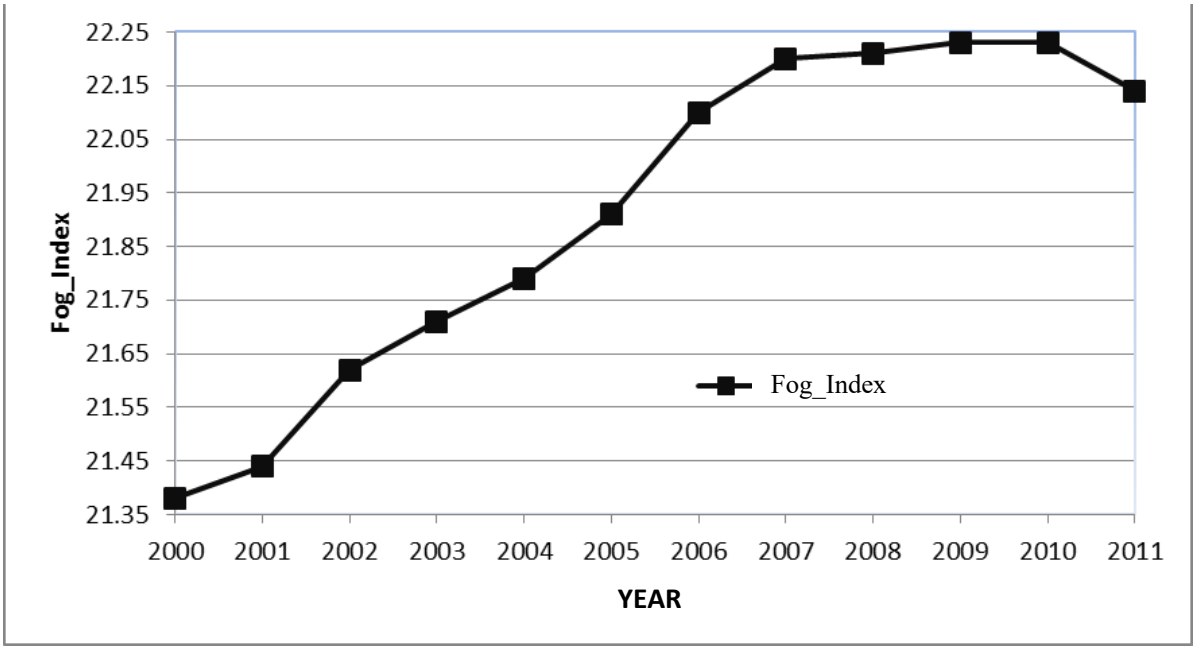


Figure showing the mean reading difficulty of annual report narratives from 2000 to 2011. YEAR is the year of analysis. Variable descriptions are defined in Appendix A1: Variable Definitions.

Panel B: Firm Year Analysis – Disclosure Complexity of Annual Reports

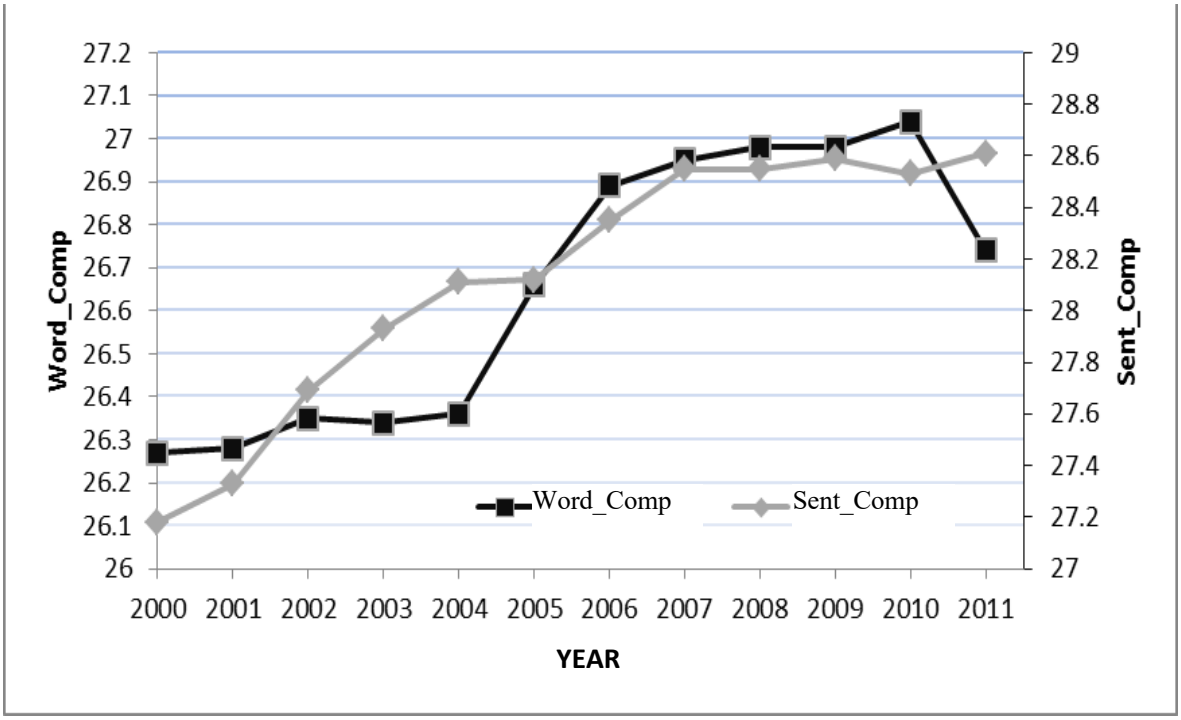


Figure showing the word complexity and sentence complexity of annual reports from 2000 to 2011. YEAR is the year of analysis. Variable descriptions are defined in Appendix A1: Variable Definitions.

Figure 2

Firm Year Analysis - Disclosure Tone of Annual Reports

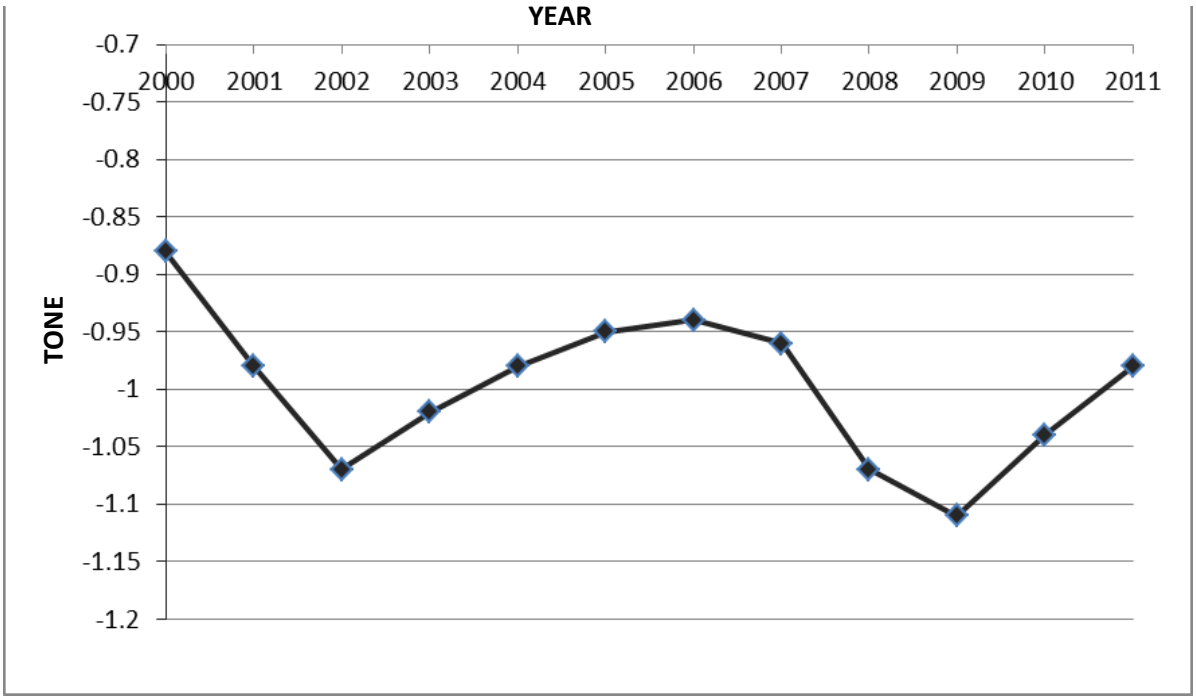
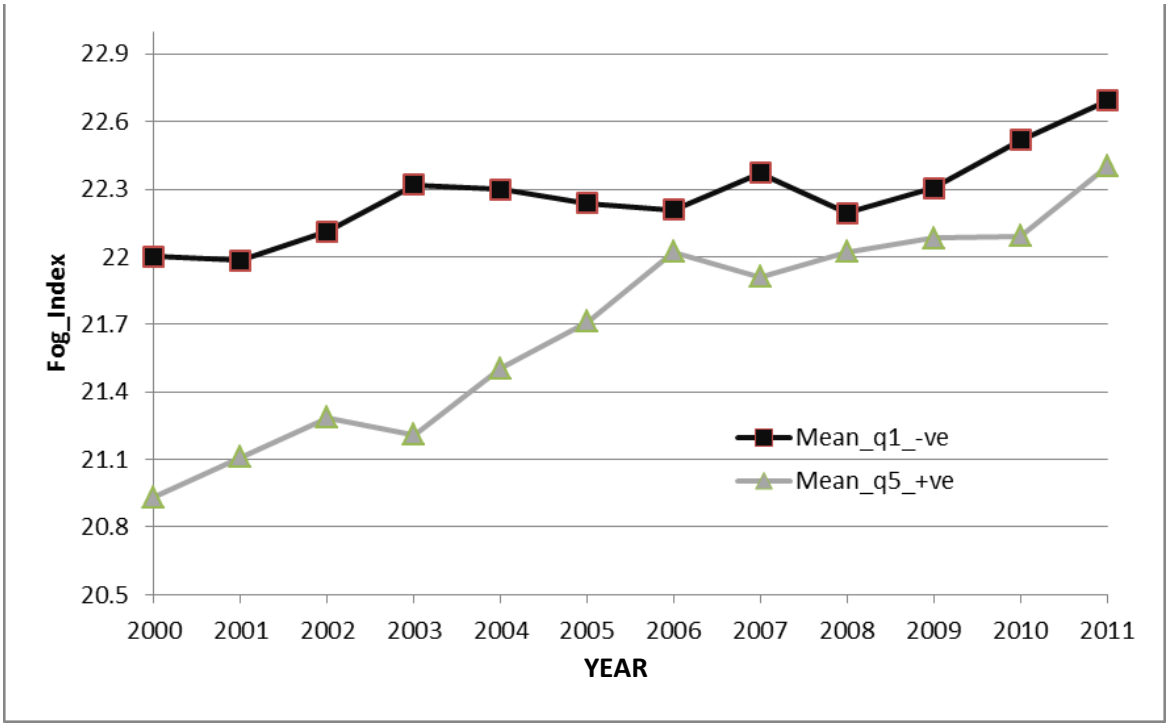


Figure shows the positive slant of annual reports from 2000 to 2011. Positive slant is the *TONE* variable as defined in Appendix A1: Variable Definitions. *YEAR* is the year of analysis.

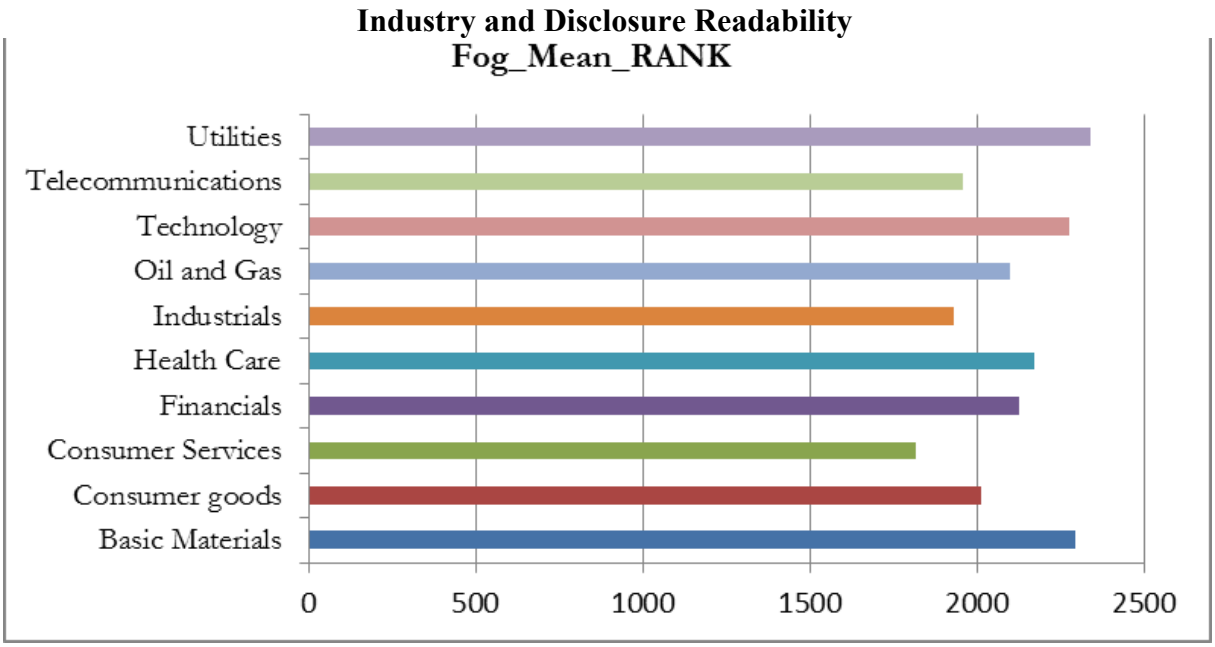
Figure 3

Reading Difficulty of Tone Quintile Groups – q1 and q5



The figure compares the Fog Indexes of the highest and lowest quintile of the TONE variable quintile groups. *YEAR* is the year of analysis. Variable descriptions are defined in Appendix A1: Variable Definitions.

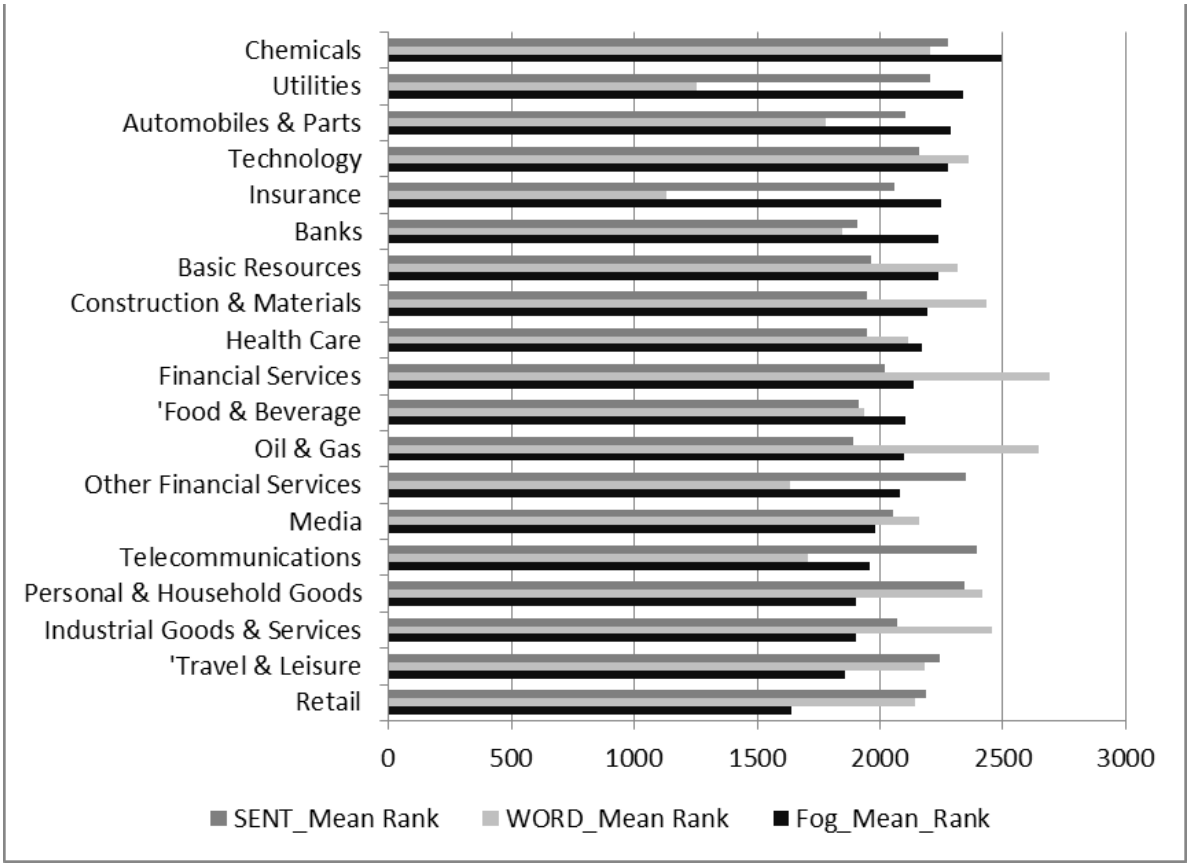
Figure 4



Disclosure Readability plot using mean rank of Kruskal–Wallis test; Mean rank difference significant at < 0.01 level; Grouping variable = ICB_Industry (IND) as defined in Table 9: Key for Industry Classification.

Figure 5

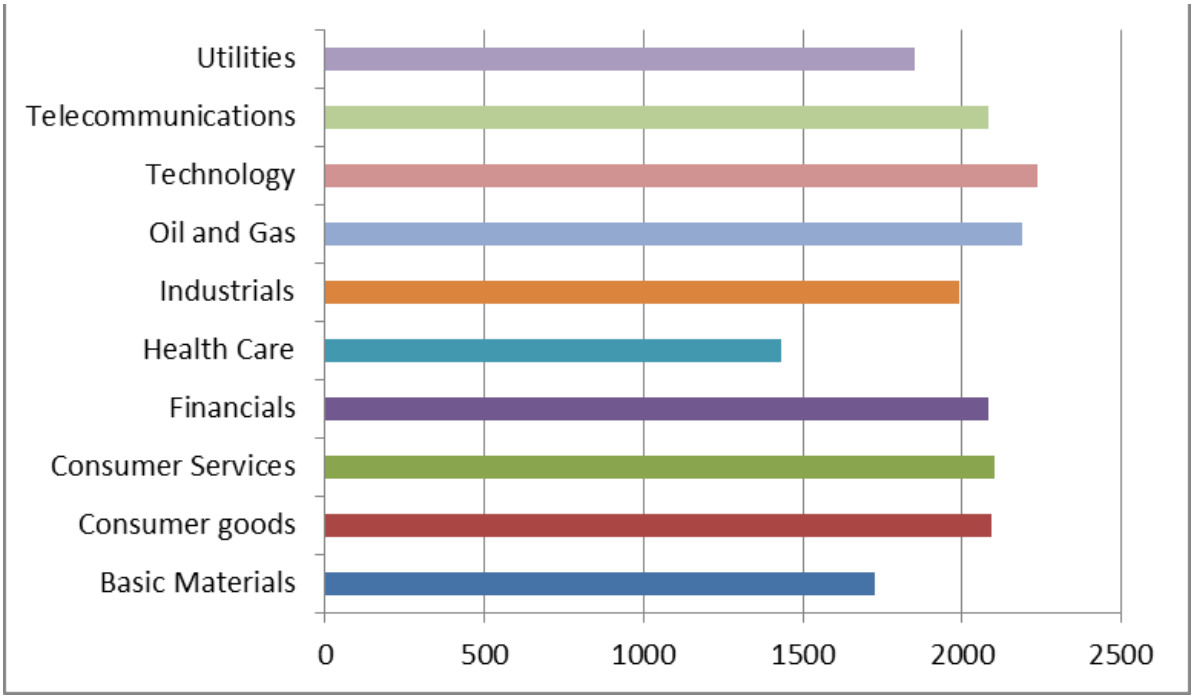
Super Sector and Disclosure Readability



SENT_Mean Rank is the mean rank of sentence complexity as measured using the variable Sent_Comp. Word_Mean Rank is the mean rank of word complexity as measured using the variable Word_Comp. Fog_Mean_Rank is the mean rank of Disclosure readability as measured using the variable Fog_Index. The figure uses the mean rank from Kruskal–Wallis test; Mean rank difference significant at < 0.01 level; Grouping variable = ICB_SUPER SECTOR as defined in Table 9: Key for Industry Classification. Variable descriptions are defined in Appendix A1: Variable Definitions.

Figure 6

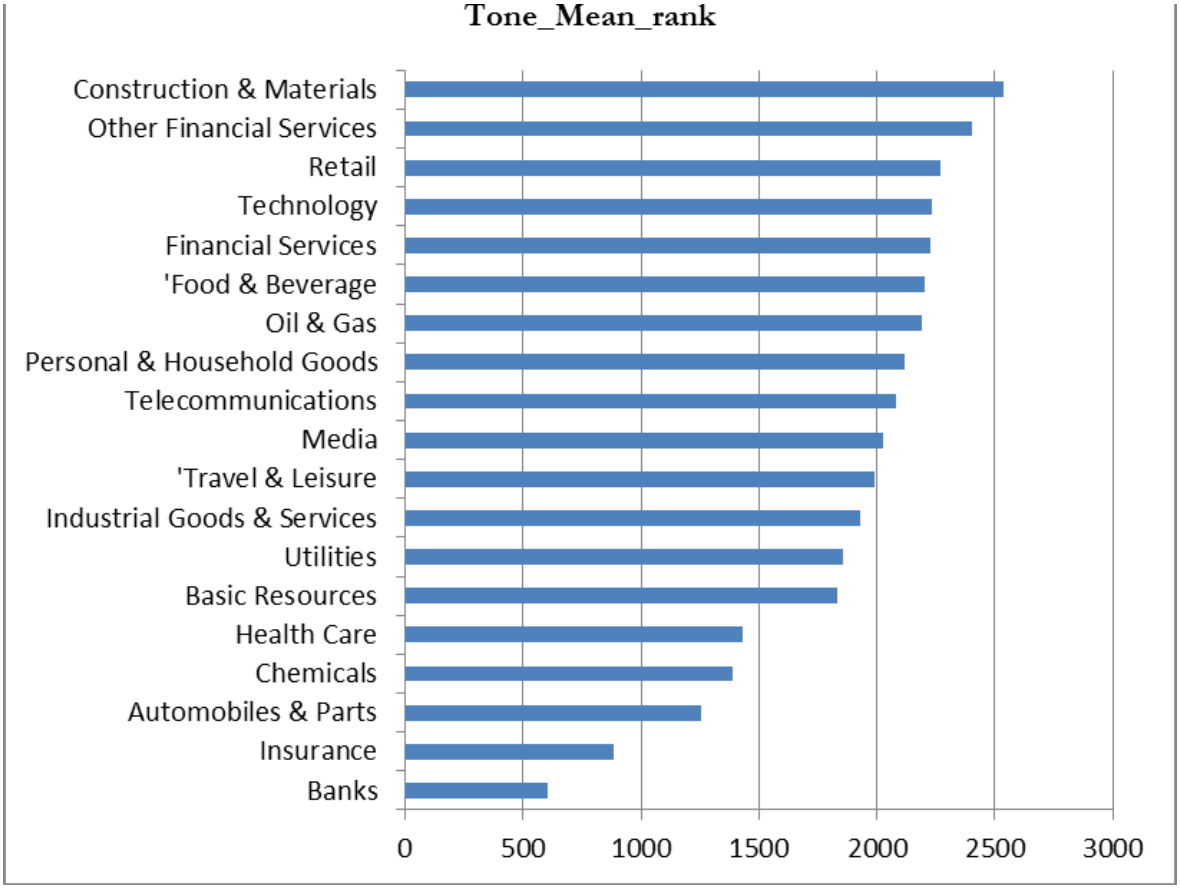
Industry and Disclosure Tone



Disclosure tone plot using mean rank from Kruskal–Wallis test; Mean rank difference significant at < 0.01 level; Grouping variable = ICB_INDUSTRY (IND) as defined in Table 9: Key for Industry Classification. It depicts the mean rank of Disclosure Tone as measured using the variable TONE. Variable descriptions are defined in Appendix A1: Variable Definitions.

Figure 7

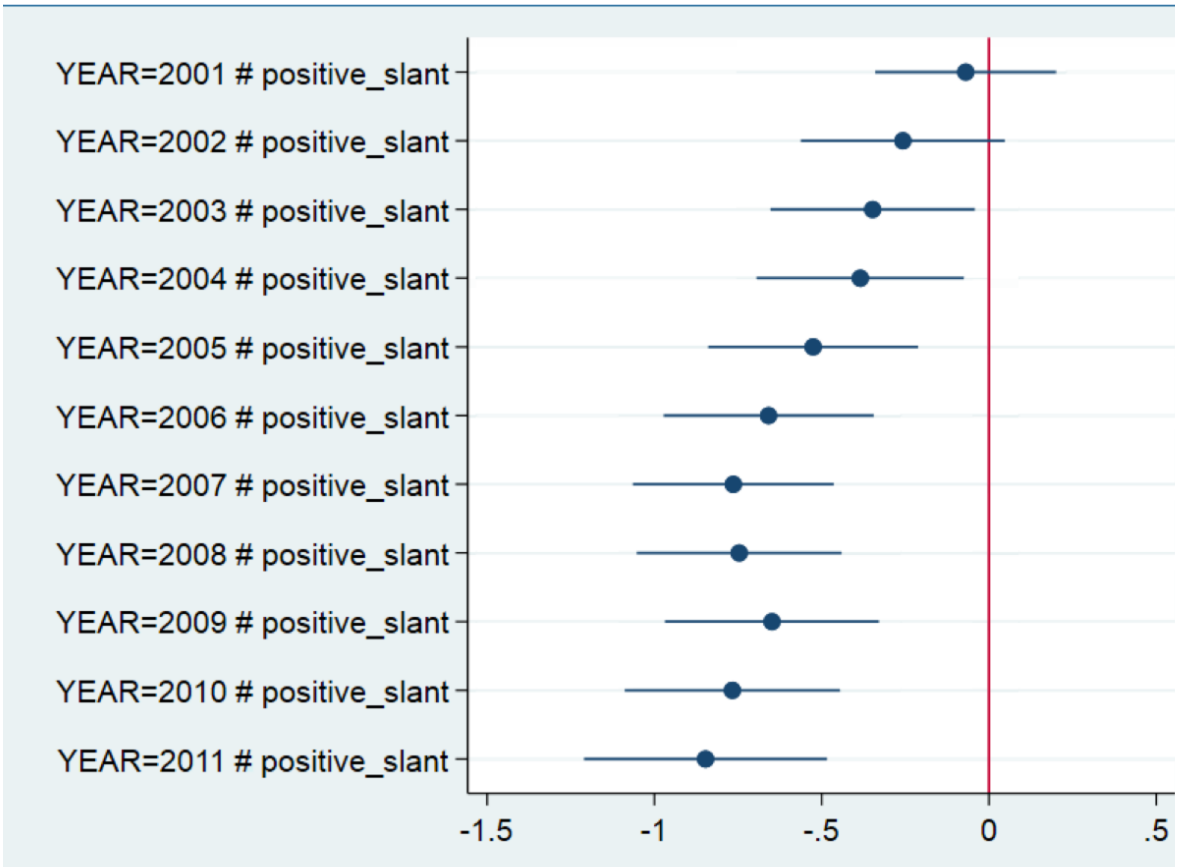
Super Sector and Disclosure Tone



Disclosure tone plot using mean rank from Kruskal–Wallis test; Mean rank difference significant at < 0.01 level; Grouping variable = ICB_SUPER SECTOR as defined in Table 9: Key for Industry Classification. Tone_Mean_rank is the mean rank of Disclosure Tone as measured using the variable TONE. Variable descriptions are defined in Appendix A1: Variable Definitions.

Figure 8

Year Analysis on the Relationship between Readability and Tone



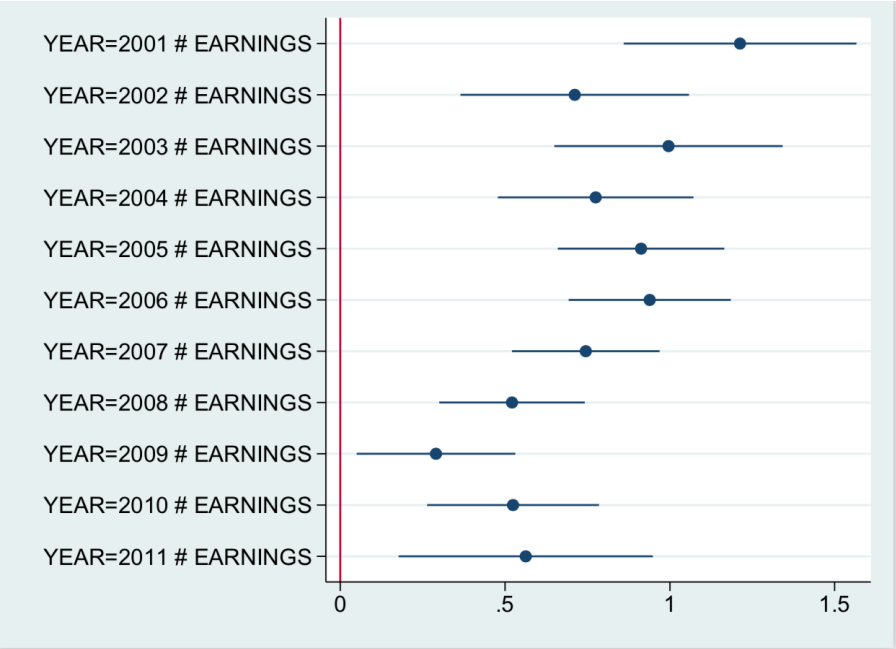
Graphical Presentation of the Year Analysis on the Relationship between Readability and Tone[12]
positive_slant is Disclosure Tone as measured using the variable TONE. Variable descriptions are defined in Appendix A1: Variable Definitions.

[12] Model includes year and firm fixed effects.

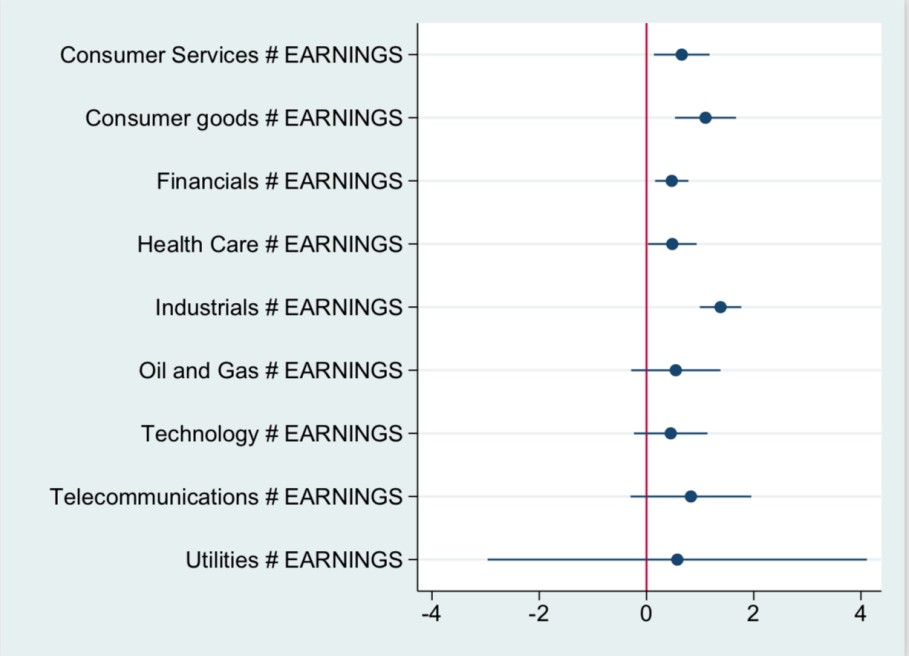
Figure 9

The Relationship between Tone and Performance

Panel A: Year Analysis



Panel B: Industry Analysis



Graphical Presentation of the Analysis on the Relationship between Tone and Performance[13]
Performance is measured using the variable EARNINGS. Tone is measured using the variable TONE. Variable descriptions are defined in Appendix A1: Variable Definitions.

[13] Model includes year and firm fixed effects.

APPENDICES

Appendix (A1): Variable Definitions

Variable	Definition
Fog_Index	Fog Index calculated as *[(Word_Complexity + Sentence_Complexity)*0.4] using Perl's Lingua En Fathom module. IMPLICATION: Firms with higher Fog Index scores have annual reports that are more difficult to read (high readability levels implies low readability of text).* *"The interpretation of the Fog Index is Score ≥ 18 = unreadable text, 14 -18 = difficult text, 12 – 14 = ideal, 10 -12 = acceptable, and 8 – 10 = childish text" (Li, 2008).*
Word_ Comp	Word complexity calculated as [(*Number of Complex Words in the Annual Report / Total Number of Words in the Same Report)*100]. Complex words being words with three or more syllables IMPLICATION: Firms with higher word complexity scores have annual reports that are more complex as measured in terms of the average syllables of words used to compose the report.*
Sent_Comp	Sentence complexity calculated as *(Total Number of Words in the Annual Report / Total Number of Sentences in the Same Report) IMPLICATION: As sentence complexity increases, the annual report includes a larger number of complex sentences.*
TONE	TONE is the positive slant calculated as *[(Total Number of Negative words in Report / Total Number of words in Same Report) * -100] IMPLICATION: As Positive slant increases, the number of negative words in an annual report decreases. This is interpreted as positive slant following Gurun and Butler (2012). Loughran and McDonald's negative word list is applied (Loughran and McDonald 2011).*
YEAR	YEAR is the year of analysis.
IND	IND is the Industry variable as defined in Table 9: Key for Industry Classification
CONTROLS	
Firm_Age	Age of the firm calculated as (Date in 2012 (precisely 02/07/2012) minus Date of Incorporation) *IMPLICATION: Older firms would have a higher firm age*.
Price_Vol	Price volatility defined as a stock's average annual price increase or decrease from a mean price for each year. For example, a stock price volatility of 20% indicates that the stock's annual high and low price has shown a historical variation of +20% to -20% from its annual average price (as defined in Thomson One Banker). *IMPLICATION: Firms with higher price volatility have share prices that are more volatile*

Variable	Definition
Bus_comp	Business complexity defined as the natural logarithm of the number of business segments. *IMPLICATION: Firms with a higher score for business complexity have a business that is more complex as indicated by the number of business segments.*
Geo_comp	Geographical complexity defined as the natural logarithm of the number of geographical segments. *IMPLICATION: Firms with a higher score for geographical complexity have structures that are more complex as indicated by a large number of geographical segments.*
Earn_Vol	Earnings volatility defined as the Standard deviation of the annual earnings (operating earnings after depreciation) of the five years prior to the year of analysis. *IMPLICATION: Firms with higher earnings volatility have earnings figures that are more likely to be volatile*
EARNINGS	Earnings calculated as a firm's operating income after depreciation scaled by the total assets of the firm. *IMPLICATION: Firms with higher earnings have better firm performance as measured by operating income*
Firm_Size	The natural logarithm *of a firm's total assets* *IMPLICATION: Larger firms have greater total assets*

Appendix (A2): Measurement Process

I measure text for readability and tone by, first, downloading the PDF files for the FTSE ALL SHARE companies. I then extract all text from the PDFs by converting the PDF files to text format.[14] This process is necessary to enable the Perl program to read the text in the files. Following text extraction, I edit the files for input into the Perl readability program. Following Miller (2010), I remove all paragraphs with more than 50 percent non-alphabetic characteristics to ensure that the analysis does not include paragraphs with only figures or tables. Moreover, I extract all text from the tables to ensure that only the text is analysed. The readability program recognises numbers and does not count them as words; therefore, keeping the numbers in the text does not change the readability results.

To parse the text for input into the readability Perl module, I write a Perl code that cleans up the text by removing text encodings and decimals. It is important to remove the text encodings to avoid misrepresenting the text. It is also important to remove all decimals because a period in the code indicates the end of a sentence. The next step is to remove paragraphs for which non-alphabetic characters make up more than 50 percent of their content using a customised Perl program. To ensure that all the annual report information is included in the test, I perform two series of readability tests as a robustness check – one that excludes the paragraphs described above and one that includes them. The reported results are based on the former because they reflect the kinds of texts read by investors.

After the file conversion and parsing process, I input each file into the Lingua EN Fathom Perl module, which reads the files and returns readability results. The Lingua EN Fathom Perl module used in accounting research (Li 2008; Miller 2010) is a Perl code written to assess the readability of English text. It takes as input a text file and calculates various text-based statistics for it. Its word-identification criteria are that a word must consist of letters and at least one vowel sound. To ensure robustness, the word-identification process does not count symbols such as '&' or abbreviations as words. It defines a sentence as a group of words and non-words followed by a period, question mark, or exclamation. Removing decimals corrects for the potential understatement of fog in numerically intensive text when using the fathom routine (Bushee et al. 2018). This study compares the result of the Perl program to manually calculated results and to results from other studies to assess the validity of the program.

[14] I use pdftotext linux software to convert PDF files to text with the option of preserving text layout.

Fog Index

The Fog Index measures readability using two sub-components: word complexity and sentence complexity metrics. It combines two measures – a measure for the complexity of the document (word complexity) and a measure for the length of the document (sentence complexity). The Lingua EN Fathom Perl module measures text complexity based on syllables per word (word complexity) and words per sentence (sentence complexity). The Fog Index formula used in the Perl module is as follows:

Fog Index = (Words per sentence + Percent complex words) *0.4.

- *Words per sentence = (num_of_words/num_of_sentences)*
- *Percent complex words = (num_of_complex words/num_of_words) * 100*
- *Complex words are words with three or more syllables*

The first part of the equation, '*Words per sentence*', measures sentence complexity, which is the number of words per sentence. Sentence complexity measures how long it will take a reader to read a sentence. The second part of the equation, '*Percent complex words*', measures word complexity, which is the proportional weight of complex words in a document. The Fog Index asserts that having more syllables per word and more words per sentence will make a document more difficult to read, all other things being equal (Li 2008).

'The interpretation of the fog index is Score ≥ 18 = unreadable text, 14 – 18 = difficult text, 12 – 14 = ideal, 10 – 12 = acceptable, and 8 – 10 = childish text' (Li, 2008).

A report scoring above 18 has unreadable text while a report with a score of between 14 and 18 is difficult to read. The higher a report's score, the more difficult it is to read (e.g. a report with a Fog Index of 21 is more difficult to read than a report with a Fog Index of 20). Difficulty levels are classified as grade levels and are interpreted as the years of education the reader needs to read the text comfortably at first reading. Every increase in the Fog Index score is representative of the need for another year of formal education to read the text with ease at the first attempt.

Tone Index

Loughran and McDonald (2011), measure the tone of a 10-K report using proportional weights, defined as negative word counts in the annual report weighted by the total number of words in the annual report. This study applies that formula using a 'bag of words' approach similar to Loughran and McDonald (2011), which requires parsing the document into vectors of words

and word counts. To obtain a word-list count, I split the annual report text into words. To this end, I employ the Lingua EN Splitter Perl Module. This module splits a document into words by defining words as a group of letters separated by a space or punctuation from another group of letters.

A customised Perl program takes as input the split word list (split words). Words featuring hyphens are considered as one word (Loughran and McDonald 2011). The code performs a loop through the Fin-Neg list to match instances of the words in the list appearing in the annual report (split words). The code produces a count of the total frequency with which the words in the Fin-Neg list appear in the annual report. The count produced, weighted by the total number of words appearing in the document, is the tone score measure used for this study. To validate the output of the program, a manual count of the Fin-Neg word instances produces the same results. The next step is to transform the tone score to a measure of positive slant by multiplying the score by -100 (Gurun and Butler 2012); this provides a score ranging between -100 and 0 (Variable *TONE*). This enables the use of a negative word list to estimate the positive slant of annual reports, as the literature on tone shows that negative words have a greater impact on readers than other kinds of words (Tetlock et al. 2008; Davis and Tama-Sweet 2012). This allows for a logical interpretation of the results; the expectation is that higher tone scores are more positive.