

Readability, Profitability, and Discretionary MD&A Text

Kurt H. Gee

Graduate School of Business
Stanford University

December 2017

I thank my dissertation committee, Mary E. Barth (Chair), Beth Blankespoor, and Ron Kasznik, for invaluable guidance; Dan Jurafsky (Stanford University Department of Linguistics) for consultation on my linguistics approach; and Anne Beyer, Dan Boneh, Yi-Chun Chen (AAA Discussant), Jung Ho Choi, Ted Christensen, Lisa De Simone, Mark Lang, Marilyn Lattin, Charles Lee, Becky Lester, Maureen McNichols, Joe Piotroski, Navdeep Sahni, participants at the 2017 AAA Annual Meeting and the 2017 AAA J. Michael Cook Doctoral Consortium, and colleagues at Stanford GSB for helpful feedback. I gratefully acknowledge financial support from Stanford GSB and the Deloitte Foundation Doctoral Fellowship.

Readability, Profitability, and Discretionary MD&A Text

Abstract

This study examines the relation between profitability and disclosure readability. Prior research finds that less profitable firms obfuscate information by decreasing readability. However, this research does not distinguish text that is non-discretionary, i.e., required disclosure or standard language given the economic environment, from text that is discretionary, i.e., voluntary disclosure or required disclosure written in a non-standard way. I measure discretion in Management Discussion & Analysis (MD&A) sentences based on their commonness relative to sentences from firms with similar profitability. I assume that sentences that are less (more) common among firms with similar profitability contain more (less) discretion. I find that less profitable firms have less readable non-discretionary text but more readable discretionary text, and this discretionary text is even more readable when firms have stronger incentives to garner a reputation for clear disclosure. My results imply that less profitable firms face an economic environment that is complicated to explain, yet exercise discretion by increasing readability.

1. Introduction

This study examines the relation between profitability and disclosure readability. Academics and regulators suggest that investors are more informed when disclosure is more readable. Prior research compares disclosure readability between high- and low-profitability firms and finds that low-profitability firms have less readable disclosure text, consistent with firms hiding bad news. However, disclosure text reflects both non-discretionary text, i.e., required disclosure or standard language given the economic environment, and discretionary text, i.e., voluntary disclosure or required disclosure written in a non-standard way. Low-profitability firms could have less readable non-discretionary text (Bloomfield 2008), which makes it difficult to identify firms' disclosure readability decisions. Dyer et al. (2016, p. 272) note: "A primary challenge for the literature is in convincingly separating disclosure choices from the effects of underlying economics and other factors." I measure discretion in Management Discussion & Analysis (MD&A) text and test how the readability of discretionary text differs between high- and low-profitability firms.

The relation between profitability and disclosure readability is unclear ex ante. When increasing readability and reducing information acquisition costs, firms face a tradeoff between reducing information risk and increasing the likelihood that investors acquire negative information about the firm. Low-profitability firms might decrease readability and hide information from investors, i.e., obfuscate, if they can prevent investors from accessing and understanding negative information. Alternatively, low-profitability firms might increase readability if they are more concerned with reducing information risk than with hiding negative information. Low-profitability firms could increase readability if doing so helps them garner a reputation for providing clear disclosure. If low-profitability firms obfuscate (provide clear

disclosure), their disclosures should be less (more) readable than those of high-profitability firms.

Li (2008) finds a positive relation between MD&A readability and profitability, i.e. firms with lower earnings have less readable MD&As, and interprets this as evidence that low-profitability firms obfuscate. Li's (2008) interpretation implies that readability is a firm choice. However, Bloomfield (2008) posits that low-profitability firms could have less readable non-discretionary text than high-profitability firms, which implies that readability is at least partially outside of firms' control. If Li's (2008) inferences are attributable to differences in the readability of non-discretionary text, this suggests that instead of obfuscating, low-profitability firms simply face an economic environment that is more complicated to explain and that leads to less readable MD&As.

I measure profitability in two ways, following prior research: (1) operating earnings over the prior twelve months scaled by total assets, and (2) an indicator for a quarterly profit or loss. Close inspection of MD&A text from high- and low-profitability firms reveals differences that are difficult to attribute to firm choice. For example, low-profitability firms discuss accumulated deficits and their ability to continue as a going concern more frequently while high-profitability firms discuss effective tax rates more frequently. These text differences, which are presumably non-discretionary, can influence readability comparisons between high- and low-profitability firms. Because the text differences are related to profitability itself, they are also difficult to address with controls.

The primary contribution of this study is to measure discretion in MD&A text and compare the readability of discretionary text between high- and low-profitability firms. I use the Gunning fog index (Li 2008) to measure readability. To measure the extent of discretion in text, I

use the empirical distribution of MD&A sentences to estimate the level of discretion in each sentence. Specifically, I assume that sentences that are common among firms with similar profitability reflect primarily required disclosure or standard language given firms' profitability, and therefore contain little discretion. Conversely, I assume that sentences that are uncommon reflect more voluntary disclosure or required disclosure written in a non-standard way and therefore contain more discretion. I calculate the commonness of each MD&A sentence relative to sentences from firms with similar profitability. Next, I eliminate sentences that are repeated by the same firm over time, i.e., boilerplate, to mitigate the influence of text related to unusual economic events or firm-specific disclosure that does not change from period to period. I rank the remaining sentences into quintiles within each MD&A based on their commonness and use the quintile rank as my sentence-level measure of discretion. I assume that less common (more common) sentences within an MD&A contain more (less) discretion.

To assess construct validity for my measure of discretion, I test for three properties. First, text with more discretion has more forward-looking statements. Forward-looking statements are generally voluntary which implies that text with more discretion contains more forward-looking statements. Consistent with this, I find that text I identify as more discretionary has significantly more forward-looking statements than text I identify as less discretionary. Second, text with more discretion has more positive tone. To the extent the average firm exercises discretion by providing more positive information or using more positive tone, text with more discretion should have more positive tone. Consistent with this, I find that text I identify as more discretionary has significantly more positive tone than text I identify as less discretionary. Third, text with less discretion is less readable over time. Dyer et al. (2017a) find that required disclosure in Form 10-K has become less readable over time, which suggests that less-

discretionary text exhibits decreasing readability over time. Consistent with this, I find that text I identify as less discretionary exhibits decreasing readability over time. Together, these three findings suggest that my measure is positively correlated with the level of discretion in text.

To address my research question, I test how the relation between readability and profitability differs as text contains more discretion. I find that low-profitability firms have less readable non-discretionary text, which suggests that low-profitability firms have less readable MD&As because of text reflecting required disclosure or standard language given the level of profitability. However, I find that discretionary text is more readable for low-profitability firms than for high-profitability firms. I also find that discretionary text is more readable when a firm has lower profitability relative to its time series. These findings are consistent with low-profitability firms providing more readable discretionary disclosure, and extend Li (2008) by demonstrating that low-profitability firms have less readable disclosures overall because of non-discretionary text.

Next, I examine differences in disclosure content to understand what low-profitability firms disclose that is more readable. First, I find that discretionary text has less positive tone for low-profitability firms than for high-profitability firms. This result is inconsistent with low-profitability firms distracting investors with more readable positive tone. Second, I find that high- and low-profitability firms differ in their use of detailed technical disclosure, i.e., disclosure that is less readable but potentially informative (Bushee et al. 2017). I use the presence of attributions, forward-looking statements, and risk statements as proxies for technical disclosure. I find that discretionary text from low-profitability firms has fewer attributions but more risk statements than such text from high-profitability firms. These results are inconsistent with low-profitability firms making more readable excuses. Collectively, the results on

differences in disclosure content are consistent with low-profitability firms discretionarily disclosing more negative, risky information in a readable way.

I also explore cross-sectional variation in discretionary readability across two dimensions. First, I test whether low-profitability firms obfuscate more when profitability is persistent. Firms might have stronger incentives to hide information about persistent low profitability. Alternatively, providing readable disclosure about persistent low profitability may be most helpful in garnering a reputation for clear disclosure. I find that when profitability is persistent, low-profitability firms have even more readable discretionary text than high-profitability firms, which is inconsistent with obfuscation. Second, I test whether discretionary readability is higher for low-profitability firms when firms are financially constrained and presumably need reputational capital to help secure future financing. I find that when firms are financially constrained, low-profitability firms have even more readable discretionary text than high-profitability firms, consistent with firms increasing readability to garner a reputation for clear disclosure.

My paper contributes to the textual analysis literature, which discusses how the inability to identify discretion and non-discretion potentially limits the inferences that can be drawn from empirical analyses (e.g., Bloomfield 2008, Berger 2011, Dyer et al. 2016, Dyer et al. 2017b). I use within-MD&A variation in the level of discretion to analyze the properties of the specific text that reflects firms' disclosure decisions. Distinguishing less-discretionary and more-discretionary text yields the opposite inference from prior research which finds that low-profitability firms obfuscate MD&A text. Instead, I find that low-profitability firms have less readable text for non-discretionary reasons yet they exercise discretion to provide more readable disclosures. This evidence contributes to a sparse area of the literature, as noted by Dyer et al.

(2017b), in which firms exercise discretion by making information more accessible to investors. Although obfuscation may occur in other settings, such as conference calls (Bushee et al. 2017), my results are inconsistent with low-profitability firms obfuscating MD&A text. My approach to measuring the level of discretion in text can help future research provide evidence of firms' disclosure decisions.

2. Hypothesis development and prior literature

2.1. Hypothesis development

Economic theory predicts that less accessible information is reflected incompletely in equilibrium stock prices (Grossman and Stiglitz 1980). Consistent with more readable disclosure being more accessible, prior research finds that more readable Forms 10-K and 10-Q are associated with higher trading volume (Miller 2010), greater stock holdings and more profitable trades for retail investors (Lawrence 2013), and lower post-earnings-announcement drift (Lee 2012). Rennekamp (2012a) finds that subjects in an experimental laboratory revise valuation assessments more strongly when new information is more readable. When information is negative (positive), subjects revise assessments downwards (upwards) more when the information is easier to read.

When determining how to present information to investors, firms trade off maximizing expectations of future cash flows, which increases stock price, and minimizing information risk, i.e., investors' concerns that they are not fully informed, which increases the cost of capital and decreases stock price. When firms have negative information, making this information less (more) accessible yields higher (lower) expectations of future cash flows, but also higher (lower) information risk. The relation between readability and profitability depends on how firms make this tradeoff.

Firms with lower profitability could make disclosures less readable, i.e., obfuscate (Bloomfield 2002), if the benefits of hiding negative information outweigh any increase in information risk. Although firms probably cannot obfuscate current profitability, less readable disclosures could make it difficult for investors to gain context regarding whether economic conditions will persist. Investors may not be able to unravel obfuscation because distinguishing firms' disclosure decisions from the effects of other factors is difficult. Overall, the obfuscation hypothesis implies that less profitable firms have lower readability, i.e., readability is positively associated with profitability.

Alternatively, firms with lower profitability could make disclosures more readable if they prioritize reducing information risk over hiding negative information. CFOs claim that garnering a reputation for clear disclosure is a primary concern when making disclosure decisions (Graham et al. 2005), and less profitable firms may have more to gain from developing a reputation for clear disclosure because they often need additional financial capital. Less profitable firms also have fewer resources to weather litigation, and could make disclosures more readable to deter potential lawsuits. Overall, the clear disclosure hypothesis implies that less profitable firms have higher readability, i.e., readability is negatively associated with profitability.

2.2. The empirical relation between readability and profitability

Li (2008) examines the relation between readability, as measured using the Gunning fog index, and firm profitability. Li (2008) finds that firms with higher (lower) earnings have more (less) readable Form 10-Ks and MD&As and interprets this as evidence that firms obfuscate negative information to make it more difficult for investors to process. Li's (2008) interpretation implies that readability is a firm choice and that less profitable firms choose lower readability.

Bloomfield (2008) posits an alternative explanation to that in Li (2008), which is that firms with lower profitability may have less readable text because negative information is inherently more complicated to explain, i.e., readability is lower for low-profitability firms for non-discretionary reasons.¹ This concern is echoed by Berger (2011, p. 210): “[T]he Fog index and report length are measures of total disclosure that do not separate discretionary and non-discretionary (i.e., mandated) components. Thus the results may merely reflect greater difficulty in explaining bad news or transitory income components...”. Rennekamp (2012b) finds that subjects in an experimental laboratory use less readable text when describing lower performance even when their only incentive is to report performance as accurately as possible. This suggests that text from lower-profitability firms may be inherently less readable.

Dyer et al. (2017a) and Guay et al. (2016) provide empirical support for readability being partially non-discretionary. Dyer et al. (2017a) finds that Form 10-Ks have become less readable over time, and that the increasingly less readable text relates to required disclosure topics. The results in Dyer et al. (2017a) suggest that readability differences across firms may be attributable to differences in required disclosure. Guay et al. (2016) finds that firms with less readable Form 10-Ks issue more management guidance. This finding is consistent with firms trying to clarify less readable disclosure, which they presumably would not do if they were obfuscating. Guay et al. (2016) finds that less profitable firms are less likely to issue management guidance following less readable Form 10-Ks. Although this result is consistent with low-profitability firms obfuscating, an alternative explanation is that low-profitability firms have less readable non-discretionary text and compensate with more readable discretionary text in the Form 10-K itself, rather than relying on management guidance to clarify less readable disclosure.

¹ Li (2008) acknowledges this possibility, but concludes that the evidence is consistent with obfuscation.

As Berger (2011) points out, the problem is that discretionary and non-discretionary text confounds comparisons of readability between high- and low-profitability firms. By discretionary text, I mean text that reflects voluntary disclosure or required disclosure written in a non-standard way. Non-discretionary text reflects required disclosure or standard language given firms' profitability. My study measures the level of discretion in text and compares the readability of text with less and more discretion between high- and low-profitability firms.

Bushee et al. (2017) recognizes that lower readability can reflect two disclosure choices: (1) informative technical disclosure, i.e., information that is less readable because it is detailed and meant to inform, and (2) obfuscation, i.e., information that is less readable because the firm tries to hide it. To separate these disclosure choices, Bushee et al. (2017) decomposes conference call readability into an informative component, which is predicted by the readability of analysts' conference call questions and a set of firm characteristics, and an obfuscation component, which is a residual. Bushee et al. (2017) finds that loss firms have lower predicted readability and lower residual readability. The former suggests that loss firms provide more technical disclosure in conference calls, while the latter suggests that they obfuscate.

My study differs from Bushee et al. (2017) in three ways. First, my study attempts to identify discretionary text, which reflects firms' disclosure decisions, whereas Bushee et al. (2017) attempts to separate two disclosure decisions, i.e., obfuscation and technical disclosure. Motivated by Bushee et al. (2017), I examine differences in proxies for technical disclosure between high- and low-profitability firms (see section 6.1). Second, I examine the MD&A while Bushee et al. (2017) examines conference calls. The MD&A may face more scrutiny by regulators because the U.S. Securities and Exchange Commission (SEC) reviews the MD&A as part of its filing review process. Firms could avoid obfuscation in the MD&A but obfuscate in

conference calls. Third, I specifically address non-discretionary differences in readability that are related to profitability, whereas profitability is not included in the model of expected readability in Bushee et al. (2017).² To the extent that low-profitability firms have less readable non-discretionary text that is not reflected in the explanatory variables used in Bushee et al. (2017), residual readability would be lower, which would be interpreted as obfuscation.

3. Research design

Following Li (2008), I examine the MD&A, which companies filing Forms 10-K or 10-Q with the SEC are required to provide. Although much of the MD&A is mandated by Regulation S-K Item 303, the narrative form of the MD&A likely affords firms more discretion than other sections in Forms 10-K or 10-Q. I use MD&As from both Forms 10-K and 10-Q to maximize linguistic variation and improve my ability to identify variation in discretion.

To compare the readability of discretionary text between high- and low-profitability firms, I require proxies for readability, profitability, and the level of discretion in text. Following prior research, I measure readability using the Gunning fog index.³ I use two proxies for profitability. The first is *Earn*, which is operating earnings defined by Compustat over the prior twelve months and scaled by total assets (Li 2008). *Earn* is ranked into quintiles over the sample and scaled between 0, the lowest profitability, and 1, the highest profitability. The second measure of profitability is *Profit*, which is an indicator equal to 1 if quarterly net income is non-

² Bushee et al. (2017) cannot both (a) use profitability as an explanatory variable to estimate predicted readability, and (b) test for differences in residual readability based on profitability. This is a consequence of the discretionary accruals framework (e.g., Jones 1991) employed by Bushee et al. (2017), in which residuals from the first stage are orthogonal to first stage explanatory variables.

³ Although the Gunning fog index has limitations as a readability measure (see Loughran and McDonald 2016, Bonsall et al. 2017), I use it for consistency with the bulk of prior research. Inferences are unchanged when using Flesch Reading Ease (Flesch 1948), an alternative readability measure used in computational linguistics research (e.g., Markowitz and Hancock 2015).

negative and to 0 if it is negative (Li 2008, Guay et al. 2016, Bushee et al. 2017). I next describe my proxy for the level of discretion in text.

3.1. Measuring discretion

Consider a hypothetical experiment in which high- and low-profitability firms are given a pre-written MD&A that reflects all required information, written in standard language, for their profitability condition, i.e., all text is non-discretionary. Each firm is permitted to adjust the pre-written MD&A before the final version is presented to investors. Firms can add or modify sentences but must retain all required disclosures. After firms make adjustments, the researcher could identify the level of discretion in text based on the extent to which firms adjust the non-discretionary text. Sentences with more (fewer) adjustments would reflect more (less) discretion.

I implement a similar concept and use the uncommonness of a sentence as a proxy for its level of discretion, i.e., sentences that are less (more) common given firms' profitability are assumed to contain more (less) discretion. This approach mimics that of a naïve reader of MD&As who uses the empirical distribution of sentences to estimate the level of discretion in each sentence. After grouping MD&As based on profitability, e.g., collecting MD&As from all loss firms, I presume that common sentences are primarily non-discretionary and thus reflect required disclosure or standard language given firms' profitability. Less common sentences likely contain more discretion as firms either add voluntary disclosure or modify the standard language of required disclosures.

The key assumption of my approach is that uncommon sentences reflect more discretion than common sentences. For example, when two loss firms have similar sentences, i.e., use approximately the same words in the same order, I presume that this reflects non-discretionary text rather than the two firms independently exercising discretion by writing the same way. Even

if both firms decide to write less readable sentences, it is unlikely that, of all the possible ways to do so, both would choose to use the same words in the same order. To the extent that some discretionary text arises from firms using the same words in the same order, this text will be misclassified and it will be more difficult to distinguish less-discretionary and more-discretionary text. Using uncommonness to measure discretion does not rely on subjective hand-classification and is therefore objective, replicable, and accommodates large samples of data.⁴

As a case study, consider two example sentences from the Q4 2002 MD&A of Swissray International Inc., a medical imaging company:⁵ (S1) “Upon adoption of FASB No 142, the Company recorded a one-time, non-cash charge of approximately \$1,120,000 to write off the carrying value of its goodwill.” and (S2) “Management of the Company hopes that its effort to attract new capital to strengthen its liquidity will have a positive result in the foreseeable future.” S2 likely contains more discretion than S1 because S2 is voluntary disclosure, while S1 reveals a material economic event about which disclosure is required. For Q4 of 2002, Swissray’s profitability is in the lowest 20% of firms, i.e., $Earn = 0$. Among the MD&As from all such firms, S1 is more common than S2, which reflects the intuition that more common sentences contain less discretion.

I construct *Discretion*, my proxy for the level of discretion in an MD&A sentence, in three steps. First, I measure sentence commonness. Second, I remove sentences that are repeated across time by the same firm, i.e., boilerplate (Brown and Tucker 2011). Third, I rank non-boilerplate sentences within each MD&A into quintiles based on their commonness. I discuss validation tests of *Discretion* in section 4.3.

⁴ My use of commonness is similar to intrinsic plagiarism detection in computational linguistics (e.g., Stamatatos 2009, 2011). In this literature, researchers use the empirical distribution of words or characters to identify portions of a document that differ from the rest and are more likely to be written by a different author.

⁵ The entire MD&A is presented in Appendix A, section A.2. The example sentences are #127 and #24, respectively.

3.1.1. Sentence commonness

Sentence commonness is a function of $P(S)$, the probability of sentence S occurring in the set of all MD&As from firms with similar profitability.⁶ I modify sentences to increase comparability and make it easier to assess commonness. For example, the sentence:

Research and development expenses were \$2,549,384 or 13.21% of net sales for the year ended June 30, 2002 compared to \$2,305,165 or 10.4% of net sales for the year ended June 30, 2001.

becomes:

<s> <s> research and development expenses were \$<d> or <d>% of net sales for the year ended <date> compared to \$<d> or <d>% of net sales for the year ended <date> </s>

The markers <s> <s> and </s> denote the beginning and end of the sentence; <d> is a marker for a number, so that <d>% denotes percentages and \$<d> denotes dollar values; and <date> denotes a date. I require all words in MD&A sentences to be contained in the Loughran McDonald (2011) Master Dictionary (hereafter, LM Dictionary).⁷ Limiting words to this dictionary excludes proper nouns such as places, company names, and product names. These types of uncommon, firm-specific words could overstate the uncommonness of a sentence relative to its level of discretion.

Next, I divide each modified sentence into overlapping three-word phrases, i.e., trigrams.

Trigrams capture both word selection and word order. The trigram {<s> <s> research} denotes

⁶ My measure of sentence commonness is based on a computational linguistics measure called perplexity (Brown 1992, Jurafsky and Martin 2009), which is an estimate of the unusualness of a sequence of words (phrases) given a probability distribution over words (phrases) from a collection of text.

⁷ I supplement the LM Dictionary with possessive forms, e.g., “companys”, contractions, e.g., “didnt”, and a set of custom category markers. These markers include: numbers, e.g. 27 becomes <d>, \$1.2 billion becomes \$<d>, and 12.5 percent becomes <d>%; SEC forms, e.g., 10-K becomes <sec_form>; accounting standards, e.g., FAS 123R becomes <acct_stndrd>; and dates, e.g., December 31, 2002 becomes <date>. See Appendix A, section A.1, for more detail about how I identify and modify MD&A sentences.

starting a sentence with the word “research”. The next two trigrams in the sentence are: {<s> research and}, {research and development}. The last trigram, {ended <date> </s>}, denotes how the sentence ends. I estimate $P(S)$ as the probability of observing the sequence of trigrams. For the example sentence:

$$P(S) = P(\text{research} | \langle s \rangle \langle s \rangle) * P(\text{and} | \langle s \rangle \text{research}) * \dots \\ * P(\langle /s \rangle | \text{ended} \langle \text{date} \rangle)$$

where $(w_i | w_{i-2} w_{i-1}) = \frac{\text{Frequency}(\{w_{i-2} w_{i-1} w_i\})}{\text{Frequency}(\{w_{i-2} w_{i-1}\})}$.

$P(\text{research} | \langle s \rangle \langle s \rangle) = \frac{\text{Frequency}(\{\langle s \rangle \langle s \rangle \text{research}\})}{\text{Frequency}(\{\langle s \rangle \langle s \rangle\})}$, i.e., the number of times {<s> <s> research} occurs divided by the number of times the bigram {<s> <s>} occurs, i.e., the number of sentences. $P(S)$ is therefore the probability of starting a sentence with “research” times the probability of continuing such a sentence with the word “and”, and so on, ultimately ending the sentence with “ended <date>”.⁸

Following the computational linguistics literature (e.g., Brown 1992, Jurafsky and Martin 2009), $\text{Commonness}(S) = \sqrt[m]{P(S)}$, which is the geometric mean of the probabilities of the m trigrams in the sentence.⁹ Sentences comprised of more uncommon (common) trigrams have lower (higher) *Commonness*, and likely have more (less) discretion.

3.1.2. Boilerplate sentences

⁸ Estimating the probability of the sentence as the probability of the sequence of trigrams assumes trigrams occur independently of one another. This is a standard assumption in the computational linguistics literature (e.g., Jurafsky and Martin 2009) and is invoked to limit the number of probabilities to estimate, which would otherwise grow exponentially with the length of the sentence. The independence assumption is more likely to be satisfied for longer phrases, which is an advantage of dividing sentences into three-word phrases, i.e., trigrams, over one-word phrases, i.e., unigrams.

⁹ Perplexity, from computational linguistics, is equal to $\sqrt[m]{\frac{1}{P(S)}}$.

A potential concern with using uncommonness to measure discretion is that some uncommon sentences may be non-discretionary, e.g., discussing the terms of an acquisition or other unusual economic events. Sentences discussing such events are often repeated from one MD&A to the next. In addition, some text is repeated across time by the same firm regardless of profitability, e.g., introductory sentences to the MD&A. This firm-specific text likely adds noise because the text does not vary with profitability. To mitigate the influence of both types of sentences, I identify and remove sentences that are repeated across time for the same firm, i.e., boilerplate. For firm i in quarter t , I classify a sentence as boilerplate if it closely resembles a sentence in the previous or subsequent MD&A.¹⁰ I use Levenshtein Distance (Levenshtein 1966, Crowley 2016) to compare sentences across MD&As for the same firm.¹¹ Given a source sentence with n words, and a target sentence, Levenshtein Distance measures the minimum number of operations, i.e., word insertion, deletion, or substitution, needed to convert the source sentence into the target sentence.¹²

For each sentence in each MD&A, I calculate the minimum Levenshtein Distance over all possible target sentences and denote this as δ . If a sentence has $\delta = 0$, it appears verbatim in the previous or subsequent MD&A. To allow for a small number of word changes in formulaic disclosures, e.g., updating “earnings” for “losses” in a sentence, I classify a sentence as boilerplate if less than 20 percent of the sentence requires editing, i.e., $\left(\frac{\delta}{n} < 0.2\right)$. Untabulated

¹⁰ To ensure comparability, if an MD&A sentence comes from a Form 10-Q (10-K) in quarter t , I compare the sentence to MD&A sentences in quarters $t-1$ and $t+1$ ($t-4$ and $t+4$).

¹¹ Levenshtein Distance reflects differences in word order within a sentence, unlike other measures of text similarity, e.g., cosine similarity (Brown and Tucker 2011). Levenshtein Distance also allows me to identify specific boilerplate sentences rather than merely summarizing an entire text (e.g., Cazier and Pfeiffer Jr. 2017).

¹² For example sentence #1, “<s> <s> earnings this period were \$<d> </s>”, and example sentence #2, “<s> <s> losses this period were \$<d> </s>”, the minimum number of operations is 1. The word “earnings” is substituted for the word “losses”.

statistics reveal that I classify 76% of the MD&A sentences in my sample as boilerplate, and 59% of MD&A sentences appear verbatim in the previous or subsequent MD&A.¹³ Removing boilerplate sentences results in MD&A text that changes from period to period, either for non-discretionary or discretionary reasons.

3.1.3. Ranking sentences

After removing boilerplate sentences, I rank the remaining sentences within each MD&A into quintiles based on their commonness. *Discretion* is the quintile rank, scaled to range from 0, i.e., the most common sentences, to 1, i.e., the least common sentences. I assume that quintiles of sentences with higher *Discretion* contain more discretion than quintiles with lower *Discretion*.

Ranking within the MD&A controls for differences in the overall level of discretion in the MD&A across firms and across time. This mitigates the potential concern that uncommon sentences, which are assumed to be more discretionary, come from a set of unique firms or from one unique time period, or, more generally, that *Discretion* is correlated with firm or time characteristics that are difficult to identify. As a result of ranking within the MD&A, *Discretion* is uncorrelated with any observable or unobservable firm or time characteristics. For example, *Discretion* is uncorrelated with business complexity because business complexity varies at the firm-quarter level, and for any level of business complexity there are the same number of sentences with *Discretion* = 0 as there are with *Discretion* = 1. Although ranking within the MD&A prevents me from comparing the level of discretion across firms, I can compare text properties, including readability, between less- and more-discretionary text and across firms.

3.2. *The relation between profitability and the readability of discretionary text*

¹³ Untabulated statistics reveal that the proportion of boilerplate text is approximately the same across different levels of *Earn* or *Profit*, as is the average value of *Commonness*. The latter result is inconsistent with less profitable firms have more varied sentences in general.

I estimate the following equation to compare the readability of text with more discretion between firms with higher and lower profitability:

$$\begin{aligned}
 \text{Readability}_{itu} = & \beta_1 \text{Profitability}_{it} + \beta_2 \text{Profitability}_{it} * \text{Discretion}_{itu} + \\
 & \Sigma \text{Industry}_{it} * \text{Discretion}_{itu} + \Delta \text{Year}_t * \text{Discretion}_{itu} + \epsilon
 \end{aligned}
 \tag{1}$$

Readability is $= -1 * \text{Fog}$, the Gunning fog index, so that higher values of *Readability* denote less foggy, i.e., more readable, text. *Readability* is calculated using all sentences in quintile u .¹⁴

Discretion is the level of discretion for quintile of sentences u and ranges from 0 to 1 within each MD&A. *Profitability* is either *Earn* or *Profit*. I estimate Equation (1) and all equations that follow by pooling firms cross-sectionally and over time, and I base test statistics on standard errors clustered by firm and calendar quarter (Gow et al. 2010). Continuous variables in Equation (1) and all equations that follow, e.g., *Readability*, are winsorized at the 1st and 99th percentile of their distributions.

Figure 1 illustrates the data structure for a hypothetical firm over two quarters. Each firm-quarter is represented by exactly five observations, one for each quintile u of sentences. *Discretion (Readability)* is the level of discretion (readability) of the sentences in quintile u . *Earn* and *Profit* have the same values for all observations from each firm-quarter, but *Discretion* and *Readability* differ across the quintiles of sentences. The distribution of *Discretion* is identical for each firm quarter.

The main effect of *Profitability* in Equation (1) is a control for differences in the readability of non-discretionary text between high- and low-profitability firms. The coefficient

¹⁴ I do not treat my custom markers, e.g., “<date>”, as words for the purposes of calculating *Readability* because it is not clear how many words they represent. A reader might treat December 31, 2009 as three words or as one word, i.e., <date>.

β_1 reflects the relation between *Readability* and *Profitability* when *Discretion* is zero, which corresponds to the least discretionary sentences in each MD&A. If $\beta_1 > 0$ ($\beta_1 < 0$), this implies that, for non-discretionary reasons, low-profitability firms have less readable (more readable) text than high-profitability firms. If $\beta_1 = 0$, this implies that there are no non-discretionary differences in readability between high- and low-profitability firms.

The coefficient of interest in Equation (1) is β_2 , which reflects how the relation between *Readability* and *Profitability* differs as text contains more discretion. The coefficient β_2 is my estimate of the difference in the readability of discretionary text between high- and low-profitability firms. I base my inferences on β_2 rather than the sum of $\beta_1 + \beta_2$, which would reflect the relation between *Readability* and *Profitability* when *Discretion* = 1. When *Discretion* = 1, sentences contain both discretionary and non-discretionary text and comparisons of readability between high- and low-profitability firms will reflect both discretionary and non-discretionary differences. Therefore, comparing readability between high- and low-profitability firms when *Discretion* = 1 does not isolate differences in the readability of discretionary text. I assume that non-discretionary text permeates the MD&A because (1) firms can exercise discretion in the way they write about non-discretionary content, and (2) voluntary disclosure is likely still related to the economic environment; firms do not disclose in a vacuum. Interpreting β_2 relies on *Discretion* being positively correlated with the level of discretion, i.e., text contains more discretion when *Discretion* = 1 than when *Discretion* = 0.

Figure 2 illustrates graphically the three possible cases for the sign of β_2 . In constructing Figure 2, I assume that $\beta_1 > 0$, i.e., low-profitability firms have less readable non-discretionary text than high-profitability firms, so that all cases are consistent with the on-average result that *Readability* is positively associated with *Profitability* (Li 2008). For parsimony, I focus on the

highest and lowest levels of *Profitability* and *Discretion*. Figure 2, Panel A, illustrates the first case, i.e., $\beta_2 > 0$. As assumed, for text with a low level of discretion, readability is lower when profitability is lower. As text contains more discretion, the difference in readability between high- and low-profitability firms increases. If $\beta_2 > 0$, this suggests that low-profitability firms have less readable discretionary text than high-profitability firms, causing the difference in readability to increase, and is consistent with obfuscation. Figure 2, Panel B, illustrates the second case, i.e., $\beta_2 < 0$. As text contains more discretion, the difference in readability between high- and low-profitability firms decreases. If $\beta_2 < 0$, this suggests that low-profitability firms have more readable discretionary text than high-profitability firms, causing the difference in readability to decrease, and is consistent with clear disclosure. Figure 2, Panel C, illustrates the third case, i.e., $\beta_2 = 0$. Low-profitability firms have less readable MD&A text, but this difference is invariant to the level of discretion in the text. Assuming *Discretion* is a valid proxy, $\beta_2 = 0$ suggests that high- and low-profitability firms use discretionary text with similar readability and that readability is higher for high-profitability firms for non-discretionary reasons.

Equation (1) includes industry * *Discretion* fixed effects, based on the Fama-French 48 industry classification, and year * *Discretion* fixed effects.¹⁵ These fixed effects serve two purposes. First, the fixed effects account for cross-sectional and time-series differences in the relation between *Readability* and *Discretion*, and help mitigate the concern that inferences are attributable to some industry- or time-specific language. Second, the fixed effects are controls for any mechanical relation between *Readability* and *Discretion*. Psycholinguistics find that

¹⁵ The main effect of *Discretion* is subsumed by the fixed effects, so it is omitted. Because *Discretion* is identically distributed across firm quarters, industry and year fixed effects that are not interacted with *Discretion* are redundant and are therefore also omitted.

readability is higher for more common words because readers more easily recall their meaning, and for more predictable sentence structures because readers can more easily link words together into logical statements (Haberlandt and Graesser 1985, Crossley et al. 2008).¹⁶ The psycholinguistics research implies that my proxies for *Readability* and *Discretion* are negatively correlated mechanically. The fixed effects are controls for any relation between *Discretion* and *Readability* that is the same for firms in the same industry or year, which would include any mechanical relation.

4. Sample, descriptive statistics, and validating *Discretion*

4.1. Sample

My sample consists of MD&As from all Forms 10-K and 10-Q on EDGAR from 1993-2015 with sufficient data from Compustat to construct my profitability proxies. I use the largest sample possible to maximize variation in sentence commonness, so that I can more easily distinguish common and uncommon sentences. I have data on *Earn (Profit)* for 407,209 (413,332) firm quarters, representing approximately 15,000 firms. To calculate *Discretion*, I require an electronically extractable MD&A and at least five non-boilerplate sentences.

Approximately 77 percent of firm quarters meet these criteria.^{17, 18} After calculating *Discretion*, I randomly sample 6,000 firms from the initial 15,000 and use all available firm quarters in the analyses that follow.¹⁹ When using *Earn (Profit)* as my measure of profitability, my final sample

¹⁶ The computational linguistics literature often uses various measures of language commonness as inputs to models predicting document readability (e.g., Feng et al. 2010).

¹⁷ When I cannot electronically extract the MD&A from Forms 10-K and 10-Q, the most common reason is that the MD&A is incorporated by reference and presented as an exhibit elsewhere in the forms and my algorithm to extract the MD&A is unable to identify the beginning and ending of the section.

¹⁸ Untabulated statistics reveal that the mean (median) MD&A is comprised of approximately 55 (35) non-boilerplate sentences, with the mean (median) sentence containing 24 (23) words.

¹⁹ Computational memory limits prevent me from utilizing the full sample of 15,000 firms. The equations I estimate would have approximately 1.5 million observations representing approximately 16.5 million MD&A sentences in total.

consists of 134,044 (135,955) firm-quarters, or 670,220 (679,775) firm-quarter-sentence quintile observations, from 5,185 (5,214) firms. All variables are defined in Appendix B.

4.2. Descriptive statistics

4.2.1. Univariate statistics and correlations

Table 1 presents descriptive statistics for the variables in my main analyses. Panel A presents statistics for the sample. Mean (median) *Readability* is equal to -18.85 (-18.81), which implies a reading level equivalent to seven years of post-secondary education with a standard deviation of approximately four grade levels. Mean (median) *Earn*, operating earnings over the prior twelve months divided by total assets, is -0.07 (0.02). Approximately two-thirds of firms have net income of zero or higher, i.e., *Profit* = 1. *Discretion* is a quintile rank scaled from 0 to 1, with mean and median equal to 0.5. Panel B presents *Readability* and *Discretion* for the two levels of *Profit*. Consistent with prior research, *Readability* is higher when *Profit* = 1 by approximately half a reading grade level. *Discretion* is identically distributed across the two levels of *Profit*, by construction.

4.2.2. Trigrams that differ with profitability

To provide descriptive evidence on how MD&A text differs with profitability, Appendix C presents the trigrams that differ between firms with *Earn* = 1 and *Earn* = 0, i.e., the 20% of firms with the highest and lowest profitability. The trigrams that are more frequent for high-profitability firms suggest that high-profitability firms discuss effective tax rates and compare performance using percentages more frequently than low-profitability firms. The trigrams that are more frequent for low-profitability firms suggest that low-profitability firms discuss net losses, accumulated deficits, and the inability to continue as a going concern more frequently than high-profitability firms. These text differences, which are presumably non-discretionary, are

difficult to account for with control variables because they are closely related to profitability itself. For example, losses result in accumulated deficits and going concern issues, while profits result in paying taxes.

4.3. Validating Discretion

To validate that *Discretion* is positively correlated with the level of discretion in text, I test for three properties. First, I examine the relation between *Discretion* and forward-looking statements. Second, I examine the relation between *Discretion* and disclosure tone. Third, I examine the readability of less-discretionary text over time.²⁰

4.3.1. Discretion and forward-looking statements

Providing forward-looking information is generally voluntary, which implies that text with more discretion contains more forward-looking statements. To test this, I estimate the following equation:

$$ForLook_{itu} = \beta_0 + \beta_1 Discretion_{itu} + \epsilon \quad (2)$$

ForLook is the proportion of sentences in quintile u that contain forward-looking statements, multiplied by 100 so that a unit represents one percent. I predict $\beta_1 > 0$, which implies that text I identify as having more discretion contains more forward-looking statements. *Discretion* is measured by grouping MD&As by *Earn*.²¹ Table 2, Panel A, reports the results. The coefficient on *Discretion* is significantly positive (coef. = 4.887, t-stat. = 32.97). The results indicate that

²⁰ Appendix A (section A.2) provides an example M&A and lists every sentence and the value of *Discretion* for each sentence. Sentences with the lowest discretion, i.e., *Discretion* = 0, discuss a new accounting standard the company implemented. Sentences with the highest discretion, i.e., *Discretion* = 1, explain increases or decreases in performance, plans for future financing, and details on the number of units sold. Because discussing the new accounting standard likely involves less discretion than discussing details on the number of units sold, these differences are consistent with *Discretion* being positively correlated with the level of discretion.

²¹ For all tests presented in Table 2, *Discretion* is measured by grouping MD&As by *Earn*. All inferences in Table 2 are identical when *Discretion* is measured by grouping MD&As by *Profit* instead (untabulated).

text with higher *Discretion* has more forward-looking statements, which is consistent with *Discretion* being positively correlated with the level of discretion in text.

4.3.2. Discretion and tone

I presume that the average firm exercises discretion by voluntarily disclosing positive information or presenting required information using more positive tone. To test this, I estimate the following equation:

$$Tone_{itu} = \beta_0 + \beta_1 Discretion_{itu} + \epsilon \quad (3)$$

Tone is net positive tone based on the LM Dictionary, calculated over quintile of sentences u for firm i in period t . To calculate *Tone*, I count the number of weighted positive and negative words, net them, and divide by the total number of words.²² I multiply this net proportion of positive words by 100 and scale it to have standard deviation of 1 across the sample. I predict $\beta_1 > 0$, which implies that text I identify as having more discretion has more positive tone. Table 2, Panel B, reports the results. The coefficient on *Discretion* is significantly positive (coef. = 0.530, t-stat. = 32.10). The results indicate that text with higher *Discretion* has more positive tone, which is consistent with *Discretion* being positively correlated with the level of discretion in text.

The results in Table 2, Panels A and B, are consistent with firms, on average, exercising discretion in a similar way by using more positive tone and more forward-looking statements. Even though firms likely use the same set of words to denote forward-looking statements, e.g., “expect”, or the same set of positive-toned words, these words occur more frequently in less

²² Following Loughran and McDonald (2011), I weight words by the inverse of the number of Form 10-Ks in which they occur, i.e., inverse document frequency (IDF), as found in the LM Dictionary. IDF weighting assumes that words that occur across more documents are less useful in distinguishing positive and negative tone and assigns them less weight.

common sentences, i.e., sentences with higher *Discretion*. Thus, my approach can accommodate, and even provides evidence of, firms exercising discretion in a similar way.

4.3.3. The readability of less-discretionary text over time

Dyer et al. (2017a) find that Form 10-K readability has decreased over time, primarily due to text related to required disclosure. This implies that less-discretionary text, which is more likely to contain required disclosure, exhibits decreasing readability over time. To test this, I estimate the following equation:

$$Readability_{itu} = \beta_0 + \beta_1 Year_t + \beta_2 Year_t * Discretion_{itu} + \beta_3 Discretion_{itu} + \epsilon \quad (4)$$

Year is the year of the fiscal quarter-end date. I predict $\beta_1 < 0$, which implies that text I identify as having the least amount of discretion is decreasing in readability over time, and that $\beta_2 > 0$, which implies that text I identify as having more discretion exhibits relatively less of a decrease over time. Table 2, Panel C, reports the results. The coefficient on *Year* is significantly negative (coef. = -0.083, t-stat. = -10.61), and the coefficient on the interaction of *Year* and *Discretion* is significantly positive (coef. = 0.073, t-stat. = 9.38). The results suggest that the least-discretionary text is decreasing in readability over time and that text with more discretion exhibits less of a decrease.²³ Both results are consistent with Dyer et al. (2017a) and with *Discretion* being positively correlated with the level of discretion in text.

5. Results

5.1. The relation between profitability and the readability of discretionary text

Table 3 presents the results of estimating Equation (1). Panel A (Panel B) presents results using *Earn (Profit)* as the proxy for *Profitability* and where *Discretion* is measured by grouping

²³ The negative coefficient on *Discretion* in Table 1, Panel C, reflects the mechanical negative relation between *Readability* and *Discretion* as discussed in Section 3.2.

MD&As based on *Earn (Profit)*. In column 1, the coefficient on *Earn* is significantly positive (coef. = 0.652, t-stat. = 10.29), which indicates that for the least-discretionary text, firms with lower earnings have less readable text than firms with higher earnings. The coefficient on the interaction of *Earn* and *Discretion* is significantly negative (coef. = -0.264, t-stat. = -4.03). This result indicates that as text contains more discretion, readability increases more for low-profitability firms than for high-profitability firms, which suggests that low-profitability firms' discretionary text is more readable. The result is inconsistent with obfuscation and instead consistent with clear disclosure.

Panel A, Column 2, presents results after including control variables that Li (2008, p. 232) finds are significantly associated with MD&A readability: an indicator for the presence of a special item, the market-to-book ratio, the number of geographical segments, the volatility of earnings, the volatility of returns, and an indicator for whether the firm engages in a seasoned equity offering. I include main effects for these controls and interact them with *Discretion* to account for any influence they have on the relation between *Discretion* and *Readability* across firms. The control for the presence of special items in particular accounts for any relation between *Discretion* and *Readability* that is attributable to unusual economic events, e.g., one-time shocks, to the extent that any text related to these events remains even after removing boilerplate sentences. The inferences are identical in Column 2, though the data requirements for the control variables reduce the sample size by approximately 40%.²⁴

To test whether the inferences from Column 1 hold when examining within-firm variation in profitability, I estimate Equation (1) with firm * *Discretion* fixed effects in place of

²⁴ Because the inferences from Column 2 are identical to those in Column 1, but the data requirements are substantially more restrictive, I do not include the controls from Li (2008) and their interactions in the remaining analyses in the paper.

the industry and year fixed effects. Some firms may have low profitability in multiple periods, which makes it difficult to separate firm characteristics and profitability. I use firm fixed effects interacted with *Discretion* to account for firm-specific differences in the readability of discretionary text over time. This allows me to interpret *Earn* as high- or low-profitability relative to the firm's own time series. Panel A, Column 3, presents the results of estimating Equation (1) with firm * *Discretion* fixed effects. The coefficient on the interaction between *Earn* and *Discretion* in Column 3 is negative but not significantly different from zero (coef. = -0.114 , t-stat. = -1.64). While the results are directionally consistent with clear disclosure, there is no statistically significant evidence.

Table 3, Panel B, presents the results using *Profit* as the measure of profitability. The inferences are identical to those in Panel A, with the exception of Column 3. When including firm * *Discretion* fixed effects, the coefficient on the interaction of *Profit* and *Discretion* is significantly negative (coef. = -0.190 , t-stat. = -5.18). This suggests that when firms have lower profitability relative to their own time series, they have more readable discretionary text.

Overall, the results in Table 3 indicate that low-profitability firms have more readable discretionary text than high-profitability firms, which implies that low-profitability firms disclose information clearly rather than obfuscating. Figure 3 illustrates the results from Table 3, Column 1 of Panels A and B. When *Discretion* = 0, which represents the least-discretionary text, low-profitability firms have less readable text. When *Discretion* is higher, the difference in readability between high- and low-profitability firms is smaller, similar to Figure 2, Panel B, and consistent with low-profitability firms having more readable discretionary text.

By measuring the level of discretion in text, I provide evidence that the overall positive relation between *Readability* and *Profitability* in prior research results from non-discretionary

differences in text, which is consistent with the explanation proposed by Bloomfield (2008) and evidence from an experimental laboratory in Rennekamp (2012b). My results are consistent with low-profitability firms exercising discretion by providing more readable disclosures.²⁵

5.2. *The readability of discretionary text for low-profitability firms specifically*

Table 3 presents evidence on the difference in discretionary readability between high- and low-profitability firms. The evidence in Table 3 could reflect low-profitability firms exercising discretion by (a) increasing readability, or (b) decreasing readability but doing so less than high-profitability firms. To distinguish these explanations, I estimate the following equation:

$$\begin{aligned}
 \text{Readability}_{itu} = & \beta_1 \text{Low Earn}_{it} + \beta_2 \text{Low Earn}_{itu} * \text{Discretion}_{it} + \\
 & \beta_3 \text{High Earn}_{it} + \beta_4 \text{High Earn}_{itu} * \text{Discretion}_{it} + \\
 & \Sigma \text{Industry}_{it} * \text{Discretion}_{itu} + \Delta \text{Year}_t * \text{Discretion}_{itu} + \epsilon
 \end{aligned} \tag{5}$$

Low Earn (*High Earn*) is an indicator equal to 1 if firm *i* in quarter *t* is in the lowest (highest) quintile of *Earn*, which corresponds to the lowest (highest) profitability, and to 0 otherwise. Firm-quarters for which both *Low Earn* and *High Earn* are 0 correspond to the three middle quintiles of *Earn*, i.e., average earnings. Relative to Equation (1), Equation (5) separates the extreme quintiles of *Earn* and benchmarks the readability of their discretionary text against the three middle quintiles. Assuming no significant relation between readability and discretion for the average firm, estimating Equation (5) allows me to determine whether the lowest-profitability firms exercise discretion by increasing readability.

²⁵ Guay et al. (2016) finds that less profitable firms are less likely to issue management guidance when their Form 10-Ks are less-readable. My results suggest that less profitable firms provide more readable discretionary text in Form 10-K to compensate for less readable non-discretionary text, and therefore have less need to clarify disclosure using management guidance.

The coefficient β_1 is the difference in *Readability* of the least-discretionary text for low-profitability firms, relative to firms with average earnings. If $\beta_1 < 0$, this is consistent with low-profitability firms having less readable non-discretionary text than firms with average earnings. The coefficient β_2 is the coefficient of interest. In Equation (5), $\beta_2 < 0$ ($\beta_2 > 0$) is consistent with obfuscation (clear disclosure) by low-profitability firms.

Table 4, Column 1, presents the results of estimating Equation (5). The coefficient on *Low Earn* is significantly negative (coef. = -0.688, t-stat. = -13.37), and the coefficient on the interaction of *Low Earn* and *Discretion* is significantly positive (coef. = 0.549, t-stat. = 9.66). These two results imply that, relative to a firm with average earnings, firms with low earnings have less readable non-discretionary text and more readable discretionary text. The coefficient on *High Earn* is not significantly different from zero (coef. = 0.037, t-stat. = 0.65), and the coefficient on the interaction of *High Earn* and *Discretion* is significantly positive (coef. = 0.243, t-stat. = 4.44), but significantly smaller than the coefficient on *Low Earn * Discretion* (p-value from F-test < 0.01). The inferences for low-profitability firms are similar in Column 2, which includes firm * *Discretion* fixed effects.

Overall, Table 4 provides evidence that readability increases with discretion for low-profitability firms and that this increase is larger than any such increase for high-profitability firms. These results suggest that low-profitability firms exercise discretion by providing investors with more readable disclosures. The results in Tables 3 and 4 are inconsistent with obfuscation and instead consistent with clear disclosure.²⁶

²⁶ The inferences in Tables 3 and 4 are robust to (a) using various phrase lengths to calculate *Discretion*, specifically unigrams, bigrams, or quadrigrams; (b) using only MD&As from Form 10-K, with *Earn* and an annual profit/loss indicator as *Profitability* measures; (c) using Flesch Reading Ease (Flesch 1948) to measure *Readability* rather than the Gunning fog index; or (d) defining boilerplate sentences as those repeated verbatim in the previous or subsequent filing, i.e., $\delta = 0$.

6. Additional analyses

6.1. Differences in disclosure content

To understand what low-profitability firms disclose that is more readable, I examine how disclosure content differs between the discretionary text of high- and low-profitability firms. First, I examine differences in tone to determine whether low-profitability firms disclose more positive information that is more readable, perhaps to distract investors away from any negative information. Tan et al. (2014) find that investors are more strongly influenced by positive tone when readability is low, and my main results imply that low-profitability firms have less readable non-discretionary text. Second, I examine differences in technical disclosure to determine whether high- and low-profitability firms provide differing amounts of technical disclosure. Bushee et al. (2017) suggests technical disclosure, i.e., detailed, less-readable information that the firm provides to inform rather than to obfuscate, can manifest as less readable text. Less profitable firms could have more readable discretionary text because they provide less technical disclosure.

6.1.1. Tone

To test for differences in the tone of discretionary text between high- and low-profitability firms, I estimate Equation (1) but use *Tone* as the dependent variable rather than *Readability*. The coefficient of interest is the coefficient on the interaction of *Profitability* and *Discretion*. A significantly negative (positive) coefficient indicates that discretionary text from firms with lower (higher) profitability has more positive tone.

Table 5 presents the results. Column 1 (Column 2) presents results using *Earn (Profit)* as the proxy for *Profitability*. In column 1, the coefficient on the interaction of *Earn* and *Discretion* is significantly positive (coef. = 0.142, t-stat. = 6.11), which indicates that when text has more

discretion, firms with lower earnings have less positive tone than firms with higher earnings. The inferences are similar in Column 2, using *Profit* as the proxy for *Profitability*. The results are inconsistent with low-profitability firms attempting to distract investors by using more positive tone in discretionary text than high-profitability firms and making such text more readable. Rather, low-profitability firms use less positive tone in discretionary text than high-profitability firms and yet this text easier to read.

6.1.2. Technical disclosure

To test whether high- and low-profitability firms use technical disclosure differently in discretionary text, I estimate Equation (1) but use as the dependent variable one of *Attrib*, *ForLook*, or *Risk*, three proxies that represent features of technical disclosure, i.e., sentences containing these features are likely less readable because of detailed information. *Attrib*, *ForLook*, and *Risk* are, respectively, the proportion of sentences in quintile u that contain attributions, forward-looking statements, i.e., the firm discloses that something is expected or likely to happen in the future, and risk statements, i.e., the firm discloses that something is possible or could happen in the future.^{27,28} The proportions are multiplied by 100 so that a unit represents one percent. In the average quintile of sentences, approximately 27 percent of the sentences contain attributions, 7 percent contain forward-looking information, and 3 percent contain risk statements (untabulated). The coefficient of interest is the coefficient on the

²⁷ To construct *Attrib*, *ForLook*, and *Risk*, I build custom dictionaries of words and phrases for each feature and programmatically search sentences for these words and phrases. To assess accuracy, I randomly sample 100 sentences and manually code whether the sentence contains each disclosure feature. The variables and my manual coding are highly correlated (Pearson correlations: *Attrib* 83%; *ForLook* 63%; *Risk* 73%).

²⁸ To validate that these features are associated with less readable text and are proxies for technical disclosure, I estimate the following equation:

$$Readability_{itu} = \beta_1 Attrib_{itu} + \beta_2 ForLook_{itu} + \beta_3 Risk_{itu} + \beta_4 Discretion_{itu} + \epsilon$$

β_1 , β_2 , and β_3 are the coefficients of interest. A negative coefficient indicates that the feature's presence is associated with lower readability. In untabulated results, all three coefficients of interest are significantly negative, which indicates that sentences containing attributions, forward-looking statements, and risk statements are less readable.

interaction of *Profitability* and *Discretion*. A significantly negative (positive) coefficient indicates that low-profitability firms use more (less) of the feature in discretionary text than high-profitability firms.

Table 6, Panels A and B, present the results. Panel A (Panel B) uses *Earn (Profit)* as the proxy for *Profitability*. In Panel A, the coefficient on the interaction of *Earn* and *Discretion* and is significantly positive when the dependent variable is *Attrib* (coef. = 2.669, t-stat. = 5.36) and not significantly different from zero in the other specifications. In Panel B, the coefficient on the interaction of *Profit* and *Discretion* is significantly positive when the dependent variable is *Attrib* (coef. = 1.778, t-stat. = 5.73). When the dependent variable is *Risk*, the coefficient is significantly negative (coef. = -0.133, t-stat. = -3.11). The results in Panel B suggest that low-profitability firms have fewer attributions and more risk statements in discretionary text than high-profitability firms. The result that low-profitability firms use fewer attributions suggests that low-profitability firms are not simply providing more readable excuses to investors. The result that low-profitability firms use more risk statements suggests that low-profitability firms are not simply providing vague, uninformative disclosures.²⁹

6.2. *Profitability persistence*

The results in Tables 3 and 4 suggest that low-profitability firms do not obfuscate on average. Firms might have stronger incentives to hide information about persistent low profitability, so obfuscation may be more likely when low-profitability persists (Li 2008).

Alternatively, if low-profitability firms exercise discretion by increasing readability to garner a

²⁹ In untabulated analyses, I re-estimate Equation (1) and include *Tone*, *Attrib*, *ForLook*, and *Risk* as control variables for differences in disclosure content. The inferences are identical to those in Table 3, i.e., low-profitability firms have more readable discretionary text than high-profitability firms. This result suggests that these differences in disclosure content do not fully explain the differences in discretionary readability between high- and low-profitability firms.

reputation for clear disclosure (Graham et al. 2005), their reputation may be enhanced the most by making disclosures readable when low-profitability persists.

I test how the readability of discretionary text differs between high- and low-profitability firms when profitability is persistent by estimating Equation (1) and interacting the variables with *Persist*, an indicator equal to 1 when *Profitability* in quarter t equals *Profitability* in quarter $t+4$, and to 0 otherwise. I compare profitability in the current quarter to profitability four quarters ahead to hold constant any seasonality. The coefficient of interest is the coefficient on the triple interaction of *Profitability* * *Discretion* * *Persist*. A positive (negative) coefficient is consistent with low-profitability firms using less readable (more readable) discretionary text than high-profitability firms when profitability is persistent.

Table 7 presents the results. Column 1 (Column 2) presents results when *Earn (Profit)* is the measure of *Profitability*. The coefficient on *Profitability* * *Discretion* * *Persist* is significantly negative in both columns, which indicates that when profitability persists, low-profitability firms use even more readable discretionary text than high-profitability firms. This is inconsistent with obfuscation, and is instead consistent with low-profitability firms making negative information more readable to garner a reputation for clear disclosure.

6.3. *Financial constraints*

If low-profitability firms use more readable discretionary text to garner a reputation for clear disclosure, this implies that low-profitability firms should use even more readable discretionary text when they have greater need for garner reputational capital. I use financial constraints as a proxy for firms' reputational capital needs, and posit that financially constrained firms, which are more likely to require financial capital in the future, are more likely to garner reputational capital by making negative information more readable.

I test how the readability of discretionary text differs between high- and low-profitability firms when firms have greater financial constraints by estimating Equation (1) and interacting all of the variables with *Constrained*, a proxy that increases in firms' financial constraints. *Constrained* is based on the SA Index, a weighted combination of firm age, firm size in assets, and firm size squared, from Hadlock and Pierce (2010), which finds that this index is the best summary measure of financial constraints. After calculating the SA Index using the weights from Hadlock and Pierce (2010), I rank index values into quintiles over the sample and scale the ranks to range between 0, i.e. less constrained, and 1, i.e., more constrained. The coefficient of interest is the coefficient on the triple interaction of *Profitability * Discretion * Constrained*. A negative coefficient is consistent with low-profitability firms using more readable discretionary text when they are more financially constrained.

Table 8 presents the results. Column 1 (Column 2) presents results when *Earn (Profit)* is the measure of *Profitability*. The coefficient on *Profitability * Discretion * Constrained* is significantly negative in both columns, which indicates that when firms are more financially constrained, low-profitability firms use even more readable discretionary text than high-profitability firms. This is consistent with low-profitability firms making information more readable in order to garner a reputation for clear disclosure.

7. Conclusion

This study provides evidence that less profitable firms have more readable discretionary MD&A text. Prior research examines whether less profitable firms obfuscate information by decreasing readability, but does not distinguish discretionary and non-discretionary text. After measuring the level of discretion in MD&A sentences, I find that the on-average result in prior research, i.e., that less profitable firms have less readable text, is attributable to differences in

non-discretionary text. Instead, low-profitability firms have more readable discretionary text than high-profitability firms.

My results suggest that low-profitability firms face an economic environment that is complicated to explain, yet exercise discretion by making disclosures more readable. Obfuscation may occur in other settings, e.g., conference calls, but my results are inconsistent with low-profitability firms obfuscating in the MD&A. Instead, my results are consistent with low-profitability firms providing clear disclosure. My inferences rely on the assumption that common (uncommon) sentences contain less (more) discretion. If this assumption holds, then my study provides insights into how firms' disclosure readability decisions differ with profitability.

Future research can use my approach for measuring discretion in settings where the variable of interest, e.g., profitability, leads to non-discretionary differences in text. For example, researchers interested in the relation between firm litigation risk and MD&A tone should worry that non-discretionary differences in MD&A text, related to litigation risk, could confound tone comparisons. Using my approach, researchers would group MD&As by litigation risk, use commonness to identify the level of discretion in sentences, and then test how the relation between tone and litigation risk differs as text contains more discretion. So long as common (uncommon) sentences contain less (more) discretion, researchers can use my approach to provide evidence of firms' disclosure decisions.

References

- Berger, P. E. 2011. "Challenges and opportunities in disclosure research—A discussion of 'the financial reporting environment: Review of the recent literature'". *Journal of Accounting and Economics* 51: 204-218.
- Bloomfield, R. 2002. "The 'Incomplete Revelation Hypothesis' and Financial Reporting". *Accounting Horizons* 16(3): 233-243.
- Bloomfield, R. 2008. "Discussion of 'Annual report readability, current earnings, and earnings persistence'". *Journal of Accounting and Economics* 45: 248-252.
- Bonsall, S. A. IV, A. J. Leone, B. P. Miller, and K. Rennekamp. 2017. "A Plain English Measure of Financial Reporting Readability". *Journal of Accounting and Economics* 63: 329-357.
- Brown, P. F., S. A. Della Pietra, V. J. Della Pietra, J. C. Lai, and R. L. Mercer. 1992. "Estimating the Entropy of English". *Computational Linguistics* 18(1): 31-40.
- Brown, S. V. and J. W. Tucker. 2011. "Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications". *Journal of Accounting Research* 49(2): 309-346.
- Bushee, B. J., I. D. Gow, and D. J. Taylor. 2017. "Linguistic Complexity in Firm Disclosures: Obfuscation or Information?" *Journal of Accounting Research* (Forthcoming).
- Cazier, R. A. and R. J. Pfeiffer Jr. 2017. "Say Again? Assessing Redundancy in 10-K Disclosures". *Journal of Financial Reporting* (Forthcoming).
- Crossley, S. A., J. Greenfield, and D. S. McNamara. 2008. "Assessing Text Readability Using Cognitively Based Indices". *TESOL Quarterly* 42(3): 475-493.
- Crowley, R. 2016. "Market Reaction to Shifting of Voluntary Disclosures across Disclosure Channels". Working paper.
- Dyer, T., M. Lang, and L. Stice-Lawrence. 2016. "Do Managers Really Guide Through the Fog? On the Challenges in Assessing the Causes of Voluntary Disclosure". *Journal of Accounting and Economics* 62: 270-276.
- Dyer, T., M. Lang, and L. Stice-Lawrence. 2017a. "The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation". *Journal of Accounting and Economics* 64: 221-245.
- Dyer, T., M. Lang, and L. Stice-Lawrence. 2017b. "What Have We Learned and Where Do We Go with Textual Research? A Discussion of Cazier and Pfeiffer (2017)". *Journal of Financial Reporting* (Forthcoming).
- Feng, L., M. Huenerfauth, M. Jansche, and N. Elhadad. 2010. "A comparison of features for automatic readability assessment". *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*: 276-284.
- Flesch, R. 1948. "A New Readability Yardstick". *Journal of Applied Psychology* 32(3): 221-233.
- Gow, I. D., G. Ormazabal, and D. J. Taylor. 2010. "Correcting for Cross-Sectional and Time-Series Dependence in Accounting Research". *The Accounting Review* 85(2): 483-512.

- Grossman, S. J. and J. E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets". *American Economic Review* 70(3): 393-408.
- Guay, W., D. Samuels, and D. Taylor. 2016. "Guiding through the fog: financial statement complexity and voluntary disclosure". *Journal of Accounting and Economics* 62: 234-269.
- Haberlandt, K. F. and A. C. Graesser. 1985. "Component Processes in Text Comprehension and Some of Their Interactions". *Journal of Experimental Psychology: General* 114(3): 357-374.
- Hadlock, C. J. and J. R. Pierce. 2010. "New Evidence on Measuring Financial Constraints: Moving Beyond the KZ Index". *The Review of Financial Studies* 23(5): 1909-1940.
- Jones, J. J. 1991. "Earnings Management During Import Relief Investigations". *Journal of Accounting Research* 29(2): 193-228.
- Jurafsky, D. and J. H. Martin. 2009. *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.). Pearson Education Inc., Upper Saddle River, NJ.
- Lawrence, A. 2013. "Individual investors and financial disclosure". *Journal of Accounting and Economics* 56: 130-147.
- Lee, Y. 2012. "The Effect of Quarterly Report Readability on Information Efficiency of Stock Prices". *Contemporary Accounting Research* 29(4): 1137-1170.
- Levenshtein, V. I. 1966. "Binary codes capable of correcting deletions, insertions, and reversals". *Soviet Physics – Doklady* 10(8): 707-710.
- Li, F. 2008. "Annual report readability, current earnings, and earnings persistence". *Journal of Accounting and Economics* 45: 221-247.
- Loughran, T. and B. McDonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks". *Journal of Finance* 66(1): 35-65.
- Loughran, T. and B. McDonald. 2016. "Measuring Readability in Financial Disclosures". *Journal of Finance* 69(4): 1643-1671.
- Markowitz, D. M. and J. T. Hancock. 2015. "Linguistic Obfuscation in Fraudulent Science". *Journal of Language and Social Psychology* 35(4): 435-445.
- Miller, B. P. 2010. "The Effects of Reporting Complexity on Small and Large Investor Trading". *The Accounting Review* 85(6): 2107-2143.
- Rennekamp, K. 2012a. "Processing Fluency and Investors' Reactions to Disclosure Readability". *Journal of Accounting Research* 50(5): 1319-1354.
- Rennekamp, K. 2012b. "The Influence of Performance and Reporting Goals on Managers' Choice of Reporting Complexity in Disclosures". Working paper.

- Stamatatos, E. 2009. "Intrinsic Plagiarism Detection Using Character n-gram Profiles". In: Proceedings of the 3rd Int. Workshop on Uncovering Plagiarism, Authorship, and Social Software Misuse. 38-46.
- Stamatatos, E. 2011. "Plagiarism Detection Using Stopword n-grams". *Journal of the American Society for Information Science and Technology*. 62(12): 2512-2527.
- Tan, H. T., E. Y. Wang, and B. 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.

Appendix A: Text processing and MD&A example

A.1. Text processing

I perform several processing tasks to convert raw MD&A text into useable data.³⁰ I divide MD&A text into sentences, where my indicators for sentences include periods; exclamation points; question marks; bullet points, based on a set of html codes observed in SEC filings; and enumerations of lists, e.g., “(a)”. When marking sentence indicators, I rule out decimal points, e.g., “1.25”, common uses of periods, e.g., “Dr.” or “Inc.”, and I attempt to rule out abbreviations, e.g., “U.S.A.”. I mark sentence indicators prior to removing html-formatted tables because companies often use html tables as a formatting tool for important written text.

After marking sentence indicators, I screen out false sentences that reflect table rows, based on a sequence of numbers in a row, to ensure that I capture only written text. I convert all characters to lowercase and remove punctuation. I generalize certain types of text into categories to facilitate comparison and include markers for these categories. The categories include: numbers, e.g., “1.23” is marked as “<d>”; percentages, marked as “<d>%”; dollar amounts, marked as “\$<d>”; dates, marked as “<date>”; SEC form identifiers, marked as “<sec_form>”; and accounting standards and codification references, marked as “<acct_stdndr>”. As with all text classifying, identifying these categories has some imperfection. To rule out proper nouns and non-word garbage produced in processing the text, I limit all words to the Loughran McDonald (2011) Master Dictionary with supplements for contractions, e.g., “didnt”, possessive forms, e.g., “companys”, and my custom category markers. Lastly, all processed sentences must contain at least 5 words or markers for inclusion in the analyses.

A.2. Example MD&A: Swissray International Inc.

The following example comes from the Form 10-K of Swissray International Inc., filed November 19, 2002.³¹ This firm-quarter is in the lowest quintile of *Earn*. This example was generated by selecting 20 MD&A texts randomly from my sample where *Earn* is in the extreme quintiles and choosing the first text with at least 100 sentences, for representativeness, and fewer than 150 sentences, for brevity.

The table below lists every sentence in Swissray’s processed MD&A text, in order. *Processed Sentence Text* is the sentence extracted from the MD&A after performing all pre-processing.³² The marker <s> <s> (</s>), which precedes (follows) each sentence to denote the beginning (ending), is omitted from the table for brevity. *Discretion* is a proxy for the level of discretion in the sentence. See section 3.1 for detail on its measurement. *Discretion* = 1 (= 0) denotes the most (least) discretionary sentences. *Discretion* does not have a value for boilerplate sentences because these are excluded from the text.

³⁰ Python code for processing text is available upon request.

³¹ <https://www.sec.gov/Archives/edgar/data/1002137/0001172665-02-000157-index.htm>

³² Sentences #28, #81, and #92 are screened out because they have consecutive \$<d> or <d>% markers in a row and resemble the row of a table.

Sent. #	Processed Sentence Text	Discretion
1	all references herein to the registrant refer to international all references herein to the company refer to international and its subsidiaries	
2	general throughout fiscal year ended <date> the company took number of important steps to strengthen its position in the field of radiology with specific emphasis upon its line of products and additions thereto so as to expand its dedicated systems and so as to offer solutions to important aspects of digital radiography	
3	market penetration of the companys flag ship product the systems continued to increase as result of the sales efforts of the companys internal sales staff as well as the sales and distribution efforts of and the companys two principal distributors	
4	furthermore the company has been able to sell its systems into new markets including and	0.5
5	while initial market penetration resulted in the sale of <d> systems during fiscal year ended <date> fiscal year ended <date> saw substantial and significant further market penetration in that the company contracted for the sale of <d> systems <d> of which were contracted for sale in the with the balance being contracted for sale outside the inclusive of <d> systems contracted for sale to the government of	
6	fiscal year ended <date> saw further continued success with the company having contracted for the sale of <d> additional systems inclusive <d> systems sold to the government of <d> systems sold to and <d> systems sold to	1
7	during fiscal year ended <date> the company contracted for <d> systems <d> were contracted for sale in the <d> in and <d> in and rest of world	1
8	the <d> systems contracted in fiscal year ended <date> all have been sold in structured markets versus <d> units of the <d> systems sold during fiscal year ended <date> and <d> units of the <d> systems sold during fiscal year ended <date>	0.75
9	with the introduction of several new performance enhancements for its direct digital radiography product line the company has succeeded in further positioning itself with high tech image and has entered the new millennium with substantial list of potential and informatics business opportunities	1
10	the new products are principally based upon the patented technology of the	
11	see also item <d> business new products for further information	
12	the companys efforts were again recognized for the fourth consecutive year by unaffiliated third parties as evidenced by its having been chosen in <date> as one of the <d> fastest growing technology companies in the new county areas	
13	this program was sponsored by touche	

Sent. #	Processed Sentence Text	Discretion
14	ongoing efforts in the companys research and development department have recently led to the company being issued patents in both the and in for the detector the	
15	these patents relate to the optical arrangement and process for transmitting and converting primary ray images generated on two dimensional primary image array	
16	furthermore the patent for the mirror optics has been approved in the for the optical arrangement and method for electronically detecting an ray image while similar patent has been approved in in <date>	
17	for the aforesaid and the company was granted design protection in	
18	additional patent applications for the design protection for both systems have been submitted in and the and are currently pending approval	
19	see item <d> business intellectual property and patents for information regarding patent numbers and dates of issuance	
20	the company received its <date> and <date> label for its systems	
21	this certification is public statement of compliance with known standard and an extremely stringent approval process involving assessment and documentation of product sample by an independent third party organization followed by on going periodic product inspections at the manufacturing site	
22	the company was also challenged within the changing financial market environment	1
23	due to lack of liquidity the company has not been able to deliver all of its backlog during fiscal year ended <date> and also had to finance its backlog by giving away prepayment interest to its distributors of \$<d> which resulted in an impact of decreased gross margin of <d>% for the twelve month period	0.75
24	management of the company hopes that its effort to attract new capital to strengthen its liquidity will have positive result in the foreseeable future	1
25	<d> management of the company believes that its unique technology and vision for the future have set the stage for continued growth in the years to come and anticipate continued increase in demand for the companys solutions which belief is based upon results of clinical successes in the as well as in	
26	year ended <date> compared to year ended <date> results of operations net sales amounted to \$<d> for the year ended <date> compared to \$<d> for the year ended <date> decrease of \$<d> or <d>% from the year ended <date>	
27	the <d>% decrease in net sales was mainly due to the decrease in sales of units by <d>% or \$<d> related information solutions decreasing by <d>% or \$<d> and the decrease in conventional business by <d>% or <d>	0.75

Sent. #	Processed Sentence Text	Discretion
29	the companys overall sales and marketing strategy includes the use of direct sales force distribution agreements with other medical imaging companies local distributors and agreements	0.75
30	the decrease in sales of units is due to the lower sell prices for sales through distribution agreements direct sales direct sales of <d> units during fiscal year ended <date> <d> units during fiscal year ended <date> and the prepayment discount of \$<d> due to tight liquidity	1
31	the decrease in information solution is due to the companies focus in the the single largest market for the company to sell its products mainly through distribution agreements who are providing similar services to the end user whereas the decease in conventional business is due to the companys conscious effort of promoting sales of systems with corresponding decline of interest in sales of conventional ray and conventional business	0.75
32	the increase in conventional ray is due to executed conventional projects	0.75
33	the increase in service is going in line with an increase installed base of units	1
34	gross profit decreased by \$<d> or <d>% to \$<d> for the year ended <date> from \$<d> for the year ended <date>	
35	gross profit as percentage of net revenues decreased to <d>% for the year ended <date> from <d>% for the year ended <date>	
36	the decrease in gross profit as percentage of net revenues is attributable to the prepayment interest which accounts for <d>% itself and fraction of higher units sales and lower average selling price due to the concentration on strong distribution agreements	0.5
37	operating expenses decreased by \$<d> or <d>% to \$<d> or <d>% of net revenues for the year ended <date> from \$<d> or <d>% of net revenues for the year ended <date>	
38	the principal items were officers and directors compensation of \$<d> or <d>% of net sales for the year ended <date> compared to \$<d> or <d>% of net sales for the year ended <date> salaries net of officers and directors compensation of \$<d> or <d>% of net sales for the year ended <date> compared to \$<d> or <d>% of net sales for the year ended <date> and selling expenses of \$<d> or <d>% of net sales for the year ended <date> compared to \$<d> or <d>% of net sales for the year ended <date>	
39	research and development expenses were \$<d> or <d>% of net sales for the year ended <date> compared to \$<d> or <d>% of net sales for the year ended <date>	
40	the slight increase is primarily due to the development of new performance enhancements for its direct digital product line	0.5
41	general and administrative expenses decreased by \$<d> or <d>% to \$<d> or <d>% of net sales for the year ended <date> from \$<d> or <d>% of net sales for the year ended <date>	

Sent. #	Processed Sentence Text	Discretion
42	the decrease in selling and general administrative expenses is due to overall savings primarily on professional fees and services	
43	other operating expenses decreased by \$<d> or <d>% to \$<d> of net sales for the year ended <date> from \$<d> or <d>% of net sales for the year ended <date>	
44	this decrease is due to the overall savings	0.5
45	interest expenses decreased to \$ for the year ended compared to \$ for the year ended	
46	this decrease is primarily due to the decrease in amortization of debenture issuance cost and conversion benefit	0.25
47	financial condition <date> compared to <date> total assets of the company on <date> decreased by \$<d> to \$<d> from \$<d> on <date> primarily due to the decrease of other assets	
48	current assets increased by \$<d> to \$<d> on <date> from \$<d> on <date>	
49	the increase in current assets is attributable to the increase of cash and cash equivalents of \$<d> the increase of accounts receivable of \$<d> the increase in inventory of \$<d> and in prepaid expenses and sundry receivables of \$<d>	0.25
50	other assets decreased \$<d> to \$<d> on <date> from \$<d> on <date>	
51	the decrease is primarily attributable to the amortization of the licensing agreement patents and trademark software development cost and the change in accounting practice in respect to goodwill	0.75
52	on <date> the company had total liabilities of \$<d> compared to \$<d> on <date>	
53	on <date> current liabilities were \$<d> compared to \$<d> on <date>	
54	working capital at <date> was \$<d> compared to \$<d> at <date>	
55	the increase in liabilities and the decrease in working capital was due to increased borrowings from bank	0.5
56	cash flow and capital expenditures year ended <date> compared to year ended <date>	
57	cash used for operating activities for the years ended <date> and <date> was \$<d> \$<d> and \$<d> respectively	
58	cash used for investing activities for years ended <date> and <date> was \$<d> \$<d> and \$<d> respectively	
59	cash flow from financing activities for years ended <date> and <date> was \$<d> \$<d> respectively	
60	liquidity on <date> the company negotiated revolving line of credit agreement with credit for up to \$<d>	0.5
61	interest on the loan is <d>% per annum and the loan is collateralized by certain accounts receivable	0.5
62	the company is currently in default of certain financial covenant and has until <date> to cure such deficiency	0.5

Sent. #	Processed Sentence Text	Discretion
63	the company anticipates that its use of cash will remain substantial for the foreseeable future	
64	in particular management of the company expects substantial expenditures in connection with the production of the planned increase of sales the continuation of the strengthening and expansion of the companys marketing organization and to lesser degree ongoing research and development projects	
65	the company expects that funding for these expenditures will be dependent to significant extent on additional debt or equity financing	0.25
66	there can be no assurance whether or not such financing will be available on terms satisfactory to management if and when needed	
67	reference is made to the companys current report on form <sec_form> filed <date> <d>	0.25
68	as stated therein the company and its major shareholders have entered an advanced stage of negotiation for the acquisition by based private equity firm of majority of the capital stock of the company pursuant to discussions begun under binding letter of intent executed in	0.75
69	the company believes that the transaction may be completed as soon as the end of if completed would provide substantial infusion of capital sufficient to permit the company to fund its day to day operations for the near future	0.5
70	transaction such as this would allow the company to continue in operation and would preserve for the companys common shareholders residual interest however modest in the companys equity value	1
71	but it is likely that the shareholders who currently are subordinated in right of payment upon liquidation to \$<d> face amount of preferred shares would as result of the proposed transaction be subordinated to much greater amount of preferred stock and could not expect to recover any substantial value in the foreseeable future if at all	1
72	the planned transaction will go far toward resolving the companys liquidity problems which have been approaching critical stage	0.75
73	the companys operations are not providing and have not provided in the past positive cash flow and as it has indicated in its public announcements to date the company will need additional financing to survive	1
74	if the planned transaction or an alternative new financing is not secured within the coming weeks the companys would not be able to continue as going concern	0.75
75	compared to cash flow and capital expenditures year ended compared to year ended	0.25
76	cash used for operating activities for the years ended <date> and <date> was \$<d> \$<d> and \$<d> respectively	
77	cash used for investing activities for years ended <date> and <date> was \$<d> \$<d> and \$<d> respectively	

Sent. #	Processed Sentence Text	Discretion
78	cash flow from financing activities for years ended <date> and <date> was \$<d> \$<d> and \$<d> respectively	
79	year ended <date> compared to year ended <date> results of operations net sales amounted to \$<d> for the year ended <date> compared to \$<d> for the year ended <date> an increase of \$<d> or <d>% from the year ended <date>	
80	the <d>% increase in net sales was mainly due to the sales of increasing by <d>% or \$<d> and related information solutions increasing by <d>% or \$<d>	
82	the decrease in conventional ray and conventional business is due to the companys conscious effort of promoting sales of systems with corresponding decline of interest in sales of conventional ray and conventional business whereas service is going in line with an increase in related service and decrease in conventional and conventional business related service	
83	in the past the company has been substantially reliant upon medical systems but at this stage of the companys maturation process and as same continues to develop reliance upon has correspondingly decreased	
84	additionally the companys agreement with relates to conventional ray equipment which has been low profit margin item	
85	more and more this type of sale is being replaced by	
86	company sale of conventional ray equipment directly to purchasing country hospital to the ultimate user thereof and	
87	more significantly and importantly by companys sales of its systems its flagship product	
88	gross profit increased by \$<d> or <d>% to \$<d> for the year ended <date> from \$<d> for the year ended <date>	
89	gross profit as percentage of net revenues increased to <d>% for the year ended <date> from <d>% for the year ended <date>	
90	the increase in gross profit as percentage of net revenues is attributable to the fact that the percentage of sales of to total sales increased to <d>% of total sales for the year ended <date> from <d>% for the year ended <date>	
91	operating expenses decreased by \$<d> or <d>% to \$<d> or <d>% of net revenues for the year ended <date> from \$<d> or <d>% of net revenues for the year ended <date>	
93	research and development expenses were \$<d> or <d>% of net sales for the year ended <date> compared to \$<d> or <d>% of net sales for the year ended <date>	
94	the increase is primarily due to the development of the and the unit as well as introduction during calendar year <date> of five additional new products	

Sent. #	Processed Sentence Text	Discretion
95	general and administrative expenses decreased by \$<d> or <d>% to \$<d> or <d>% of net sales for the year ended <date> from \$<d> or <d>% of net sales for the year ended <date>	
96	the decrease in officers and directors compensation of \$<d> or <d>% is due to the fact that such officers received during fiscal year ended <date> common stock for services rendered which was valued at \$<d> while no shares for services rendered were issued during the succeeding fiscal year	
97	the decrease in selling and general administrative expenses is due to overall savings primarily on professional fees and services	
98	other operating expenses decreased by \$<d> or <d>% to \$<d> or <d>% of net sales for the year ended <date> from \$<d> or <d>% of net sales for the year ended <date>	
99	this decrease is due to the overall savings primarily in rent and insurance costs	
100	interest expenses decreased to \$<d> for the year ended <date> compared to \$<d> for the year ended <date>	
101	this decrease is primarily due to the decrease of interest expense for accrual of penalty interest on periodic payments required by terms of financing agreements and an increase in amortization of debenture issuance cost and conversion benefit	
102	financial condition <date> compared to <date> total assets of the company on <date> decreased by \$<d> to \$<d> from \$<d> on <date> primarily due to the decrease of current assets	
103	current assets decreased \$<d> to \$<d> on <date> from <date>	
104	the decrease in current assets is attributable to the decrease of cash and cash equivalents of \$<d> restricted cash of \$<d> the decrease of accounts receivable of \$<d> of which approximately \$<d> arises from the receipt of cash from the sale of to the decrease in inventory of \$<d> and the decrease in prepaid expenses and sundry receivables of \$<d>	
105	other assets decreased \$<d> to \$<d> on <date> from \$<d> on <date>	
106	the decrease is primarily attributable to the amortization of the licensing agreement patents trademark software development cost and the goodwill	
107	on <date> the company had total liabilities of \$<d> compared to \$<d> on <date>	
108	on <date> current liabilities were \$<d> compared to \$<d> on <date>	
109	working capital at <date> was \$<d> compared to \$<d> at <date>	
110	cash flow and capital expenditures year ended <date> compared to year ended <date>	
111	cash used for operating activities for the years ended <date> and <date> was \$<d> \$<d> and \$<d> respectively	

Sent. #	Processed Sentence Text	Discretion
112	cash used for investing activities for years ended <date> and <date> was \$<d> \$<d> and \$<d> respectively	
113	cash flow from financing activities for years ended <date> and <date> was \$<d> \$<d> and \$<d> respectively	
114	<d> effect of currency on results of operations the results of operations and the financial position of the companys subsidiaries outside of the united states are reported in the relevant foreign currency primarily in francs and then translated into us dollars at the applicable foreign exchange rate for inclusion in the companys consolidated financial statements	
115	accordingly the results of operations of such subsidiaries as reported in us dollars can vary significantly as result of changes in currency exchange rates in particular the exchange rate between the franc and the us dollar	
116	inflation inflation can affect the costs of goods and services used by the company	
117	the competitive environment in which the company operates limits somewhat the companys ability to recover higher costs through increasing selling prices	
118	moreover there may be differences in inflation rates between countries in which the company incurs the major portion of its costs and other countries in which the company sells its products which may limit the companys ability to recover increased costs if not offset by future increase of selling prices	
119	to date the companys sales to high inflation countries have either been made in francs or us dollars	
120	accordingly inflationary conditions have not had material effect on the companys operating results	
121	seasonality the companys business has historically experienced slight amount of seasonal variation with sales in the first fiscal quarter slightly lower than sales in the other fiscal quarters due to the fact that the companys first quarter coincides with the summer vacations in certain of the companys markets	
122	backlog management estimates that as of the end of fiscal year ended company had an order backlog of \$<d> which consisted of \$<d> in conventional ray equipment and services and \$<d> in digital and information solutions as compared to an order backlog of \$<d> which consisted of \$<d> in conventional ray equipment and \$<d> in as of the fiscal year ended <date>	
123	order backlog as of <date> amounted to \$<d> of which digital backlog accounted for \$<d>	0.5
124	new accounting pronouncements in <date> the financial accounting standards board issued statements of financial accounting standards no	

Sent. #	Processed Sentence Text	Discretion
	<d> business combinations and no <d> goodwill and other intangible assets effective for fiscal years beginning after <date>	
125	under the new rules goodwill and intangible assets deemed to have indefinite lives will no longer be amortized but will be subject to annual impairment tests in accordance with the statements	
126	other intangible assets will continue to be amortized over their useful lives	
127	upon adoption of no <d> the company recorded one time cash charge of approximately \$<d> to write off the carrying value of its goodwill	0.25
128	such charge is operational in nature and is reflected as cumulative effect of an accounting change in the accompanying consolidated statement of operations	0.25
129	in <date> the issued <acct_stndrd> accounting for asset retirement obligations	0
130	this statement addresses financial accounting and reporting for obligations associated with the retirement of tangible long lived assets and the associated retirement costs	0
131	this statement applies to all entities	0.25
132	it applies to legal obligations associated with the retirement of long lived <d> assets that result from the acquisition construction development and or the normal operation of long lived asset except for certain obligations of lessees	0.25
133	this statement is effective for financial statements issued for fiscal years beginning after <date>	0
134	the company does not believe the adoption of this standard will have material impact on the companys financial statements	0
135	in <date> the issued <acct_stndrd> accounting for the impairment or disposal of long lived assets	0
136	<d> this statement addresses financial accounting and reporting for the impairment or disposal of long lived assets and supersedes no <d> accounting for the impairment of long lived assets and long lived assets to be disposed of	0
137	the company does not believe the adoption of these standards will have material impact on the companys financial statements	0
138	in <date> the issued statement no <d> <acct_stndrd> accounting for costs associated with exit or disposal activities	0
140	<acct_stndrd> requires companies to recognize costs associated with exit activities when they are incurred rather than at the date of commitment to an exit or disposal plan	0
141	is to be applied prospectively to exit or disposal activities initiated after	0
142	the company is currently evaluating this standard	0.25

Appendix B: Variable definitions

Variable	Definition
<i>Attrib</i>	The proportion of sentences in a quintile of MD&A sentences that contain attributions. Sentences are classified using a custom dictionary, available upon request, and the proportion is calculated by dividing the number of sentences with at least one attribution by the total number of sentences in the quintile. The proportion is multiplied by 100 so that a unit increase corresponds to an increase of one percent.
<i>Constrained</i>	The proxy for financial constraints, following the SA Index from Hadlock and Pierce (2010). $SA\ Index = -0.737 * Size + 0.043 * Size^2 - 0.040 * Age$ where <i>Size</i> is the log of assets (<i>atq</i> from Compustat) and <i>Age</i> is the number of years the firm is listed with a non-missing stock price on Compustat. In calculating this index, unlogged <i>Size</i> is capped at \$4.5 billion, and <i>Age</i> is capped at thirty-seven years. The sample is ranked into quintiles by the SA Index and scaled to range from 0 (low financial constraints) to 1 (high financial constraints).
<i>Discretion</i>	The proxy for the level of discretion in a sentence, based on its commonness. To calculate commonness, MD&As are divided into groups based on profitability (using <i>Earn</i> or <i>Profit</i>) and split into sentences. Commonness for each sentence is calculated by dividing the sentence into a sequence of overlapping trigrams and estimating trigram probabilities based on the sentences in each MD&A group. After removing boilerplate sentences, the remaining sentences are ranked within each MD&A into quintiles by commonness. This ranking is scaled to range from 0, i.e., most common, to 1, i.e., least common. Uncommon sentences are assumed to contain more discretion than common sentences.
<i>Earn</i>	Operating earnings over the prior 12 months ($oeeps12 * cshprq$ from Compustat) scaled by total assets (<i>atq</i> from Compustat). <i>Earn</i> is scaled to range from 0 to 1.
<i>ForLook</i>	The proportion of sentences in a quintile of MD&A sentences that contain forward-looking statements. Sentences are classified using a custom dictionary, available upon request, and the proportion is calculated by dividing the number of sentences with at least one forward-looking statement by the total number of sentences in the quintile. The proportion is multiplied by 100 so that a unit increase corresponds to an increase of one percent.

Variable	Definition
<i>High Earn</i>	An indicator equal to 1 if <i>Earn</i> = 1, i.e., the firm-quarter is in the highest quintile of earnings, and to 0 otherwise.
<i>Low Earn</i>	An indicator equal to 1 if <i>Earn</i> = 0, i.e., the firm-quarter is in the lowest quintile of earnings, and to 0 otherwise.
<i>Persist</i>	An indicator equal to 1 if the measure of profitability, <i>Earn</i> or <i>Profit</i> , is the same in quarter <i>t</i> as in quarter <i>t+4</i> , and to 0 otherwise.
<i>Profit</i>	An indicator equal to 1 if quarterly net income (<i>niq</i> from Compustat) is non-negative, and 0 otherwise.
<i>Readability</i>	The readability of a quintile of MD&A sentences. Calculated as -1 multiplied by the Gunning fog index, which is calculated according to prior literature as $Fog = 0.4 * \left[\left(\frac{\#words}{\#sentences} \right) + 100 * \left(\frac{\#complex\ words}{\#words} \right) \right]$. Higher values of <i>Readability</i> denote more readable text. To identify complex words, i.e., words with 3+ syllables, I use syllable counts provided in the Loughran McDonald (2011) Master Dictionary (LM Dictionary). Because some syllable counts in the LM Dictionary are determined programmatically and may be incorrect, syllable counts from the LM Dictionary are compared to syllables determined from the Carnegie Mellon Pronunciation Dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict) and counts from the latter are used if the two differ.
<i>Risk</i>	The proportion of sentences in a quintile of MD&A sentences that contain risk statements. Sentences are classified using a custom dictionary, available upon request, and the proportion is calculated by dividing the number of sentences with at least one risk statement by the total number of sentences in the quintile. The proportion is multiplied by 100 so that a unit increase corresponds to an increase of one percent.
<i>Tone</i>	Using the LM Dictionary, this is the weighted sum of positive-toned words less the weighted sum of negative-toned words, divided by the total number of words in each quintile of MD&A sentences. The words are weighted using the inverse of their document frequency (IDF) based on the number of Form 10-Ks in which they occur, as given by the LM Dictionary. The net proportion of positive words is multiplied by 100 and scaled to have standard deviation of 1 across the sample.
<i>Year</i>	The calendar year in which the Form 10-K or 10-Q was filed. The base year, 1993, is equal to 0 so that the variable measures years since the base year.

Appendix C: Trigrams that differ with profitability

This appendix provides descriptive evidence on how trigrams, i.e., three-word phrases, differ between MD&As from firms in the highest and lowest quintiles of *Earn*. I calculate trigram frequencies in each group of firms prior to removing boilerplate sentences. To highlight substantial differences between groups of high- and low-profitability firms, I require relative frequencies to differ by at least a factor of 5 between the groups, where the relative frequency of trigram *j* for low-profitability firms is the frequency of trigram *j* divided by the total frequency of all trigrams in low-profitability firms' MD&As. I present the 50 most frequent trigrams in each group, sorted from most frequent to least frequent, from the set of trigrams with substantial differences.

C.1. Trigrams that are more frequent in highest-profitability firms

1 effective tax rate	26 in net income
2 earnings per share	27 \$<d> billion in
3 year to date	28 and <d> percent
4 <s> net income	29 <d> percentage points
5 income from operations	30 tax rate was
6 the effective tax	31 our effective tax
7 <d> basis points	32 of foreign currency
8 per diluted share	33 income from continuing
9 income tax rate	34 of <d> percent
10 <s> operating income	35 increased <d>% in
11 tax rate for	36 <d> percent for
12 <d> percent in	37 tax rate of
13 income before income	38 net income and
14 diluted earnings per	39 <d> percent </s>
15 \$<d> per diluted	40 percent of sales
16 effective income tax	41 <d> percent and
17 <s> the effective	42 income attributable to
18 net sales increased	43 net income per
19 net income for	44 rate was <d>%
20 of \$<d> billion	45 <s> income from
21 comparable store sales	46 percent for the
22 increased <d> percent	47 <d> percent to
23 to \$<d> billion	48 percent in the
24 sales increased <d>%	49 operating income for
25 to <d> percent	50 same store sales

C.2. Trigrams that are more frequent in lowest-profitability firms

1 net loss of	26 we have incurred
2 our research and	27 going concern </s>
3 series preferred stock	28 raise additional capital

4 our product candidates
5 convertible preferred stock
6 to raise additional
7 as going concern
8 loss for the
9 warrants to purchase
10 exercise price of
11 continue as going
12 the net loss
13 to continue as
14 will need to
15 development and commercialization
16 net loss for
17 phase <d> clinical
18 clinical trials and
19 product candidates </s>
20 to net loss
21 we will need
22 the series preferred
23 an exercise price
24 <s> net loss
25 clinical trials </s>

29 our net loss
30 deficit of \$<d>
31 preferred stock and
32 raise additional funds
33 and commercialization of
34 purchase <d> shares
35 loss from continuing
36 and development programs
37 of our research
38 an accumulated deficit
39 accumulated deficit of
40 since our inception
41 of series preferred
42 of the series
43 of the warrants
44 continue to incur
45 series convertible preferred
46 and clinical trials
47 shares of series
48 phase clinical trial
49 our clinical trials
50 at an exercise

Figure 1 -- Data structure for a hypothetical firm over two quarters

Firm	Quarter	Quintile of Sentences	<i>Earn</i>	<i>Profit</i>	<i>Discretion</i>	<i>Readability</i>
1	1	1	0.0	0	0.00	-16.0
1	1	2	0.0	0	0.25	-16.5
1	1	3	0.0	0	0.50	-17.0
1	1	4	0.0	0	0.75	-17.5
1	1	5	0.0	0	1.00	-18.0
1	2	1	0.5	1	0.00	-14.6
1	2	2	0.5	1	0.25	-15.2
1	2	3	0.5	1	0.50	-15.8
1	2	4	0.5	1	0.75	-16.4
1	2	5	0.5	1	1.00	-17.0

This figure illustrates the data structure for a hypothetical firm over two quarters. Each firm-quarter is represented by exactly 5 observations, one for each quintile of MD&A sentences. Sentences are ranked into quintiles based on their commonness, with more common sentences having a lower rank.

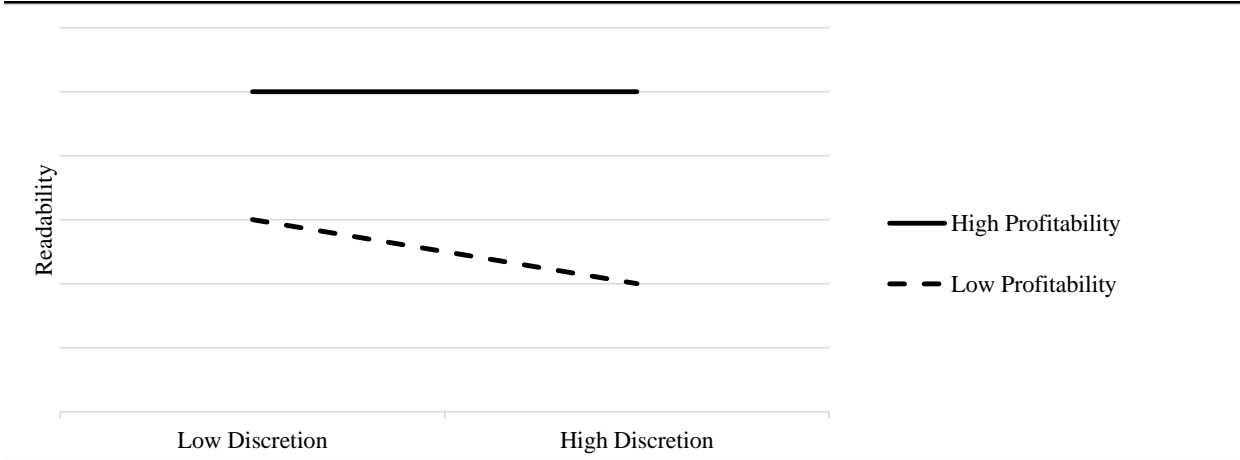
Readability is the readability of all sentences in a quintile and equal to -1 multiplied by the Gunning fog index. *Earn* is operating earnings over the prior twelve months scaled by total assets, ranked into quintiles, and scaled between 0 and 1. *Profit* is an indicator equal to 1 when quarterly net income is non-negative and to 0 when there is a loss. *Discretion* is a measure of the level of discretion in a quintile of MD&A sentences, and is a quintile ranking that ranges from 0 to 1 within each MD&A. See Appendix B for detailed variable definitions.

Figure 2 -- Predictions for estimating Equation (1)

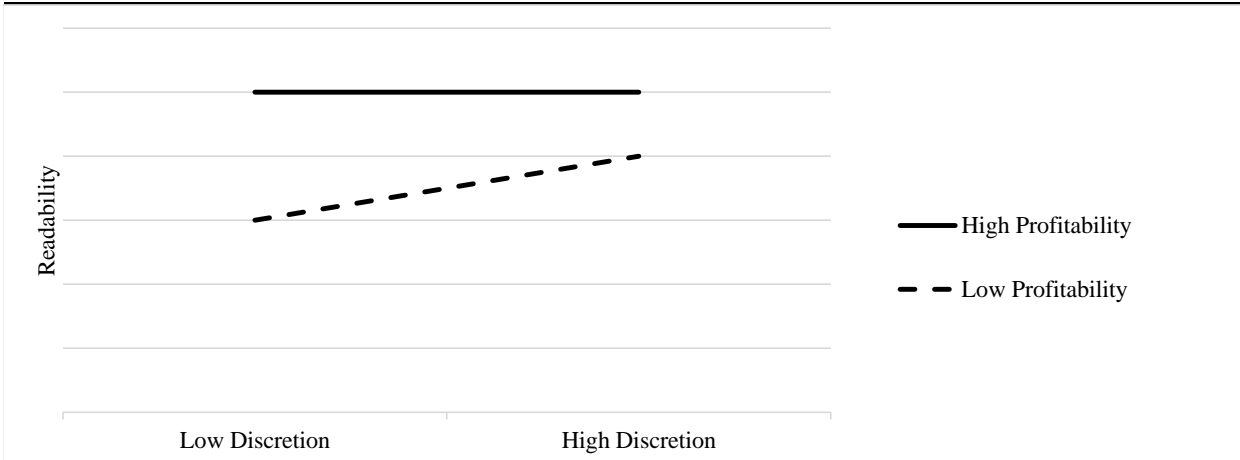
$$Readability = \beta_1 Profitability + \beta_2 Profitability * Discretion + \sum Industry * Discretion + \Delta Year * Discretion + \epsilon$$

Assuming $\beta_1 > 0$

Panel A: $\beta_2 > 0$ (Obfuscation)



Panel B: $\beta_2 < 0$ (Clear disclosure)



Panel C: $\beta_2 = 0$ (Non-discretionary differences only)

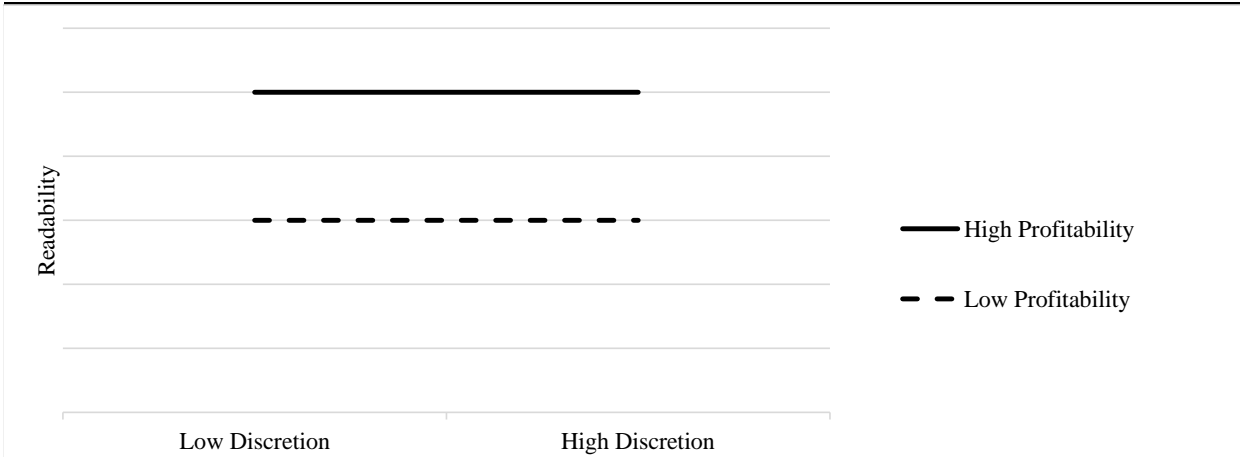
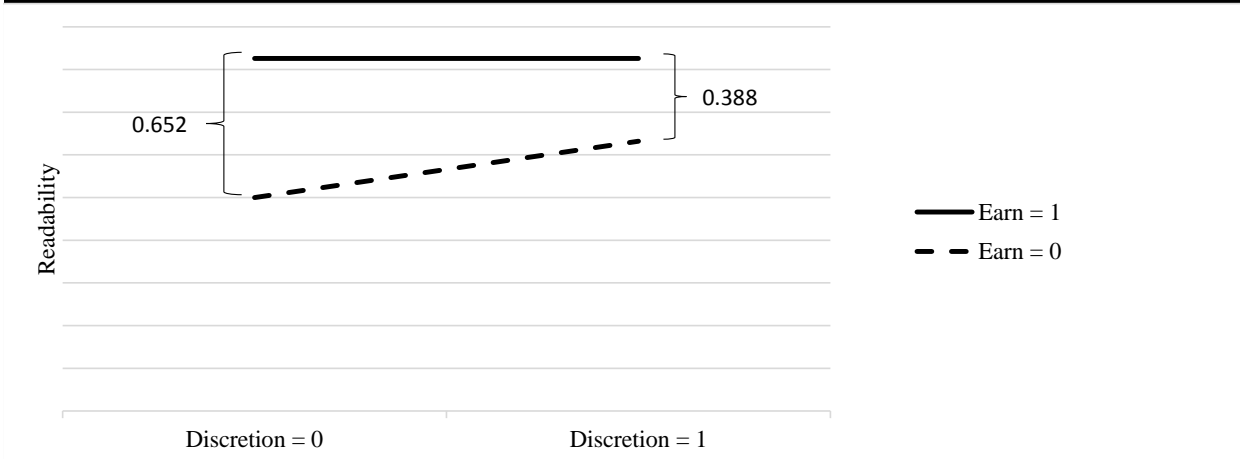


Figure 3 -- Results from estimating Equation (1)

$$Readability = \beta_1 Profitability + \beta_2 Profitability * Discretion + \sum Industry * Discretion + \Delta Year * Discretion + \epsilon$$

Panel A: Profitability = Earn



Panel B: Profitability = Profit

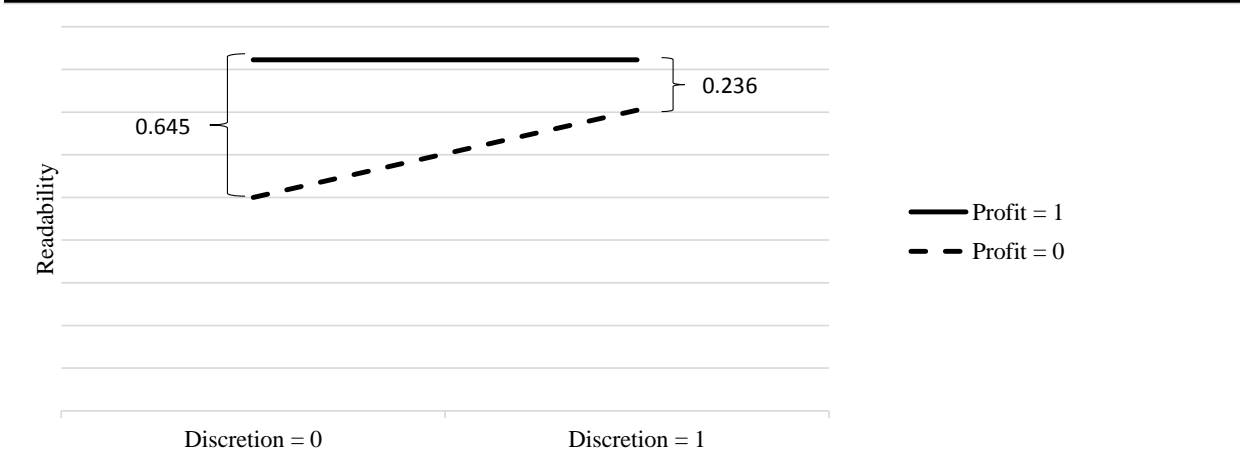


Table 1 -- Descriptive statistics

<i>Panel A: Univariate statistics for the sample</i>				
Variable	N	Mean	Std. Dev.	Median
<i>Readability</i>	679,775	-18.85	3.55	-18.81
<i>Earn (raw)</i>	670,220	-0.07	0.35	0.02
<i>Profit</i>	679,775	0.66	0.47	1
<i>Discretion</i>	679,775	0.50	0.35	0.50

<i>Panel B: Univariate statistics by Profit</i>				
<i>Profit = 0</i>				
Variable	N	Mean	Std. Dev.	Median
<i>Readability</i>	229,585	-19.22	3.51	-19.14
<i>Discretion</i>	229,585	0.50	0.35	0.50

<i>Profit = 1</i>				
Variable	N	Mean	Std. Dev.	Median
<i>Readability</i>	450,190	-18.67	3.56	-18.63
<i>Discretion</i>	450,190	0.50	0.35	0.50

This table presents descriptive statistics. Each firm-quarter is represented by exactly 5 observations, one for each quintile of MD&A sentences. Sentences are ranked into quintiles based on their commonness, with more common sentences having a lower rank. *Readability* is the readability of all sentences in a quintile and equal to -1 multiplied by the Gunning fog index. *Earn* is operating earnings over the prior twelve months scaled by total assets. In subsequent tables, *Earn* is a quintile ranking scaled between 0 and 1. *Profit* is an indicator set to 1 when quarterly net income is non-negative and to 0 when there is a loss. *Discretion* is a measure of the level of discretion in a quintile of MD&A sentences, and is a quintile ranking that ranges from 0 to 1 within each MD&A. See Appendix B for detailed variable definitions.

Table 2 -- Validating *Discretion*

<i>Panel A: Discretion and forward-looking statements</i>		
Variable	Prediction	Depvar. = <i>ForLook</i>
Intercept	?	4.608 *** (29.31)
<i>Discretion</i>	+	4.887 *** (32.97)
N		670,220
Adj. R-square		2.25%
<i>Panel B: Discretion and tone</i>		
Variable	Prediction	Depvar. = <i>Tone</i>
Intercept	?	-5.607 *** (-278.16)
<i>Discretion</i>	+	0.531 *** (32.10)
N		670,220
Adj. R-square		3.53%
<i>Panel C: The readability of less-discretionary text over time</i>		
Variable	Prediction	Depvar. = <i>Readability</i>
Intercept	?	-16.778 *** (-154.27)
<i>Year</i>	-	-0.083 *** (-10.61)
<i>Year * Discretion</i>	+	0.073 *** (9.38)
<i>Discretion</i>	?	-3.035 *** (-27.32)
N		670,220
Adj. R-square		5.32%

This table presents the results of validation tests for *Discretion*. Panel A (Panel B) [Panel C] presents the results of estimation Equation (2) (Equation (3)) [Equation (4)]. *Discretion* is the level of discretion in each quintile of sentences, scaled to range from 0 and 1. *Discretion* is calculated by grouping MD&As according to *Earn*. *ForLook* is the percent of sentences containing forward-looking statements and multiplied by 100 so that a unit increase corresponds to an increase of one percent. *Tone* is net positive tone based on the Loughran-McDonald (2011) dictionary multiplied by 100 and scaled to have standard deviation of 1. *Readability* is the the Gunning fog index multiplied by -1 so that higher values denote higher readability. *Year* is the calendar year of the fiscal quarter-end date. All variables are defined in Appendix B. T-statistics are presented in parentheses beneath coefficient estimates. Coefficients of interest are presented in bold type. Standard errors are clustered by firm and calendar quarter. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 3 -- The relation between profitability and the readability of discretionary text

<i>Panel A: Profitability = Earn</i>			
	(1)	(2)	(3)
Variable	Depvar. = Readability	Depvar. = Readability	Depvar. = Readability
<i>Earn</i>	0.652 *** (10.29)	0.578 *** (7.57)	0.427 *** (7.86)
<i>Earn * Discretion</i>	-0.264 *** (-4.03)	-0.229 *** (-2.66)	-0.114 (-1.64)
Industry * <i>Discretion</i> FE	Yes	Yes	No
Year * <i>Discretion</i> FE	Yes	Yes	No
Firm * <i>Discretion</i> FE	No	No	Yes
Controls	No	Yes	No
Controls * <i>Discretion</i>	No	Yes	No
N	670,220	395,790	670,220
Adj. R-square	12.20%	13.06%	26.62%
<i>Panel B: Profitability = Profit</i>			
	(1)	(2)	(3)
Variable	Depvar. = Readability	Depvar. = Readability	Depvar. = Readability
<i>Profit</i>	0.645 *** (16.76)	0.379 *** (8.58)	0.321 *** (11.67)
<i>Profit * Discretion</i>	-0.409 *** (-9.83)	-0.245 *** (-4.65)	-0.190 *** (-5.18)
Industry * <i>Discretion</i> FE	Yes	Yes	No
Year * <i>Discretion</i> FE	Yes	Yes	No
Firm * <i>Discretion</i> FE	No	No	Yes
Controls	No	Yes	No
Controls * <i>Discretion</i>	No	Yes	No
N	679,775	397,545	679,775
Adj. R-square	12.79%	13.41%	27.07%

This table presents the results of estimating Equation (1) with two proxies for *Profitability*. *Earn* (Panel A) is operating earnings over the prior twelve months scaled by total assets, ranked into quintiles, and scaled between 0 and 1. *Profit* (Panel B) is an indicator set to 1 when quarterly net income is non-negative and to 0 when there is a loss. All variables are defined in Appendix B. Control variables included in Column 2 of both panels are from Li (2008): an indicator for the presence of a special item, the market-to-book ratio, the number of geographical segments, the volatility of earnings, the volatility of returns, and an indicator for whether the firm engages in a seasoned equity offering. T-statistics are presented in parentheses beneath coefficient estimates. Coefficients of interest are presented in bold type. Standard errors are clustered by firm and calendar quarter. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 4 -- The readability of discretionary text for low-profitability firms specifically

Variable	(1) Depvar. = <i>Readability</i>	(2) Depvar. = <i>Readability</i>
<i>Low Earn</i>	-0.688 *** (-13.37)	-0.279 *** (-6.52)
<i>Low Earn * Discretion</i>	0.549 *** (9.66)	0.245 *** (4.36)
<i>High Earn</i>	0.037 (0.65)	0.094 ** (2.25)
<i>High Earn * Discretion</i>	0.243 *** (4.44)	0.063 (1.23)
Industry * <i>Discretion</i> FE	Yes	No
Year * <i>Discretion</i> FE	Yes	No
Firm * <i>Discretion</i> FE	No	Yes
N	670,220	670,220
Adj. R-square	12.25%	26.59%
<i>High Earn * Discretion</i> – <i>Low Earn * Discretion</i>	-0.306 ***	-0.182 ***

This table presents the results of estimating Equation (5). *Low Earn* (*High Earn*) is an indicator equal to 1 if *Earn* = 0 (*Earn* = 1), and to 0 otherwise, i.e., the lowest (highest) quintile of earnings. All variables are defined in Appendix B. T-statistics are presented in parentheses beneath coefficient estimates. Coefficients of interest are presented in bold type. Standard errors are clustered by firm and calendar quarter. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 5 -- Tone

Variable	(1)	(2)
	<i>Profitability = Earn</i> Depvar. = Tone	<i>Profitability = Profit</i> Depvar. = Tone
<i>Profitability</i>	0.077 *** (3.16)	0.070 *** (4.51)
<i>Profitability * Discretion</i>	0.063 *** (2.67)	0.027 * (1.78)
Industry * <i>Discretion</i> FE	Yes	Yes
Year * <i>Discretion</i> FE	Yes	Yes
N	670,220	679,775
Adj. R-square	7.61%	6.99%

This table presents the results of estimating Equation (1) but with *Tone* as the dependent variable. *Tone* is net positive tone based on the Loughran-McDonald (2011) dictionary multiplied by 100 and scaled to have standard deviation of 1. All variables are defined in Appendix B. T-statistics are presented in parentheses beneath coefficient estimates. Coefficients of interest are presented in bold type. Standard errors are clustered by firm and calendar quarter. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 6 -- Technical disclosure

<i>Panel A: Profitability = Earn</i>			
	(1)	(2)	(3)
Variable	<i>Depvar. = Attrib</i>	<i>Depvar. = ForLook</i>	<i>Depvar. = Risk</i>
<i>Earn</i>	2.776 *** (6.84)	-1.269 *** (-10.00)	-1.100 *** (-17.64)
<i>Earn * Discretion</i>	2.669 *** (5.36)	0.163 (0.86)	-0.094 (-0.99)
Industry * <i>Discretion</i> FE	Yes	Yes	Yes
Year * <i>Discretion</i> FE	Yes	Yes	Yes
N	670,220	670,220	670,220
Adj. R-square	6.26%	4.47%	3.69%
<i>Panel B: Profitability = Profit</i>			
	(1)	(2)	(3)
Variable	<i>Depvar. = Attrib</i>	<i>Depvar. = ForLook</i>	<i>Depvar. = Risk</i>
<i>Profit</i>	2.650 *** (10.67)	-1.326 *** (-16.23)	-0.943 *** (-19.29)
<i>Profit * Discretion</i>	1.778 *** (5.73)	0.005 (0.04)	-0.133 ** (-2.12)
Industry * <i>Discretion</i> FE	Yes	Yes	Yes
Year * <i>Discretion</i> FE	Yes	Yes	Yes
N	679,775	679,775	679,775
Adj. R-square	6.43%	4.54%	3.80%

This table presents the results of estimating Equation (1) with features of technical disclosure as the dependent variables. *Attrib*, *ForLook*, and *Risk* are the proportion of sentences containing attributions, forward-looking statements, and risk disclosures, respectively. The variables are multiplied by 100 so that a unit increase corresponds to an increase of one percent. All variables are defined in Appendix B. T-statistics are presented in parentheses beneath coefficient estimates. Coefficients of interest are presented in bold type. Standard errors are clustered by firm and calendar quarter. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 7 -- Profitability persistence

Variable	(1) <i>Profitability = Earn</i> Depvar. = Readability	(2) <i>Profitability = Profit</i> Depvar. = Readability
<i>Profitability * Discretion</i>	-0.101 (-1.24)	-0.232 *** (-3.73)
<i>Profitability * Discretion * Persist</i>	-0.271 *** (-2.71)	-0.297 *** (-3.87)
<i>Profitability</i>	0.357 *** (5.49)	0.275 *** (5.78)
<i>Profitability * Persist</i>	0.459 *** (5.26)	0.523 *** (8.41)
<i>Discretion * Persist</i>	0.335 *** (5.30)	0.274 *** (4.79)
<i>Persist</i>	-0.279 *** (-5.30)	-0.290 *** (-6.43)
Industry * <i>Discretion</i> FE	Yes	Yes
Year * <i>Discretion</i> FE	Yes	Yes
N	572,140	582,275
Adj. R-square	12.16%	12.76%

This table presents the results of estimating Equation (1) after interacting all variables with *Persist*. *Persist* is an indicator equal to 1 when *Profitability* in quarter t is equal to *Profitability* in quarter $t+4$. All variables are defined in Appendix B. T-statistics are presented in parentheses beneath coefficient estimates. Coefficients of interest are presented in bold type. Standard errors are clustered by firm and calendar quarter. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Table 8 -- Financial constraints

Variable	(1) <i>Profitability = Earn</i> Depvar. = Readability	(2) <i>Profitability = Profit</i> Depvar. = Readability
<i>Profitability * Discretion</i>	0.177 (1.40)	-0.367 *** (-5.07)
<i>Profitability * Discretion * Constrained</i>	-0.927 *** (-5.21)	-0.218 * (-1.95)
<i>Profitability</i>	0.367 *** (2.86)	0.544 *** (8.11)
<i>Profitability * Constrained</i>	0.749 *** (4.29)	0.373 *** (3.69)
<i>Discretion * Constrained</i>	-0.180 (-1.54)	-0.515 *** (-5.03)
<i>Constrained</i>	0.495 *** (4.19)	0.628 *** (6.35)
Industry * <i>Discretion</i> FE	Yes	Yes
Year * <i>Discretion</i> FE	Yes	Yes
N	619,360	625,835
Adj. R-square	12.15%	12.71%

This table presents the results of estimating Equation (1) after interacting all variables with *Constrained*. *Constrained* is a proxy for financial constraints, based on the SA Index by Hadlock and Pierce (2010). I rank the SA Index into quintiles and scale the ranking to range from 0 to 1. *Constrained* = 1 denotes the most financially constrained firms and *Constrained* = 0 denotes the least financially constrained firms. All variables are defined in Appendix B. T-statistics are presented in parentheses beneath coefficient estimates. Coefficients of interest are presented in bold type. Standard errors are clustered by firm and calendar quarter. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.