
Unverifiable net assets ratio and annual report reading difficulty

Wray Bradley*, Wen-Chyuan Chiang and
Li Sun

Collins College of Business,
The University of Tulsa,
Oklahoma, USA

Email: wray-bradley@utulsa.edu

Email: wen-chyuan-chiang@utulsa.edu

Email: li-sun@utulsa.edu

*Corresponding author

Abstract: The Financial Accounting Standards Board (FASB) has greatly expanded the use of fair value instruments, which amplifies an increase of unverifiable assets and liabilities that do not have actively traded market prices. Prior research indicates that managers may have strong incentives to use discretion in estimating fair values of such assets and liabilities for his/her self-serving interests, leading to more agency conflicts. This study examines the impact of unverifiable net assets on reading difficulty of annual reports. We predict that annual reports of firms with more unverifiable assets and liabilities are more difficult to read because prior research links agency conflicts to increased reading difficulty. Our results show a significant positive relation between unverifiable net assets and reading difficulty, consistent with the agency theory. Our study has implications for revenue management because prior research links earnings management to annual report readability.

Keywords: unverifiable assets and liabilities; fair value; readability; annual reports.

Reference to this paper should be made as follows: Bradley, W., Chiang, W-C. and Sun, L. (2021) 'Unverifiable net assets ratio and annual report reading difficulty', *Int. J. Revenue Management*, Vol. 12, Nos. 1/2, pp.104–131.

Biographical notes: Wray Bradley is a Full Professor of Accounting at the University of Tulsa. He teaches taxation, fraud detection and prevention, and forensic accounting. He has practiced as a CPA and tax attorney. He is also a Certified Management Accountant and is a former arbitrator for the American Arbitration Association.

Wen-Chyuan Chiang is the Doris L Flournoy Professor of Operations Management at The University of Tulsa. He received his PhD in Operations Management and his MBA from the University of Texas at Austin and his BA from the National Taiwan Normal University. His research interests include revenue management, supply chain management, distribution and logistics, manufacturing operations, and AI applications to operations management problems. His research appears in *Management Science*, *International Journal of Revenue Management*, *IIE Transactions*, *Decision Sciences*, *International*

Journal of Production Research, European Journal of Operational Research, Annals of Operations Research, Journal of the Operational Research Society and INFORMS Journal on Computing, among others. He has been a recipient of the Mayo Excellence in Research Award seven times. He is an editor of *International Journal of Revenue Management* and a senior editor of *Production and Operations Management*.

Li Sun is an Associate Professor of Accounting at the University of Tulsa. He teaches accounting information systems and cost accounting.

1 Introduction

After the Financial Accounting Standards Board (FASB) substantially expanded and accelerated the use of fair value instruments starting in the early 1990s, more and more assets and liabilities require managerial estimation in determining their market prices (i.e., fair values). Such assets and liabilities are known as unverifiable assets and liabilities in the context of financial accounting and it is relatively difficult to determine their fair values without managerial estimation. Unverifiable assets and liabilities consist of long-term assets, receivables, payables, and financial instruments that are less likely to have actively traded market prices. Prior studies such as Ramanna (2008) and Ramanna and Watts (2012) suggest that managers exploit discretion in estimating fair values of unverifiable assets and liabilities for their self-serving interests, consistent with the agency theory.

The purpose of this study is to examine the association between the level of unverifiable assets and liabilities and the readability of annual reports. Since 1998, the Securities Exchange Commission (SEC) has been persistently encouraging companies to improve the readability of their filings (SEC 1998; SEC 2013; SEC 2014). For example, in 1998 the SEC released a staff legal bulletin suggesting that written communications to investors be in plain English and easily understandable. Previous research on annual report readability suggests that:

- 1 firm operating performance (e.g., earnings), managerial incentives to obfuscate poor performance (e.g., earnings management), agency problems (e.g., corporate governance), and managerial characteristics (e.g., managerial ability) can influence the readability of annual reports and other documents
- 2 more readable documents including annual reports can bring benefits to a firm (e.g., favourable loan contracts).

Despite a growing body of research on the impact of readability, few studies including De Franco et al. (2015) and Hasan (2020) investigate factors that can influence the level of readability. Thus, this study should contribute to a more comprehensive understanding of factors that may influence the readability of documents by investigating the relation between unverifiable net assets and annual report readability.

Using a large panel sample of US firms consisting of more than 15,000 observations from 1993 to 2016, we document a significant and positive relation between unverifiable net assets ratio and reading difficulty of annual reports, which suggests that annual reports of firms with more unverifiable assets and liabilities are more difficult for average information users to read and understand. We perform various additional tests to check the robustness of our primary findings and to curtail concerns about possible endogeneity issues in our study, and still obtain consistent results supporting our hypothesis. In other words, our findings are robust to alternative readability measures, different sample time periods and empirical specifications, changes analysis, lagged independent variables, and the two-stage ordinary least squares regression analysis (2SLS). In addition, subsample comparison tests reveal that our primary results become stronger for firms in low technology industries or firms with higher bond credit ratings (i.e., investment grade ratings).

By documenting a significant relation between the level of unverifiable net assets and the readability of annual reports, our study makes several important contributions. First, our study contributes to the research on the readability of corporate filings and studies on fair value instruments in the accounting literature. In particular, few studies examine the impact of maintaining a certain level of unverifiable assets and liabilities on firm characteristics such as the readability of corporate filings. Second, our study directly examines the association between unverifiable net assets ratio and the readability of annual reports. Therefore, our study relates to the SEC's effort to promote more readable documents but also reveals that the level of corporate unverifiable assets and liabilities may reflect the level of the complexity of annual reports. The results imply that the SEC may use the level of unverifiable assets and liabilities to identify firms with less readable annual reports and other documents which can be a proxy for various red flags such as intent to obfuscate or mislead investors. Third, our results have practical implications to various stakeholder groups. For example, to average investors and analysts, they may need to devote more time and effort to understand annual reports of firms with a higher level of unverifiable assets and liabilities. Our study also has implications for revenue management because prior research links earnings management to annual report readability. In other words, it is documented that firms with a higher level of earnings management activities tend to make their annual reports harder to read and understand, implying reading difficulty is highly related to earnings management. Therefore, investors may pay attention to revenue management activities of firms with less readable annual reports because earnings and revenue are highly correlated.

The remainder of this paper is organised as follows. Section 2 reviews related studies on the readability of corporate filings (including annual reports) and develops our hypothesis. Section 3 presents the research design including empirical specifications, variable definitions, and sample characteristics. Section 4 reports our primary findings and Section 5 reports results of additional tests. Section 6 concludes our study.

2 Literature review and hypothesis development

2.1 Readability of annual reports and other documents

A growing number of studies have investigated the readability of corporate filings such as annual reports [Bonsall et al., (2017), p.331]. Studies on readability can be divided into studies on outcomes of readability and causes of readability with the former representing the bulk of the literature. The following review suggests a variety of outcomes based on the readability of reports and other documents. Bloomfield (2002) argues that easy-to-read documents can facilitate decision making. You and Zhang (2009) find that less readable annual reports can cause investors' under-reaction stronger. Lehavy et al. (2011) document that analysts have to devote more time and effort to understand less readable annual reports. Loughran and McDonald (2011) suggest that earnings volatility is positively associated with the complexity of annual reports, and such reports can result in more dispersed analyst forecasts. Lawrence (2013) argues that individual investors are less likely to invest in firms with less readable corporate filings. Bozanic and Thevenot (2015) find that complex annual reports (i.e., less readable annual reports) increase uncertainties in analyst reports. Bonsall and Miller (2017) find that firms with more readable 10-K reports receive higher bond credit ratings. Luo et al. (2018) suggest that poor document readability may reflect a higher level of agency problems or conflicts. Ertugrul et al. (2017) find that complex annual reports often result in stricter loan contract terms and greater stock price crash risks in the future. Guay et al. (2016) suggest that in order to mitigate any potential risks from complex annual reports, managers often provide voluntary disclosures to the public. Finally, Lo et al. (2017) investigate the relation between earnings management and readability, and find that the readability of the management discussion and analysis section of the 10-K reports (MD&A) decreases when firms engage in earnings management activities.

Relatively few studies have investigated the determinants of annual report readability. For example, Li (2008) finds that annual reports of firms with higher level of earnings and earnings persistence are easier to read and understand (i.e., more readable), suggesting that firms make their annual reports less readable when the managers seek to obfuscate poor corporate performance (i.e., lower earnings). De Franco et al. (2015) examine and find that less capable analysts produce less readable analyst reports. Hasan (2020) finds that managerial ability plays an important role in the readability of annual reports.

2.2 Hypothesis development

Prior research including Ramanna (2008) and Ramanna and Watts (2012) finds that managers tend to engage in more opportunistic discretion especially when their firms employ more unverifiable assets and liabilities. Because fair values (i.e., market prices) of unverifiable assets and liabilities are relatively difficult to estimate, managers often opportunistically use the discretion for their self-serving interests, consistent with the

agency theory. The FASB has significantly increased the use of fair value accounting and instruments such as goodwill impairment testing, derivatives, financial assets, and employee stock options since the 1990s. Such instruments often require managerial estimation of the current market prices. Hence, firms with more unverifiable assets and liabilities give their managers more opportunities to exercise discretion for his/her own benefits when estimating current market prices or fair values, potentially causing more agency conflicts. We rely on the agency theory to posit a positive relation between unverifiable net assets ratio and reading difficulty of annual reports (i.e., less readable annual reports) because recent research (e.g., Luo et al., 2018) documents a significant positive association between agency problems and reading difficulty. Together, we propose the following hypothesis on the association between unverifiable net assets ratio and annual report readability.

H1 The annual reports of firms with greater unverifiable net assets ratio are more difficult to read and understand.

3 Research design

3.1 *Measuring unverifiable net assets ratio*

Following prior research (e.g., Richardson et al., 2005; Ramanna, 2008; Ramanna and Watts, 2012), we first calculate the following ratio to measure the percentage of assets and liabilities that have easily verifiable market prices.

$$\text{Verifiable Net Assets} = \frac{\text{Cash} + \text{Short-term Investment} + \text{Long-term Investment} - \text{Long-term Debt} - \text{Debt in Current Debt} - \text{Preferred Stock}}{\text{Total Assets} - \text{Total Liabilities}}$$

In the above equation, items excluded from the numerator are those that carry fair value estimates are less verifiable (e.g., relatively difficult for managers to estimate the current market value). Therefore, such items include fixed assets, receivables, payables, and inventories. In concept, they are less likely to have actively traded market prices, requiring more subjectivity in estimating their fair values. Thus, as this ratio increases (decreases), subjectivity in estimating fair values of net assets decreases (increases).

To construct a variable of unverifiable net assets that increase in subjectivity, we then multiply the ratio by -1 and use the result as the unverifiable net assets (UNVA) in our study. Moreover, we use the percentile rank of UNVA, namely UNVA_RANK, in our empirical analysis, following Ramanna and Watts (2012). A high value of UNVA_RANK reflects a high level of subjectivity of managers in estimating the market prices of the unverifiable assets and liabilities. In other words, managers of firms with a high value of UNVA_RANK need to use more discretion when estimating fair values.

3.2 Measuring reading difficulty

There are several measures to capture the readability of documents. Prior research (e.g., Li, 2008; Loughran and McDonald, 2014) mainly uses the Fog Index and the total file size to measure the readability of narrative disclosures in annual reports. A recent study, Bonsall et al. (2017), introduces a more comprehensive readability measure known as the

Bog Index, which evaluates the readability of narrative disclosures from more perspectives such as passive voice and sentence complexity, relative to the Fog Index and the file size. Recent studies (e.g., Hasan, 2020) conclude that the Bog Index is a good readability measure. Hence, we use the Bog Index in our study as our primary readability measure.

The Bog Index is multifaceted and is calculated by an editing software called *Stylewriter*. Specifically, the formula to calculate the Bog Index is as follows:

$$\text{Bog Index} = \text{Sentence Bog} + \text{Word Bog} - \text{Pep} \quad (1)$$

According to Bonsall et al. (2017), the first component (Sentence Bog) captures the readability of sentence length. The second component (Word Bog) captures the readability of plain English problems (passive verbs, hidden verbs, overwriting, legal terms, clichés, abstract words, and wordy phrases) and word difficulty (heavy words, abbreviations, and specialist terms)¹. The third component (Pep) captures readability attributes (good writing) that can help readers understand. A high value of the Bog Index indicates that the narrative disclosures are difficult to read and understand (i.e., less readable). Thus, a low value of the Bog Index would indicate a higher level of readability, which is preferred. In order to aid in interpretation, Bonsall et al. (2017) breakdown the values of the Bog Index as follows: 1,000 and above = Gobbledygook (not readable); 131 – 1,000 = dreadful; 101 – 130 = bad; 71 – 100 = poor; 41 – 70 = average; 21 – 40 = good; 0 – 20 = excellent (most readable).

3.3 Empirical specification

To explore the association between unverifiable net assets ratio and the reading difficulty of annual reports, we use the following equation:

$$\begin{aligned} \text{DIFFICULTY} = & \alpha_0 + \alpha_1 \text{UNVA_RANK} + \alpha_2 \text{SIZE} + \alpha_3 \text{LEV} + \alpha_4 \text{ROA} \\ & + \alpha_5 \text{MTB} + \alpha_6 \text{CFO} + \alpha_7 \text{CAPINT} + \alpha_8 \text{FIRMAGE} \\ & + \alpha_9 \text{BONDRATING} + \alpha_{10} \text{MABILITY} + \alpha_{11} \text{VOLATILITY} \\ & + \alpha_{12} \text{SPECIAL} + \alpha_{13} \text{MERGE} + \alpha_{14} \text{EARNINGS} \\ & + \alpha_{15} \text{ACCRUALS} + \alpha_{16} \text{LOSS} + \alpha_{17} \text{SEGMENT} + \alpha_{18} \text{BIG4} \\ & + \text{Year Indicators} + \text{Industry Indicators} + \varepsilon \end{aligned} \quad (2)$$

DIFFICULTY is the Bog Index of narrative disclosures in annual reports, reflecting reading difficulty. UNVA_RANK shows the level of unverifiable assets and liabilities. If the annual reports of firms with more unverifiable assets and liabilities are more difficult to read and understand, we expect a significant and positive coefficient on UNVA_RANK (α_1).

Following recent studies on readability, we control for other factors that may influence the readability of documents. Specifically, based on Lo et al. (2017) and Hasan (2020), we control for total firm assets (SIZE), leverage ratio (LEV), firm profitability (ROA), firm growth or market to book ratio (MTB), capital intensity (CAPINT), the age of a firm in the CompuStat database (FIRMAGE), managerial ability (MABILITY), the volatility of sales (VOLALITY), special items (SPECIAL), whether a firm-year observation engages in merger & acquisitions activities (MERGE), earnings (EARNINGS), discretionary accruals (ACCRUALS), whether a firm-year observation experiences a loss (LOSS), and the number of business segments (SEGMENT). For example, Lo et al. (2017) find that annual reports of firms with a higher level of earnings management activities are less readable. Hence, we use the absolute values of discretionary accruals to capture the level of earnings management activities. We follow Kothari et al. (2005) to calculate discretionary accruals. We control for bond credit ratings (BONDRATING) in equation (2) because Bonsall and Miller (2017) document a significant association between bond ratings and readability. We also control for whether a firm use a Bug 4 accounting firm as their auditor. Lastly, we include industry and year dummy variables in equation (2). All continuous variables in our baseline regression model are winsorised at the 1% and 99% levels. Following Petersen (2009), we employ the clustered (by year and firm) standard errors OLS regression as our primary regression technique.

3.4 Sample distribution and descriptive statistics

Our initial sample is an intersection of reading difficulty data (BOG index) from Professor Bonsall's website and financial statement data from the Compustat database². After removing observations with insufficient data to construct the primary independent variable and control variables in Equation 2, our final sample consists of 15,229 firm year observations representing 1,921 unique publicly traded US firms from 1993 to 2016³. Panel A of Table 1a presents the sample distribution by year. For instance, there are 394 observations in 1993, 715 observations in 2000, 613 observations in 2010, and 542 observations in 2016. From 1995 to 2016, the numbers of observations are in the range of 540 and 770. Panel B of Table 1b reports the sample distribution by industry based on the Fama and French 48 industry classifications. For example, there are 62 observations in agriculture industries (FF48 = 1), and 777 observations in transportation industries (FF48 = 40). In our sample, the most represented industry is oil, petroleum and gas (FF48 = 30) with 1,290 observations, followed by retail (FF48 = 42) with 1,234 observations and telecommunications (FF48 = 32) with 965 observations. Panel C of Table 1c presents the industry distribution by the value of reading difficulty

(DIFFICULTY) in column 1 and by the value of unverifiable net assets ratio (UNVA_RANK) in column 2. The industry with the most difficulty-to-read annual reports is defence (FF48 = 26), while the industry with the highest unverifiable net assets ratio is pharmaceutical products (FF48 = 13).

Table 1a Unverifiable net assets ratio and annual report reading difficulty: sample distribution by year

<i>Panel A</i>			
<i>Year</i>	<i>Number of observations</i>	<i>Percentage</i>	<i>Cumulative percent</i>
1993	394	2.59%	2.59%
1994	334	2.19%	4.78%
1995	563	3.70%	8.48%
1996	667	4.38%	12.86%
1997	756	4.96%	17.82%
1998	765	5.02%	22.84%
1999	699	4.59%	27.43%
2000	715	4.69%	32.13%
2001	710	4.66%	36.79%
2002	701	4.60%	41.39%
2003	718	4.71%	46.11%
2004	733	4.81%	50.92%
2005	684	4.49%	55.41%
2006	653	4.29%	59.70%
2007	636	4.18%	63.88%
2008	623	4.09%	67.97%
2009	608	3.99%	71.96%
2010	613	4.03%	75.99%
2011	613	4.03%	80.01%
2012	608	3.99%	84.00%
2013	621	4.08%	88.08%
2014	661	4.34%	92.42%
2015	612	4.02%	96.44%
2016	542	3.56%	100.00%
	15,229	100.00%	

Note: This panel reports the sample distribution by fiscal year for the complete sample, which consists of 15,229 firm-year observations from 1993 to 2016. There are 1,921 unique US public companies in our sample.

Table 1b Unverifiable net assets ratio and annual report reading difficulty: sample distribution by industry

<i>Panel B</i>				
<i>Industry</i>	<i>Description (FF48)</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative percent</i>
1	Agriculture	62	0.41%	0.41%
2	Food products	377	2.48%	2.88%
3	Candy and soda	85	0.56%	3.44%
4	Beer and liquor	76	0.50%	3.94%
5	Tobacco products	54	0.35%	4.29%
6	Recreation	71	0.47%	4.76%
7	Entertainment	366	2.40%	7.16%
8	Printing and publishing	155	1.02%	8.18%
9	Household and consumer goods	285	1.87%	10.05%
10	Clothing and apparel	27	0.18%	10.23%
11	Healthcare	357	2.34%	12.57%
12	Medical equipment	270	1.77%	14.35%
13	Pharmaceutical products	487	3.20%	17.55%
14	Chemicals	743	4.88%	22.42%
15	Rubber and plastic products	138	0.91%	23.33%
16	Textiles	114	0.75%	24.08%
17	Construction materials	438	2.88%	26.96%
18	Construction	203	1.33%	28.29%
19	Steel works	399	2.62%	30.91%
20	Fabricated products	46	0.30%	31.21%
21	Machinery	746	4.90%	36.11%
22	Electrical equipment	146	0.96%	37.07%
23	Automobiles and trucks	312	2.05%	39.12%
24	Aircraft	244	1.60%	40.72%
25	Shipbuilding and railroad	75	0.49%	41.21%
26	Defence	83	0.55%	41.76%
27	Precious metals	42	0.28%	42.03%
28	Mines and non-metal	81	0.53%	42.56%
29	Coal	28	0.18%	42.75%
30	Oil, petroleum and gas	1,290	8.47%	51.22%
31	Utilities	67	0.44%	51.66%
32	Telecommunications	965	6.34%	57.99%
33	Personal services	205	1.35%	59.34%
34	Business services	959	6.30%	65.64%

Note: This panel presents the sample distribution by industry (based on Fama and French 48 industry classifications). The sample consists of 15,229 observations from 1993 to 2016, representing 1,921 unique public companies in the USA.

Table 1b Unverifiable net assets ratio and annual report reading difficulty: sample distribution by industry (continued)

<i>Panel B</i>				
<i>Industry</i>	<i>Description (FF48)</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative percent</i>
35	Computers	427	2.80%	68.44%
36	Chips and electronic equipment	730	4.79%	73.24%
37	Measuring and control	250	1.64%	74.88%
38	Business supplies	446	2.93%	77.81%
39	Boxes and shipping containers	144	0.95%	78.75%
40	Transportation	777	5.10%	83.85%
41	Wholesale	634	4.16%	88.02%
42	Retail	1,234	8.10%	96.12%
43	Meals, restaurants and hotels	272	1.79%	97.91%
44	Banking	23	0.15%	98.06%
45	Insurance	16	0.11%	98.16%
46	Real estate	23	0.15%	98.31%
47	Trading	69	0.45%	98.77%
48	Other	188	1.23%	100.00%

Note: This panel presents the sample distribution by industry (based on Fama and French 48 industry classifications). The sample consists of 15,229 observations from 1993 to 2016, representing 1,921 unique public companies in the USA.

Table 1c Unverifiable net assets ratio and annual report reading difficulty: sample distribution by values of DIFFICULTY and UNVA_RANK

<i>Panel C</i>					
<i>Column 1</i>			<i>Column 2</i>		
<i>Based on Bog index</i>			<i>Based on UNVA_RANK</i>		
<i>DIFFICULTY</i>	<i>Industry</i>	<i>Description (FF48)</i>	<i>UNVA_RANK</i>	<i>Industry</i>	<i>Description (FF48)</i>
89.897	26	Defense	49.721	13	Pharmaceutical products
89.359	12	Medical equipment	49.451	36	Chips and electronic equipment
88.673	36	Chips and electronic equipment	49.382	35	Computers
88.473	11	Healthcare	44.010	34	Business services
88.281	13	Pharmaceutical products	42.512	37	Measuring and control
87.396	37	Measuring and control	41.432	22	Electrical equipment
86.714	35	Computers	40.593	5	Tobacco products

Note: This panel presents the sample distribution by values of DIFFICULTY and UNVA_RANK. The sample consists of 15,229 observations from 1993 to 2016, representing 1,921 unique public companies in the USA.

Table 1c Unverifiable net assets ratio and annual report reading difficulty: sample distribution by values of DIFFICULTY and UNVA_RANK (continued)

<i>Panel C</i>					
<i>Column 1</i>			<i>Column 2</i>		
<i>Based on Bog index</i>			<i>Based on UNVA_RANK</i>		
<i>DIFFICULTY</i>	<i>Industry</i>	<i>Description (FF48)</i>	<i>UNVA_RANK</i>	<i>Industry</i>	<i>Description (FF48)</i>
85.791	14	Chemicals	38.822	12	Medical equipment
85.777	34	Business services	38.268	33	Personal services
85.216	32	Telecommunications	36.738	27	Precious metals
85.008	48	Other	36.704	6	Recreation
84.457	20	Fabricated products	36.519	9	Household and consumer goods
84.335	47	Trading	36.458	26	Defense
84.309	28	Mines and non-metal	35.993	21	Machinery
84.199	22	Electrical equipment	35.669	45	Insurance
84.001	45	Insurance	34.890	8	Printing and publishing
83.992	24	Aircraft	34.850	32	Telecommunications
83.414	30	Oil, petroleum and gas	34.714	48	Other
83.374	41	Wholesale	34.124	47	Trading
83.121	44	Banking	34.040	10	Clothing and apparel
83.013	21	Machinery	34.001	44	Banking
82.856	46	Real estate	33.179	42	Retail
82.821	29	Coal	32.891	20	Fabricated products
82.463	33	Personal services	32.888	46	Real estate
82.434	40	Transportation	32.766	24	Aircraft
82.163	23	Automobiles and trucks	32.753	17	Construction materials
82.111	31	Utilities	32.525	31	Utilities
82.008	7	Entertainment	32.069	40	Transportation
82.000	15	Rubber and plastic products	31.667	25	Shipbuilding and railroad
81.972	19	Steel works	31.138	19	Steel works
81.944	5	Tobacco products	30.469	28	Mines and non-metal
81.747	25	Shipbuilding and railroad	29.721	43	Meals, restaurants and hotels
81.461	17	Construction materials	29.256	14	Chemicals

Note: This panel presents the sample distribution by values of DIFFICULTY and UNVA_RANK. The sample consists of 15,229 observations from 1993 to 2016, representing 1,921 unique public companies in the USA.

Table 1c Unverifiable net assets ratio and annual report reading difficulty: sample distribution by values of DIFFICULTY and UNVA_RANK (continued)

<i>Panel C</i>					
<i>Column 1</i>			<i>Column 2</i>		
<i>Based on Bog index</i>			<i>Based on UNVA_RANK</i>		
<i>DIFFICULTY</i>	<i>Industry</i>	<i>Description (FF48)</i>	<i>UNVA_RANK</i>	<i>Industry</i>	<i>Description (FF48)</i>
80.438	10	Clothing and apparel	28.365	30	Oil, petroleum and gas
80.315	18	Construction	28.063	41	Wholesale
79.736	39	Boxes and shipping containers	27.901	18	Construction
79.387	1	Agriculture	27.803	4	Beer and liquor
79.208	42	Retail	26.452	23	Automobiles and trucks
79.071	27	Precious metals	26.422	39	Boxes and shipping containers
78.971	38	Business supplies	25.714	2	Food products
78.746	43	Meals, restaurants and hotels	25.457	15	Rubber and plastic products
78.684	16	Textiles	25.250	29	Coal
78.254	6	Recreation	25.216	7	Entertainment
77.674	9	Household and consumer goods	24.645	1	Agriculture
77.013	4	Beer and liquor	23.210	11	Healthcare
76.918	3	Candy and soda	21.667	16	Textiles
76.775	2	Food products	19.125	38	Business supplies
75.929	8	Printing and publishing	15.588	3	Candy and soda

Note: This panel presents the sample distribution by values of DIFFICULTY and UNVA_RANK. The sample consists of 15,229 observations from 1993 to 2016, representing 1,921 unique public companies in the USA.

Table 2 presents the summary of sample descriptive statistics. The mean and median values of DIFFICULTY are 83.121 and 83.000, respectively. Based on Bonsall and Miller (2017), these numbers suggest that on average the readability of annual reports for our sample firms is poor. In addition, these numbers are very close to those in Hasan (2020), who reports the mean and median values of the Bog Index of 82.849 and 83.000, respectively. The mean values of leverage ratio (LEV), firm profitability (ROA), and firm growth (MTB) are 0.316, 0.027, and 2.826, respectively, which suggests that the overall operating performance of our sample firms seems normal and is in line with general expectations. The mean value of ACCRUALS is 0.046. The mean value of LOSS is 0.227, suggesting that on average approximately 22% of total observations experience a loss in a given year. The mean value of BIG4 is 0.968, which suggests that the majority of our sample firms use Big4 accounting firms as their auditors.

Table 2 Unverifiable net assets ratio and annual report reading difficulty sample descriptive statistics

<i>Variable</i>	<i>Obs.</i>	<i>Mean</i>	<i>Std Dev</i>	<i>25th Pctl</i>	<i>50th Pctl</i>	<i>75th Pctl</i>
DIFFICULTY	15,229	83.121	7.462	78.000	83.000	88.000
UNVA_RANK	15,229	33.900	23.820	17.000	30.000	45.000
SIZE	15,229	7.961	1.357	6.985	7.843	8.862
LEV	15,229	0.316	0.203	0.176	0.280	0.416
ROA	15,229	0.027	0.095	0.007	0.040	0.072
MTB	15,229	2.826	4.982	1.293	2.172	3.511
CFO	15,229	0.093	0.069	0.053	0.090	0.131
CAPINT	15,229	0.062	0.063	0.023	0.042	0.075
FIRMAGE	15,229	3.207	0.650	2.639	3.258	3.829
BONDRATING	15,229	12.356	3.456	10.000	12.000	15.000
MABILITY	15,229	0.538	0.307	0.300	0.500	0.800
VOLATILITY	15,229	0.728	0.575	0.327	0.571	0.944
SPECIAL	15,229	-0.016	0.049	-0.015	-0.003	0.000
MERGE	15,229	0.522	0.500	0.000	1.000	1.000
EARNINGS	15,229	0.015	0.220	0.005	0.041	0.084
ACCRUALS	15,229	0.046	0.045	0.015	0.034	0.062
LOSS	15,229	0.227	0.419	0.000	0.000	0.000
SEGMENT	15,229	1.987	1.106	1.099	1.946	2.565
BIG4	15,229	0.968	0.176	1.000	1.000	1.000

Note: This table presents the number of observations, pooled means, standard deviations, 25th percentile, median, and 75th percentile of the dependent variables, independent variable of interest, and control variables. Appendix 1 provides detailed variable definitions in this table.

3.5 Correlations

Table 3 reports the Pearson (below the diagonal) and Spearman (above the diagonal) correlations among key variables used in equation (2). As shown in Table 3, both correlation matrixes show a significant and positive correlation between UNVA_RANK and DIFFICULTY. For example, the correlation coefficient is 0.133 with a p-value of less than 0.0001 in Pearson matrix. This evidence suggests that it is likely that annual reports of firms with more unverifiable assets and liabilities are more difficult to understand or less readable, providing initial support to our hypothesis. Many correlation coefficients in Table 3 are reasonably small, suggesting that multicollinearity may not be a major concern. In addition, many pairs of associations are statistically significant, highlighting the importance of testing our hypothesis in a multivariate setting.

Table 3 Unverifiable net assets ratio and annual report reading difficulty (correlation matrix)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
DIFFICULTY	1	0.119 <.0001	0.125 <.0001	0.047 <.0001	-0.123 <.0001	-0.016 <.0001	-0.097 <.0001	-0.224 <.0001	-0.074 <.0001	-0.191 <.0001	-0.045 <.0001	0.059 <.0001	-0.172 <.0001	0.134 <.0001	-0.025 <.0001	0.078 <.0001	0.115 <.0001	0.144 <.0001	-0.039 <.0001
UNVA_RANK	2	0.133 <.0001	1	0.115 <.0001	-0.554 <.0001	0.264 <.0001	-0.075 <.0001	0.212 <.0001	-0.059 <.0001	0.086 <.0001	0.208 <.0001	-0.033 <.0001	0.022 <.0001	0.004 <.0001	0.216 <.0001	0.103 <.0001	-0.110 <.0001	0.059 <.0001	0.044 <.0001
SIZE	3	0.115 <.0001	0.054 <.0001	1	-0.302 <.0001	0.232 <.0001	0.229 <.0001	0.154 <.0001	-0.022 <.0001	0.419 <.0001	0.581 <.0001	-0.133 <.0001	0.009 <.0001	0.137 <.0001	0.278 <.0001	-0.098 <.0001	-0.213 <.0001	0.287 <.0001	0.124 <.0001
LEV	4	0.066 <.0001	-0.166 <.0001	-0.313 <.0001	1	-0.394 <.0001	-0.206 <.0001	-0.245 <.0001	0.041 <.0001	-0.287 <.0001	-0.511 <.0001	0.116 <.0001	-0.065 <.0001	-0.082 <.0001	-0.286 <.0001	-0.021 <.0001	0.272 <.0001	-0.157 <.0001	-0.102 <.0001
ROA	5	-0.125 <.0001	0.013 <.0001	0.223 <.0001	-0.307 <.0001	1	0.458 <.0001	0.574 <.0001	0.060 <.0001	0.196 <.0001	0.535 <.0001	-0.099 <.0001	0.357 <.0001	0.073 <.0001	0.827 <.0001	0.001 <.0001	0.699 <.0001	0.072 <.0001	0.095 <.0001
MTB	6	-0.022 <.0001	-0.250 <.0001	0.103 <.0001	0.111 <.0001	0.176 <.0001	1	0.342 <.0001	0.037 <.0001	0.127 <.0001	0.398 <.0001	-0.067 <.0001	0.047 <.0001	0.079 <.0001	0.395 <.0001	0.058 <.0001	-0.277 <.0001	0.028 <.0001	0.067 <.0001
CFO	7	-0.105 <.0001	0.137 <.0001	0.156 <.0001	-0.198 <.0001	0.444 <.0001	0.177 <.0001	1	0.313 <.0001	0.083 <.0001	0.376 <.0001	-0.092 <.0001	0.111 <.0001	-0.040 <.0001	0.482 <.0001	0.144 <.0001	-0.361 <.0001	0.042 <.0001	0.054 <.0001
CAPINT	8	-0.136 <.0001	-0.079 <.0001	-0.057 <.0001	0.104 <.0001	-0.075 <.0001	-0.066 <.0001	0.287 <.0001	1	-0.053 <.0001	0.066 <.0001	0.040 <.0001	0.138 <.0001	-0.205 <.0001	0.030 <.0001	-0.015 <.0001	-0.017 <.0001	-0.154 <.0001	0.019 <.0001
FIRMGAGE	9	-0.091 <.0001	0.016 <.0001	0.412 <.0001	-0.295 <.0001	0.188 <.0001	0.067 <.0001	0.083 <.0001	-0.128 <.0001	1	0.444 <.0001	-0.323 <.0001	0.000 <.0001	0.050 <.0001	0.147 <.0001	-0.136 <.0001	-0.173 <.0001	0.511 <.0001	0.085 <.0001
BONDRATING	10	-0.219 <.0001	0.082 <.0001	0.582 <.0001	-0.495 <.0001	0.451 <.0001	0.186 <.0001	0.352 <.0001	-0.057 <.0001	0.440 <.0001	1	-0.232 <.0001	0.117 <.0001	0.170 <.0001	0.474 <.0001	-0.076 <.0001	-0.433 <.0001	0.233 <.0001	0.160 <.0001
MABILITY	11	-0.057 <.0001	0.145 <.0001	0.265 <.0001	-0.202 <.0001	0.266 <.0001	0.106 <.0001	0.255 <.0001	0.061 <.0001	0.168 <.0001	0.321 <.0001	-0.056 <.0001	0.063 <.0001	-0.007 <.0001	0.258 <.0001	0.091 <.0001	-0.203 <.0001	0.060 <.0001	0.048 <.0001
VOLATILITY	12	0.073 <.0001	-0.021 <.0001	-0.137 <.0001	0.123 <.0001	-0.131 <.0001	-0.013 <.0001	-0.124 <.0001	0.075 <.0001	-0.333 <.0001	-0.241 <.0001	-0.069 <.0001	0.013 <.0001	0.013 <.0001	-0.108 <.0001	0.055 <.0001	0.090 <.0001	-0.147 <.0001	-0.063 <.0001
SPECIAL	13	-0.091 <.0001	-0.050 <.0001	0.090 <.0001	-0.064 <.0001	0.618 <.0001	0.034 <.0001	0.121 <.0001	0.056 <.0001	0.064 <.0001	0.144 <.0001	-0.034 <.0001	0.105 <.0001	0.106 <.0001	-0.069 <.0001	0.317 <.0001	-0.090 <.0001	-0.336 <.0001	0.027 <.0001
MERGE	14	0.132 <.0001	-0.031 <.0001	0.139 <.0001	-0.094 <.0001	0.078 <.0001	0.016 <.0001	-0.043 <.0001	-0.224 <.0001	0.047 <.0001	0.166 <.0001	-0.007 <.0001	0.014 <.0001	-0.010 <.0001	0.082 <.0001	0.016 <.0001	-0.114 <.0001	0.158 <.0001	0.001 <.0001
EARNINGS	15	-0.076 <.0001	-0.027 <.0001	0.172 <.0001	-0.229 <.0001	0.793 <.0001	0.104 <.0001	0.300 <.0001	-0.105 <.0001	0.157 <.0001	0.328 <.0001	-0.142 <.0001	0.415 <.0001	0.069 <.0001	<.0001	<.0001	-0.010 <.0001	-0.725 <.0001	0.068 <.0001
ACCRUALS	16	0.066 <.0001	0.138 <.0001	-0.111 <.0001	0.015 <.0001	-0.097 <.0001	0.042 <.0001	0.138 <.0001	0.060 <.0001	-0.153 <.0001	-0.090 <.0001	0.104 <.0001	0.082 <.0001	-0.182 <.0001	-0.078 <.0001	0.199 <.0001	0.087 <.0001	-0.083 <.0001	-0.032 <.0001
LOSS	17	0.122 <.0001	-0.015 <.0001	-0.214 <.0001	0.297 <.0001	-0.649 <.0001	-0.123 <.0001	-0.351 <.0001	0.032 <.0001	-0.175 <.0001	-0.420 <.0001	-0.202 <.0001	0.113 <.0001	-0.406 <.0001	-0.520 <.0001	0.107 <.0001	<.0001	-0.092 <.0001	-0.086 <.0001
SEGMENT	18	<.0001	0.062 <.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
SEGMENT	18	0.138 <.0001	0.011 <.0001	0.276 <.0001	-0.144 <.0001	0.076 <.0001	-0.009 <.0001	-0.038 <.0001	-0.174 <.0001	0.432 <.0001	0.193 <.0001	0.056 <.0001	-0.125 <.0001	0.011 <.0001	0.141 <.0001	0.073 <.0001	-0.096 <.0001	-0.087 <.0001	0.013 <.0001
BIG4	19	-0.040 <.0001	0.024 <.0001	0.123 <.0001	-0.104 <.0001	0.079 <.0001	0.245 <.0001	-0.047 <.0001	-0.027 <.0001	0.084 <.0001	0.153 <.0001	0.049 <.0001	-0.058 <.0001	0.011 <.0001	0.068 <.0001	-0.030 <.0001	-0.086 <.0001	0.001 <.0001	0.001 <.0001
BIG4	19	<.0001	0.003 <.0001	<.0001	<.0001	<.0001	0.000 <.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.164 <.0001	0.939 <.0001	<.0001	0.000 <.0001	<.0001	0.892 <.0001	

Note: This table reports the Pearson correlations (below the diagonal) and the Spearman correlations (above the diagonal) of the variables for the full sample. Correlation coefficients and related (two-tailed) p-values are provided for each pair of variables. P-values are italicized. Appendix 1 provides detailed variable definitions.

4 Primary results

To test our hypothesis, we use the clustered standard errors OLS regression to estimate equation (2) and report results in Table 4. Specifically, we only include UNVA_RANK in column 1, and UNVA_RANK and industry/year indicators in Column 2. We include the entire equation (2) in column 3. in column 1 (column 2) of Table 4, the coefficient on UNVA_RANK is 0.042 (0.008) with a t-value of 17.71 (3.96), indicating a significant positive relation between unverifiable net assets ratio and the complexity of annual reports. More importantly, the coefficient on UNVA_RANK in column 3 is 0.018 with a t-value of 8.08, showing that UNVA_RANK is positively and significantly related to DIFFICULTY. This finding suggests that annual reports of firms with more unverifiable assets and liabilities are more difficult for readers to read and understand. Thus, empirical results strongly support our hypothesis.

Column 3 reports that DIFFICULTY is positively related to SIZE, LEV, MTB, VOLATILITY, MERGE, EARNINGS, ACCRUALS, LOSS, SEGMENT, and BIG4. In addition, the coefficients on ROA, CFO, CAPINT, FIRMAGE, and BONDRATING are significant negative. Several of the above findings are in line with general expectations. For instance, the positive relation between DIFFICULTY and SEGMENT suggests that annual reports of firms with more business segments are harder to read and understand because these firms have more diversified operations. The negative relation between DIFFICULTY and BONDRATING is consistent with Bonsall and Miller (2017), which finds that firms with easier to read annual reports receive higher bond ratings. The positive DIFFICULTY and ACCRUALS suggests that annual reports of firms engaging in more earnings management activities are less readable, consistent with the findings in Lo et al. (2017).

Our results also appear economically meaningful. Column 1 reports that the adjusted R squared is 0.1778, which suggests that approximately 18% of the variation in DIFFICULTY can be explained by the variation in UNVA_RANK alone. Similarly, the value of the adjusted R squared in Column 3 shows that 44% of the variation in DIFFICULTY can be explained by all the variables in equation (2). In addition, based on the numbers in Column 3 of Table 4, a one standard deviation increase in UNVA_RANK is associated with an increase of 5.74% in DIFFICULTY⁴. We also check the values of variance inflation factor (VIF) and find that these values are relatively small, suggesting that multicollinearity should not be a major concern in our study.

Table 4 Unverifiable net assets ratio and annual report reading difficulty (primary results)

	<i>Dependent variable = DIFFICULTY</i>						
	<i>Column 1</i>		<i>Column 2</i>		<i>Column 3</i>		
	<i>Estimate</i>	<i>t value</i>	<i>Estimate</i>	<i>t value</i>	<i>Estimate</i>	<i>t value</i>	<i>VIF</i>
Intercept	81.705***	795.57	86.757***	224.00	84.930***	139.04	
UNVA_RANK	0.042***	17.71	0.008***	3.96	0.018***	8.08	1.37
SIZE					0.747***	14.72	2.32
LEV					0.539*	1.89	1.63
ROA					-2.983**	-2.57	5.96
MTB					0.028***	2.76	1.24
CFO					-3.730***	-4.05	1.95
CAPINT					-4.129***	-4.25	1.84
FIRMAGE					-1.016***	-10.92	1.79
BONDRATING					-0.404***	-16.79	3.39
MABILITY					0.102	0.60	1.32
VOLATILITY					1.060***	11.94	1.27
SPECIAL					0.880	0.67	2.09
MERGE					1.046***	10.55	1.20
EARNINGS					0.970***	2.74	2.97
ACCRUALS					2.182**	1.98	1.23
LOSS					0.802***	5.41	1.88
SEGMENT					0.633***	12.92	1.44
BIG4					0.749***	2.81	1.07
Industry Indicators	No		Yes		Yes		
Year Indicators	No		Yes		Yes		
Observations	15,229		15,229		15,229		
Adjusted R ²	0.1778		0.3805		0.4438		

Notes: This table presents the results of the baseline regression model using clustered standard errors OLS regression. The baseline model is as follows. $DIFFICULTY = \alpha_0 + \alpha_1 \times UNVA_RANK + \alpha_x \times \text{Control Variables} + \text{Industry Indicators} + \text{Year Indicators} + \varepsilon$. The continuous variables in the baseline regression model are winsorised at the 1% and 99% percentiles each year before entering the regression analysis. *, **, and *** denote significance at the 10, 5 and 1 percent (two-tailed) confidence levels, respectively. Appendix 1 provides detailed variable definitions for variables used in this table.

5 Additional tests

5.1 *Alternative readability measures*

Loughran and McDonald (2014) use the total file size of a document to indicate its readability, illustrating that the file size and the readability of a document is highly correlated. In addition, prior research (e.g., Li, 2008) uses the FOG Index to capture the readability of documents. The Fog Index is calculated using the following formula:

$$\text{Fog Index} = (\text{Average number of words in sentences} + \text{Percentage of words of three or more syllables}) \times 0.4 \quad (3)$$

This measure reflects the number of years of education for an average person to understand a document. For example, a Fog Index score of 17 means the text requires an education of 17 years, which suggests the reader needs to have a college degree. The relation between the Fog and reading ease is as follows: FOG > 18 means the text is unreadable; 14–18 means the text is difficult; 12–14 means the text is ideal; 10–12 means the text is acceptable; and 8–10 means childish. Therefore, a higher value of the Fog Index indicates more years of education needed to understand a document, suggesting poor readability.

Recent studies such as Bonsall and Miller (2017) and Bonsall et al. (2017) argue that both measures (total file size and the Fog Index) have shortcomings. For example, the total file size may be driven by unrelated text in an annual report. The Fog Index is often misspecified because the formula [equation (3)] itself fails to capture the sufficient perspectives of readability (Loughran and McDonald, 2014).

In addition to the above two measures, we also include FLESCHE reading ease score, which is calculated using the following formula.

$$\text{FLESCHE} = 206.835 - (1.015 \times \text{Average Sentence Length}) - (84.6 \times \text{Average Number of Syllables per word}) \quad (4)$$

A higher FLESCHE score indicates higher readability (i.e., easier to read). For example, a range of 90–100 indicates ‘very easy’, while a range of 0–29 suggests ‘very confusing’.

Because of the popularity of these measures in prior readability studies, we use these three measures, namely FSIZE, FOG, and FLESCHE, as our alternative readability measures⁵. Specifically, we re-estimate equation (2) using these three measures as dependent variables and report results in Table 5. As shown in Table 5, the coefficients on UNVA_RANK are 0.0003 with a t-value of 2.26 and 0.001 with a t-value of 2.69 where the dependent variables are FSIZE and FOG, respectively. Column 3 shows that the coefficient on UNVA_RANK is –0.005 with a t-value of –2.37 where FLESCHE is the dependent variable. Taken together, these results suggest annual reports of firms with more unverifiable assets and liabilities are less readable. Hence, results of Table 5 suggest our primary findings still hold when using different readability measures.

Table 5 Unverifiable net assets ratio and annual report reading difficulty (alternative readability measures)

	<i>Column 1</i>		<i>Column 2</i>		<i>Column 3</i>	
	<i>Dep. var. = FSIZE</i>		<i>Dep. var. = FOG</i>		<i>Dep. var. = FLESCH</i>	
	<i>Estimate</i>	<i>t value</i>	<i>Estimate</i>	<i>t value</i>	<i>Estimate</i>	<i>t value</i>
Intercept	12.344***	216.51	15.503***	29.51	31.518***	22.26
UNVA_RANK	0.0003**	2.26	0.001**	2.69	-0.005**	-2.37
SIZE	0.153***	29.14	0.241***	10.61	-0.502***	-9.65
LEV	0.101***	3.42	0.174	1.52	-0.282	-1.02
ROA	-0.393***	-3.49	-0.598	-1.16	4.138***	3.65
MTB	0.002**	2.14	0.002	0.33	-0.032**	-2.39
CFO	-0.134	-1.42	-0.582	-1.59	0.649	0.73
CAPINT	-0.155	-1.51	-0.655*	-1.72	5.941***	6.44
FIRMAGE	-0.070***	-7.81	-0.118***	-2.66	0.325***	3.20
BONDRATING	-0.045***	-17.81	-0.054***	-5.51	0.048**	1.98
MABILITY	-0.035**	-2.06	0.055	0.72	-0.069	-0.39
VOLATILITY	0.023**	2.47	0.104**	2.48	-0.400***	-4.37
SPECIAL	0.135	1.14	0.812	1.52	-3.560***	-2.86
MERGE	0.062***	6.30	0.023	0.50	-0.157	-1.54
EARNINGS	0.126***	3.87	0.222	1.39	-0.856**	-2.44
ACCRUALS	-0.054	-0.46	0.842*	1.82	-0.916	-0.84
LOSS	0.007	0.47	0.146**	2.33	-0.114	-0.80
SEGMENT	0.050***	11.12	0.090***	3.69	-0.333***	-6.12
BIG4	-0.026	-1.14	0.134	0.64	-0.465	-1.15
Industry indicators	Yes		Yes		Yes	
Year indicators	Yes		Yes		Yes	
Observations	12,052		9,763		9,763	
Adjusted R ²	0.2417		0.0776		0.1982	

Notes: This table presents the results of the baseline regression model using three alternative readability measures, namely FSIZE, FOG, and FLESCH. FSIZE is the natural log of the size of annual reports. FOG is the FOG reading score of annual reports. FLESCH is the FLESCH reading ease score of annual reports. The continuous variables in the baseline regression model are winsorised at the 1% and 99% percentiles each year before entering the regression analysis. *, **, and *** denote significance at the 10, 5 and 1 percent (two-tailed) confidence levels, respectively. Detailed variable definitions are provided in Appendix 1.

5.2 Alternative sample periods

We also check whether our primary findings hold across different time periods. Specifically, we randomly divide the sample period (1993–2016) into three periods, namely 1993–1999, 2000–2009, and 2010–2016, and re-estimate equation (2) for each period. Table 6 shows that the coefficients on UNVA_RANK are significant and positive in each period. For instance, the coefficient on UNVA_RANK for the period of 2010–2016 is 0.013 with a t-value of 3.53. Results of Table 6 suggest that our primary findings are less likely to subject to changes in macroeconomic factors and are clearly robust over time.

Table 6 Unverifiable net assets ratio and annual report reading difficulty (alternative sample periods)

	<i>Column 1</i>		<i>Column 2</i>		<i>Column 3</i>	
	<i>1993–1999</i>		<i>2000–2009</i>		<i>2010–2016</i>	
	<i>Estimate</i>	<i>t value</i>	<i>Estimate</i>	<i>t value</i>	<i>Estimate</i>	<i>t value</i>
Intercept	81.012***	58.14	80.711***	105.99	85.627***	86.66
UNVA_RANK	0.027***	5.27	0.016***	5.18	0.013***	3.53
SIZE	0.737***	6.55	0.608***	8.48	0.827***	8.59
LEV	-0.162	-0.26	0.213	0.56	1.942***	3.71
ROA	-5.397*	-1.85	-1.184	-0.83	-2.078	-1.00
MTB	0.055*	1.87	0.019	1.20	0.021*	1.73
CFO	-3.167*	-1.66	-0.725	-0.59	-12.766***	-6.40
CAPINT	-5.219***	-2.76	-8.078***	-4.72	-1.555	-0.91
FIRMAGE	-1.437***	-6.30	-0.924***	-7.26	-1.014***	-6.65
BONDRATING	-0.718***	-13.77	-0.263***	-7.54	-0.232***	-4.70
MABILITY	-0.540	-1.47	0.786***	3.20	0.157	0.53
VOLATILITY	1.386***	8.08	0.765***	5.93	0.848***	5.09
SPECIAL	-3.975	-1.12	-1.707	-1.02	1.398	0.51
MERGE	0.882***	4.03	0.914***	6.50	0.900***	5.27
EARNINGS	2.614**	2.37	0.764*	1.94	0.399	0.66
ACCRUALS	0.329	0.15	-0.197	-0.13	6.732***	3.11
LOSS	0.983***	2.92	0.851***	4.21	0.126	0.47
SEGMENT	1.180***	7.55	0.542***	8.06	0.497***	6.75
BIG4	1.422	1.62	1.413***	3.47	-0.180	-0.56
Industry Indicators	Yes		Yes		Yes	
Year Indicators	Yes		Yes		Yes	
Observations	4,178		6,781		4,270	
Adjusted R ²	0.4103		0.359		0.3257	

Notes: This table presents the results of the baseline regression model for three alternative time periods, namely 1993–1999, 2000–2009, and 2010–2016. The continuous variables in the baseline regression model are winsorised at the 1% and 99% percentiles each year before entering the regression analysis. *, **, and *** denote significance at the 10, 5 and 1 percent (two-tailed) confidence levels, respectively. Detailed variable definitions are provided in Appendix 1.

5.3 Addressing omitted correlated variables

Determining the level of the readability of annual reports or/and the level of unverifiable assets and liabilities is a firm-specific decision. In addition, it is possible that some unknown firm characteristics may influence the level of readability and unverifiable net assets ratio simultaneously, causing a significant positive relation. To curtail the above concerns about omitted correlated variables, we first employ a changes analysis to examine the relation between the changes in net unverifiable assets ratio and the changes in reading difficulty of annual reports. Specifically, based on equation (2), we regress the changes in reading difficulty from year $t-1$ to year t (i.e., $\Delta\text{DIFFICULTY}_t = \text{DIFFICULTY}_t - \text{DIFFICULTY}_{t-1}$) on the changes in unverifiable net assets ratio from year $t-1$ to year t (i.e., $\Delta\text{UNVA_RANK}_t = \text{UNVA_RANK}_t - \text{UNVA_RANK}_{t-1}$) and

changes in control variables from year-1 to year. Column 1 of Table 7 presents that the coefficient on Δ UNVA_RANK is 0.001 with a t-value of 1.79, suggesting a significant positive relation between the changes in unverifiable net assets ratio and the changes in reading difficulty. This evidence shows that an increase (a decrease) in unverifiable net assets ratio can lead to an increase (a decrease) in the reading difficulty of annual reports, strengthening our primary findings.

Table 7 Unverifiable net assets ratio and annual report reading difficulty (addressing omitted correlated variables)

	<i>Clustered standard errors OLS</i>		<i>Firm fixed effects regression</i>	
	<i>Column 1</i>		<i>Column 2</i>	
	<i>Dep. var. = ADIFFICULTY</i>		<i>Dep. var. = DIFFICULTY</i>	
	<i>Estimate</i>	<i>t value</i>	<i>Estimate</i>	<i>t value</i>
Intercept	0.122	0.73	Intercept	
Δ UNVA_RANK	0.001*	1.79	UNVA_RANK	0.003** 2.32
Δ SIZE	1.206***	7.67	SIZE	1.057*** 12.17
Δ LEV	0.512	1.39	LEV	1.012*** 3.45
Δ ROA	-1.350*	-1.86	ROA	-3.237*** -3.67
Δ MTB	-0.006	-0.74	MTB	0.005 0.60
Δ CFO	-0.891*	-1.66	CFO	-1.588** -2.24
Δ CAPINT	-0.898	-0.99	CAPINT	0.057 0.06
Δ FIRMAGE	2.030**	2.33	FIRMAGE	-2.648*** -9.22
Δ BONDRATING	-0.197***	-5.33	BONDRATING	-0.398*** -14.36
Δ MABILITY	0.099	0.96	MABILITY	0.063 0.44
Δ VOLATILITY	0.088	0.88	VOLATILITY	0.505*** 6.46
Δ SPECIAL	0.005	0.01	SPECIAL	1.645* 1.73
Δ MERGE	0.116**	2.19	MERGE	0.421*** 5.44
Δ EARNINGS	-0.190	-0.59	EARNINGS	0.629** 2.25
Δ ACCRUALS	0.487	0.80	ACCRUALS	1.096 1.35
Δ LOSS	0.089	1.21	LOSS	0.285*** 2.71
Δ SEGMENT	0.082**	2.23	SEGMENT	0.255*** 4.84
Δ BIG4	-1.440	0.15	BIG4	-0.739** -2.30
Industry indicators	Yes		Industry indicators	Yes
Year indicators	Yes		Year indicators	Yes
Observations	13,308		Observations	15,229
Adjusted R ²	0.0479		Adjusted R ²	0.8151

Notes: This table first presents the results of the baseline regression model using changes in dependent variable, in primary independent variable and in control variables.

The model is as follows. Δ DIFFICULTY = $\beta_0 + \beta_1 \times \Delta$ UNVA_RANK + $\beta_x \times \Delta$ Control Variables + Industry Indicators + Year Indicators + ϵ . Next, this table reports results of the baseline regression model using firm fixed effects regression. The continuous variables in the baseline regression model are winsorised at the 1% and 99% percentiles each year before entering the regression analysis.

*, **, and *** denote significance at the 10, 5 and 1 percent (two-tailed) confidence levels, respectively. Detailed variable definitions are provided in Appendix 1.

Table 8 Unverifiable net assets ratio and annual report reading difficulty (addressing reverse causality)

	Using lagged UNVA measures				Two-stage OLS regression (2SLS)			
	Dependent variable = DIFFICULTY				Dep. var. = DIFFICULTY			
	Column 1	Column 2	Column 3	Column 4	Column 5	Column 4	Column 5	
Estimate	t value	Estimate	t value	Estimate	t value	Estimate	t value	
Intercept	85.018***	134.58	85.175***	125.29	85.479***	118.67	83.148***	144.67
Lag_UNVA1	0.019***	8.46	0.023***	9.59	0.022***	8.51	0.793***	19.75
Lag_UNVA2								
Lag_UNVA3								
UNVA_MEAN								
UNVA_INSTRUMENTAL								
SIZE	0.741***	13.28	0.720***	12.33	0.697***	11.47	0.018***	8.15
LEV	0.445	1.46	0.388	1.17	0.160	0.45	0.789***	15.59
ROA	-4.157***	-3.35	-4.758***	-3.54	-5.626***	-3.96	0.533*	1.88
MTB	0.008	0.84	0.001	0.09	-0.006	-0.58	-3.090***	-2.65
CFO	-3.229***	-3.21	-4.132***	-3.79	-4.134***	-3.58	0.023***	2.29
CAPIT	-4.115***	-3.54	-3.505***	-2.77	-3.596***	-2.56	-3.658***	-3.96
FIRMAGE	-1.039***	-10.16	-1.070***	-9.57	-1.134***	-9.35	-3.471***	-3.57
BONDRATING	-0.370***	-13.69	-0.333***	-11.75	-0.299***	-10.04	-1.118***	-12.00
MABILITY	0.133	0.72	0.217	1.11	0.352*	1.71	-0.436***	-18.18
VOLATILITY	1.099***	11.15	1.205***	10.96	1.193***	9.81	0.088	0.52
SPECIAL	0.644	0.46	0.548	0.36	1.057	0.67	0.976***	10.98
MERGE	0.898***	8.47	0.928***	8.28	0.932***	7.85	1.021***	3.19
EARNINGS	0.966***	2.68	1.034**	2.50	1.326***	2.94	8.787*	1.84
ACCRUALS	3.098**	2.52	3.415**	2.53	4.368***	3.03	-3.733***	-10.46
LOSS	0.656***	4.12	0.547***	3.18	0.401**	2.19	-12.072***	-9.46
SEGMENT	0.618***	11.86	0.592***	10.76	0.593***	10.13	38.661***	9.72
BIG4	0.657**	2.37	0.550*	1.77	0.404	1.23	2.176***	1.96
Industry Indicators	Yes	Yes	Yes	Yes	Yes	Yes	0.443	0.83
Year Indicators	Yes	Yes	Yes	Yes	Yes	Yes	-0.186	-1.05
Observations	13,308		11,623		10,161		2.327**	2.42
Adjusted R ²	0.4406		0.4363		0.4276		83.148***	144.67

Notes: This table first presents the results of the baseline regression model using three lagged UNVA_RANK measures, namely lag_UNVA1, lag_UNVA2, and lag_UNVA3. The model is as follows: DIFFICULTY = $\alpha_0 + \alpha_1 \times \text{lagged UNVA_RANK} + \alpha_2 \times \text{Control Variables} + \text{Industry Indicators} + \text{Year Indicators} + \varepsilon$. Next, this table reports results of two-stage OLS regression analysis (2SLS). The continuous variables in the baseline regression model are winsorized at the 1% and 99% percentiles each year before entering the regression analysis. *, **, and *** denote significance at the 10, 5 and 1 percent (two-tailed) confidence levels, respectively. Detailed variable definitions are provided in Appendix 1.

Next, we re-estimate equation (2) using firm fixed effects regression and report results in Column 2 of Table 7. The coefficient on UNVA_RANK is still significant and positive, indicating that annual reports of firms with more unverifiable assets and liabilities are more difficult for average persons to read and understand. In summary, the results of Table 7 not only strengthen the robustness of our primary findings but also effectively mitigate concerns about omitted correlated variables in our empirical analysis.

5.4 Addressing reverse causality

To curtail any concerns about reverse causality in our study, we re-estimate equation (2) using lagged values of UNVA_RANK and report results in Table 8. As shown in columns 1–3 of Table 8, the coefficient on Lag_UNVA1 (i.e., unverifiable net assets ratio in year $t-1$), Lag_UNVA2 (i.e., unverifiable net assets ratio in year $t-2$), and Lag_UNVA3 (i.e., unverifiable net assets ratio in year $t-3$) are all significant and positive, which suggests that our primary independent variables of interests (i.e., unverifiable net assets ratio) drive the dependent variable (i.e., DIFFICULTY).

We further perform a two-stage OLS regression analysis (2SLS) in which we estimate instrumented variables (UNVA_INSTRUMENTAL) using mean values of UNVA_RANK of firms in the same industry based on the first two digits of the SIC code in the first stage, and re-estimate equation (2) using UNVA_INSTRUMENTAL (obtained from the first stage of 2SLS) in the second stage. Column 4 and 5 report the results of the first stage and the second stage of 2SLS, respectively. As shown in column 5 of Table 8, the coefficient on UNVA_INSTRUMENTAL is 0.018 with a p-value of 8.15. Taken together, results of Table 8 curtail concerns about reverse causality.

5.5 High tech firms vs. low tech firms

Recent anecdotal evidence suggests that high tech firms are more transparent in financial reporting, due to the help of technology. If this is the case, we expect that the volatility in reading difficulty of high tech firms may be smaller than that of low tech firms, suggesting that our primary findings may become stronger for firms in low tech industries. In this test, we divide our sample into two subsamples (high tech firms and low tech firms) and re-estimate equation (2) for each subsample. We follow Kile and Phillips (2009) to identify high tech firms which are in one of the following industries: SIC = 283, 357, 366, 382, 384, 481, 482, 489, 737, and 873. Table 9 reports that the coefficient on UNVA_RANK is 0.010 with a t-value of 2.31 for high tech firms and 0.019 with a t-value of 7.16 for low tech firms. The coefficient comparison tests (0.010 vs. 0.019) show that the difference between the two coefficients is statistically significant (i.e., p-value = 0.0077), which suggests that the coefficient on UNVA_RANK for low tech firms are significantly larger than those for high tech firms. In other words, the significant positive relation between unverifiable net assets ratio and the reading difficulty of annual reports becomes stronger for low tech firms.

Table 9 Unverifiable net assets ratio and annual report reading difficulty (high tech firms vs. low tech firms)

	<i>Clustered standard errors OLS</i>			
	<i>Column 1</i>		<i>Column 2</i>	
	<i>High tech firms</i>		<i>Low tech firms</i>	
	<i>Estimate</i>	<i>t value</i>	<i>Estimate</i>	<i>t value</i>
Intercept	90.420***	74.99	85.360***	128.22
UNVA_RANK	0.010**	2.31	0.019***	7.16
SIZE	0.628***	5.50	0.752***	12.32
LEV	-0.509	-0.80	0.950***	2.99
ROA	-3.520	-1.45	-1.283	-0.96
MTB	-0.001	-0.03	0.032***	2.87
CFO	-11.111***	-5.14	-3.480***	-3.28
CAPINT	2.459	0.86	-6.699***	-5.96
FIRMAGE	-0.326	-1.45	-1.173***	-11.62
BONDRATING	-0.374***	-6.76	-0.420***	-14.27
MABILITY	0.390	0.98	-0.015	-0.08
VOLATILITY	0.457**	2.20	1.092***	11.14
SPECIAL	0.362	0.12	-1.070	-0.72
MERGE	0.734***	3.36	1.148***	10.28
EARNINGS	2.382***	3.93	0.289	0.74
ACCRUALS	1.813	0.86	3.005**	2.25
LOSS	0.374	1.10	0.809***	4.90
SEGMENT	-0.007	-0.07	0.743***	13.29
BIG4	0.818	1.22	0.655**	2.41
Industry indicators	Yes		Yes	
Year indicators	Yes		Yes	
Observations	3,152		12,147	
Adjusted R ²	0.2676		0.4244	
Coefficient comparison test				
Coefficient on UNVA_RANK of high tech vs. coefficient of UNVA_RANK of low tech				
F-Stat. = 7.00; P-value = 0.0077				

Notes: This table presents the results of the baseline regression model using clustered standard errors OLS regression for two subsamples, namely firms in high tech industries and firms in low tech industries. The baseline model is as follows. DIFFICULTY = $\alpha_0 + \alpha_1 \times \text{UNVA_RANK} + \alpha_x \times \text{Control Variables} + \text{Industry Indicators} + \text{Year Indicators} + \varepsilon$. The continuous variables in the baseline regression model are winsorised at the 1% and 99% percentiles each year before entering the regression analysis. *, **, and *** denote significance at the 10, 5 and 1 percent (two-tailed) confidence levels, respectively. In Appendix 1, we provide detailed variable definitions.

Table 10 Unverifiable net assets ratio and annual report reading difficulty
(firms with investment grade rating vs. firms without investment grade rating)

	Clustered standard errors OLS			
	Column 1		Column 2	
	Firms with investment grade rating		Firms without investment grade rating	
	Estimate	t value	Estimate	t value
Intercept	88.011***	70.10	84.018***	103.93
UNVA_RANK	0.034***	6.20	0.016***	6.08
SIZE	0.549***	6.44	0.866***	12.75
LEV	1.090	1.53	0.925***	2.94
ROA	-6.599**	-2.26	-1.663	-1.34
MTB	0.063***	3.66	0.027**	2.14
CFO	-5.320***	-3.14	-1.693	-1.49
CAPINT	-0.331	-0.16	-4.062***	-3.34
FIRMAGE	-0.754***	-4.90	-1.109***	-9.56
BONDRATING	-0.629***	-12.62	-0.412***	-8.60
MABILITY	0.617**	2.34	-0.006	-0.02
VOLATILITY	1.646***	9.83	0.751***	7.33
SPECIAL	-5.608*	-1.71	1.457	1.04
MERGE	0.866***	5.59	1.005***	7.83
EARNINGS	2.888***	3.01	0.103	0.29
ACCRUALS	0.795	0.42	1.598	1.18
LOSS	0.609*	1.76	0.964***	5.85
SEGMENT	0.732***	9.24	0.547***	8.92
BIG4	1.409*	1.85	0.369	1.41
Industry Indicators	Yes		Yes	
Year Indicators	Yes		Yes	
Observations	7,179		8,050	
Adjusted R ²	0.5041		0.3954	

Coefficient comparison test

Coefficient on UNVA_RANK of firms with investment grade rating vs.
coefficient of UNVA_RANK of firms without investment grade rating

F-Stat. = 98.46; P-value < 0.0001

Notes: This table presents the results of the baseline regression model using clustered standard errors OLS regression for two subsamples, namely firms with investment grade bond rating and firms without investment grade bond rating. The baseline model is as follows. $DIFFICULTY = \alpha_0 + \alpha_1 \times UNVA_RANK + \alpha_x \times \text{Control Variables} + \text{Industry Indicators} + \text{Year Indicators} + \epsilon$. The continuous variables in the baseline regression model are winsorised at the 1% and 99% percentiles each year before entering the regression analysis. *, **, and *** denote significance at the 10, 5 and 1 percent (two-tailed) confidence levels, respectively. In Appendix 1, we provide detailed variable definitions.

5.6 *Investment grade bond ratings vs. non-investment grade bond ratings*

Bonsall and Miller (2017) document a positive association between bond credit ratings and the reading ease of annual reports, which suggests that annual reports of firms with higher ratings (e.g., investment grade ratings and above) have more readable. Hence, we expect that the volatility in reading difficulty of firms with higher bond ratings may be larger than that of firms with lower bond ratings because annual reports of firms with lower bond ratings are already difficult to read and understand.

We consider a firm to have a high bond rating if the rating is investment grade and above (i.e., BBB- and above). We divide our sample into two subsamples (firms with investment grade rating and firms without investment grade rating), re-estimate equation (2) for each subsample and report results in Table 10. As shown in Table 10, the coefficient on UNVA_RANK is 0.034 (0.016) with a p-value of 6.20 (6.08) for firms with investment grade rating (firms without investment grade rating). Based on the F-values and p-values in the coefficient comparison tests, we conclude that our primary results are stronger for firms with investment grade rating.

6 Conclusions

In this study, we examine the impact of unverifiable assets and liabilities on the reading difficulty of annual reports utilising the BOG index, a new measure to capture reading difficulty. In support of our hypothesis, we find a significant positive relation between unverifiable net assets ratio and the reading difficulty of annual reports for a large sample of US public firms. Our results suggest that annual reports of firms with more unverifiable assets and liabilities are more difficult for average persons to read and understand (i.e., less readable). Our results still hold after a battery of robustness checks and additional tests. Moreover, we find that our primary results become stronger for firms in low technology industries or firms with high bond ratings (i.e., investment grade rating and above).

Our results have important implications to different stakeholders, namely the SEC. The SEC has been persistently pushing companies to make their regulatory filings, such as annual reports, easier for people to read and understand because the overall readability of regulatory filings is still poor. This study helps shed some light on perhaps why readability is still low and what can be done about it. First, the outcome of this study provides the SEC with compelling evidence regarding the determinants of readability of annual reports. Second, we provide empirical evidence of the utility of the BOG index to identify firms with less readable annual reports. Third, we provide empirical evidence for the utility of unverifiable assets and liabilities as a proxy measure for report readability.

Just like many other studies, our study has several limitations. First, our sample firms are large public US firms. Whether our results hold for private firms or international firms remains unknown. Second, we use the ratio of unverifiable net assets to show the level of unverifiable assets and liabilities. Other measures to capture the level of such assets and liabilities may also exist. Third, the readability measure (i.e., BOG index) is an approximate measure that is not perfect. For example, it is possible that our readability measure captures more reading difficulty of narrative disclosures than that of financial statements. Therefore, more precise measures of readability may lead to stronger results. The above issues can be further investigated in future research.

References

- Bloomfield, R. (2002) 'The incomplete revelation hypothesis: implications for financial reporting', *Accounting Horizons*, Vol. 16, No. 3, pp.233–244.
- Bonsall, S., Leone, A.J., Miller, B.P. and Rennekamp, K. (2017) 'A plain English measure of financial reporting readability', *Journal of Accounting and Economics*, Vol. 63, No. 2, pp.329–357.
- Bonsall, S.B. and Miller, B.P. (2017) 'The impact of narrative disclosure readability on bond ratings and the cost of debt capital', *Review of Accounting Studies*, Vol. 22, No. 2, pp.608–643.
- Bozanic, Z. and Thevenot, M. (2015) 'Qualitative disclosure and changes in sell-side financial analysts' information environment', *Contemporary Accounting Research*, Vol. 32, No. 4, pp.1595–1616.
- De Franco, C., Hope, O., Vyas, D. and Zhou, Y. (2015) 'Analyst report readability', *Contemporary Accounting Research*, Vol. 32, No. 1, pp.76–104.
- Demerjian, P., Lev, B. and McVay, S. (2012) 'Quantifying managerial ability: a new measure and validity tests', *Management Science*, Vol. 58, No. 7, pp.1229–1248.
- Dunn, P. (2004) 'The impact of insider power on fraudulent financial reporting', *Journal of Management*, Vol. 30, No. 3, pp.397–412.
- Ertugrul, M., Lei, J., Qiu, J. and Wan, C. (2017) 'Annual report readability, tone ambiguity, and the cost of borrowing', *Journal of Financial and Quantitative Analysis*, Vol. 52, No. 2, pp.811–836.
- Ghosh, D. and Olsen, L. (2009) 'Environmental uncertainty and managers' use of discretionary accruals', *Accounting, Organizations and Society*, Vol. 34, No. 2, pp.188–205.
- Guay, W., Samuels, D. and Taylor, D. (2016) 'Guiding through the fog: financial statement complexity and voluntary disclosure', *Journal of Accounting and Economics*, Vol. 62, Nos. 2–3 pp.234–269.
- Hasan, M.M. (2020) 'Readability of narrative disclosure in 10-K reports: does managerial ability matter?', *European Accounting Review*, Vol. 29, No. 1, pp.147–168.
- Kile, C.O. and Phillips, M.E. (2009) 'Using industry classification codes to sample high-technology firms: Analysis and recommendations', *Journal of Accounting, Auditing and Finance*, Vol. 24, No. 1, pp.35–58.
- Kothari, S.P., Leone, A.J. and Wasley, C.E. (2005) 'Performance matched discretionary accrual measures', *Journal of Accounting and Economics*, Vol. 39, No. 1, pp.163–197.
- Lawrence, A. (2013) 'Individual investors and financial disclosure', *Journal of Accounting and Economics*, Vol. 56, No. 1, pp.130–147.
- Lehavy, R., Li, F., Merkley, K. (2011) 'The effect of annual report readability on analyst following and the properties of their earnings forecasts', *The Accounting Review*, Vol. 86, No. 3, pp.1087–1115.
- Li, F. (2008) 'Annual report readability, current earnings, and earnings persistence', *Journal of Accounting and Economics*, Vol. 45, No. 2, pp.221–247.
- Lo, K., Ramos, F. and Rogo, R. (2017) 'Earnings management and annual report readability', *Journal of Accounting and Economics*, Vol. 63, No. 1, pp.1–25.
- Loughran, T. and McDonald, B. (2011) 'When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks', *Journal of Finance*, Vol. 66, No. 1, pp.1–25.
- Loughran, T. and McDonald, B. (2014) 'Measuring readability in financial disclosures', *Journal of Finance*, Vol. 69, No. 4, pp.1643–1671.
- Luo, J.H., Li, X. and Chen, H. (2018) 'Annual report readability and corporate agency costs', *China Journal of Accounting Research*, Vol. 11, No. 3, pp.187–212.
- Petersen, M.A. (2009) 'Estimating standard errors in finance panel data sets: comparing approaches', *Review of Financial Studies*, Vol. 22, No. 1, pp.435–480.

- Ramanna, K. (2008) 'The implications of unverifiable fair-value accounting: evidence from the political economy of goodwill accounting', *Journal of Accounting and Economics*, Vol. 45, No. 2, pp.253–281.
- Ramanna, K. and Watts, R.L. (2012) 'Evidence on the use of unverifiable estimates in required goodwill impairment', *Review of Accounting Studies*, Vol. 17, No. 4, pp.749–780.
- Richardson, S.A., Sloan, R.G., Soliman, M.T. and Tuna, I. (2005) 'Accrual reliability, earnings persistence and stock prices', *Journal of Accounting and Economics*, Vol. 39, No. 3, pp.437–485.
- You, H. and Zhang, X. (2009) 'Financial disclosure complexity and investor underreaction to 10-K information', *Review of Accounting Studies*, Vol. 14, No. 4, pp.559–586.

Notes

- 1 A hidden verb is a verb that is converted into a noun. If a sentence does not provide anything new, this is known as overwriting. An example of hidden verb and overwriting in the 2018 10-K report of Northrop Grumman is as follows. "We offer a broad portfolio of capabilities and technologies that enable us to deliver innovative platforms, systems and solutions for applications that range from undersea to outer space and into cyberspace." This sentence can be rewritten as "We deliver innovative platforms, systems and solutions for applications that range from undersea to outer space and into cyberspace." Heavy words refer to difficult words such as 'literally' and 'ironic'. An example of heavy words in the 2018 10-K report of Northrop Grumman is as follows. "We provide capabilities in autonomous systems; cyber; command, control, communications and computers, intelligence, surveillance and reconnaissance (C4ISR); space; strike; and logistics and modernization."
- 2 <https://kelley.iu.edu/bpm/activities/bogindex.html>
- 3 2016 is the last year of our sample because the Compustat database discontinued reporting S&P bond ratings in 2017.
- 4 $0.018 \times 23.82 \div 7.462$.
- 5 We thank Professor Feng Li for providing the Fog Index data and Flesch reading ease data and Professor McDonald for providing file size data.

Appendix 1

Table A1 Variable definitions

<i>Main variables</i>	<i>Definition</i>
DIFFICULTY =	The Bog Index by Bonsall et al. (2017) to measure the reading difficulty
UNVA_RANK =	The ratio of cash (CHE) to total firm assets (AT)
SIZE =	The natural log of total firm assets (AT)
LEV =	Leverage ratio, measure as the ratio of long-term liabilities (DLTT) to total assets (AT)
ROA =	Profitability, measured as income before extraordinary items (IB) scaled by total assets (AT)
MTB =	Market to book ratio, measured as market value of outstanding common shares [Outstanding common shares (CSHO) \times price at fiscal year-end (PRCC_F)] divided by total book value of outstanding common shares (CEQ)
CFO =	Cash flows from operating activities (OANCF) scaled by total assets (AT)

Table A1 Variable definitions (continued)

<i>Main variables</i>		<i>Definition</i>
CAPINT	=	Capital expenditures (CAPX) scaled by total assets (AT)
FIRMAGE	=	The natural log of the number of firms since the firm has been listed in the Compustat database
BONDRATING	=	S&Ps' bond credit ratings
MABILITY	=	Managerial ability ranks by Demerjian et al. (2012)
VOLALITY	=	Five year rolling average sales (SALE) volatility based on the procedure in Ghosh and Olsen (2009)
SPECIAL	=	Special items (SPI) scaled by total assets (AT)
MERGE	=	An indicator variable that equals one if a firm engages in merger & acquisition transactions and otherwise zero
EARNINGS	=	Earnings scaled by total assets (AT)
ACCRUALS	=	The absolute value of discretionary accruals using performance-matched modified Jones model
LOSS	=	An indicator variable that equals one if a firm reports a loss otherwise zero
SEGMENT	=	The natural log of the number of business segments
BIG4	=	An indicator variable that equals one if a firm uses a Big 4 accounting firm as their auditor otherwise zero
<i>Other variables</i>		
FILESIZE	=	The natural log of the gross file size of 10-K reports by Loughran and McDonald (2011)
FOG	=	The Fog Index
FLESCH	=	The Flesch Reading Ease Index
Δ DIFFICULTY	=	DIFFICULTY in year t – DIFFICULTY in year t – 1
Δ UNVA_RANK	=	UNVA_RANK in year t – UNVA_RANK in year t – 1
Δ SIZE	=	SIZE in year t – SIZE in year t – 1
Δ LEV	=	LEV in year t – LEV in year t – 1
Δ ROA	=	ROA in year t – ROA in year t – 1
Δ MTB	=	MTB in year t – MTB in year t – 1
Δ CFO	=	CFO in year t – CFO in year t – 1
Δ CAPINT	=	CAPINT in year t – CAPINT in year t – 1
Δ FIRMAGE	=	FIRMAGE in year t – FIRMAGE in year t – 1
Δ MABILITY	=	MABILITY in year t – MABILITY in year t – 1
Δ BONDRATING	=	BONDRATING in year t – BONDRATING in year t – 1
Δ VOLALITY	=	VOLALITY in year t – VOLALITY in year t – 1
Δ SPECIAL	=	SPECIAL in year t – SPECIAL in year t – 1
Δ MERGE	=	MERGE in year t – MERGE in year t – 1
Δ EARNINGS	=	EARNINGS in year t – EARNINGS in year t – 1
Δ ACCRUALS	=	ACCRUALS in year t – ACCRUALS in year t – 1
Δ LOSS	=	LOSS in year t – LOSS in year t – 1
Δ SEGMENT	=	SEGMENT in year t – SEGMENT in year t – 1
Δ BIG4	=	BIG4 in year t – BIG4 in year t – 1.