Information Disclosure in 10-Ks – The Role of External Auditors

Abstract

This paper disaggregates the role of auditors and management in disclosing information in the 10-K reports. It employs various readability and similarity measures and relates them to audit and financial outcomes to show that both sections react differently depending on who prepares them. The results show that the MD&A (item 7) section of the financial statement is associated with financial outcomes, such as earnings but not audit outcomes, such as whether or not the financial statements are qualified. The reverse is the case for the financial statements (item 8). The paper finds that the financial statements become more challenging to read when the auditors are big four firms, the firm is large, or the financial statements are qualified. It also finds that based on the similarity score, new information is introduced in item 7 and item 8 of the 10-K reports when the financial statements are qualified.

1. Introduction

According to the Public Company Accounting Oversight Board, the financial statements are the management's responsibility. The auditor's responsibility is to express an opinion on the financial statements. The independent auditor may make suggestions about the form or content of the financial statements or draft them, in whole or part, based on information from management during the audit. In essence, the auditor's responsibility for the financial statements he or she has audited is confined to the expression of his or her opinion on them.

However, in reality, the auditor can prepare and present the information contained in the financial statements. We argue that a firm's auditors prepare some sections of the 10-K, and information contained therein and communicated to FS users is under the auditors' control. Thus, auditors play a role in the communication process in the financial markets, outside just expressing an opinion. This implies that prior and future literature that uses the 10-K as a whole inadvertently observes the effect of two different actors driven by different objectives.

When the auditor prepares financial statements, it is considered a non-attest service. According to the technical standards, the auditor's service of preparing or assisting in preparing the financial statements must be evaluated and appropriately documented. This is usually included in the management letter for an audit engagement. Non-attest services include financial statement preparation, cash-to-accrual conversions, reconciliations, and tax return preparation. The American Institute of Certified Public Accountants presents guidance for when such services are rendered.

The auditor can offer this service if certain professional and practical considerations are met: management must assign an individual with the necessary skills, knowledge, and experience to oversee the auditor's non-attest service of preparing financial statements, management must also acknowledge and accept responsibility in writing for the auditor's preparation of financial statements. Before concluding the audit, the auditor must obtain a representation letter from management confirming responsibility for preparing and presenting the financial statements per the applicable financial reporting framework.

Based on the discrepancy in the preparation role observed above, a significant observation is that auditors and not management control information disclosure to

interested stakeholders, especially regarding GAAPs. This paper investigates this discrepancy by postulating that some sections of the 10-K, like Item 8, are under the purview of auditors who choose what to and what not to disclose.

Going by the assertion presented above, auditors have a significant role in disclosing information to shareholders. This can be seen as a competition between auditors and management to disclose more or less information in the 10-Ks. We assume that the Notes to the Financial Statement (NTFS) are the most crucial section for auditors to communicate information. This paper investigates whether item 8 (Financial Statements and Notes to the Financial Statements) and item 7 (Management Discussion and Analysis) are significantly different from each other and, if so, what information is put forth by auditors.

This paper aims to disaggregate the role of auditors in information disclosure on the face of 10-K filings. Specifically, we attempt to answer the following questions: Do the information in the MD&A and Notes to the Financial Statements (NTFS) section respond differently to audit and financial outcomes levers? This would be indicative of the variation in preparers (and objective). We test this in relation to accounting and performance outcomes. We also test the role of the Big 4 (Deloitte, Ernst & Young, KPMG, and PricewaterhouseCoopers) accounting firms on items seven and eight of the financial statement. We also extend our analysis to understand whether auditor change affects the similarity and readability of the NTFS but not the MD&A.

Based on the tests we have conducted, we find that the readability stats of the MDA and NTFS sections of the 10-K respond to audit and financial outcomes differently. The results show that the NTFS has a higher readability score (worse readability) and similarity stats

(increased repetition) on average than the MD&A. This is not surprising, given that the section uses technical accounting language. However, since this section is a large portion of the 10-Ks and is more likely to respond to changing accounting standards and not management performance, previous papers such as Li (2008), Biddle, Hilary, and Verdi (2009), Miller (2010), Cohen, Malloy, and Nguyen(2020) that use the entire 10-K readability and similarity are inadvertently capturing the effect of a different actor (the auditors, guided and responding to changing accounting and auditing standards).

The results show that the 10-K MD&A sections are associated with financial outcomes such as EBIT but not audit outcomes. The NTFS readability, on the other hand, is associated with audit outcomes but not financial outcomes. These results conform with the prior expectation that the NTFS becomes less readable given qualified financial statements and internal control reports. Auditors increase substantive testing to communicate and disclose more information (detailed information) to cover their downside and reduce future litigation incidences. Also, the NTFS produced by the big four firms is less readable than its counterparts, perhaps evidence of more information disclosure, although more complicated.

Also, the authors find that the similarity of both sections is associated with audit outcomes. In the presence of a qualified FS, similarity reduces for the MD&A section and the NTFS as both management and auditors disclose more information and explanation for the qualified FS. Again, we find that the similarity scores of the NTFS are not related to the financial outcomes meaning that auditors do not base information disclosure on the profitability of their clients.

The dataset used for this analysis was assessed from cleaned 10-K SEC/ EDGAR data files from the University of Notre Dame. Several cleaning and preprocessing procedures have been applied to the data discussed in section 3 below. The paper also uses data from Compustat for relevant fundamental data.

The remainder of the paper is organized as follows. Section 2 describes prior literature. Section 3 discusses data, procedures, and transformations applied, along with sample selection procedures. Section 4 reports our empirical results, and section 5 concludes.

2. Literature Review

Existing literature on readability and similarity measures of the 10-k is quite extensive. Readability stats have been associated with insider trading, firm performance (Li, 2008), capital investment (Biddle, Hilary, and Verdi 2009), stock price crashes (Kim, Wang, and Zhang 2019), and investor trading behavior (Miller 2010). However, few have focused on the role of auditors, who audit the information presented in the 10-Ks for relevance and reliability. These actors also perform non-attest services by preparing an essential section of the 10-Ks, item 8, financial statements, and notes to the financial statement.

Some papers have investigated the effects of auditor's choice and style on the 10-Ks. The paper by Johnston and Zhang (2021) titled Auditor Style and Financial Reporting Similarity examined whether auditor style is related to financial reporting similarity based on accounting items disclosed in eXtensible Business Reporting Language (XBRL) 10-K filings. Their findings have significant implications for the role of auditors in the 10-Ks. They define financial reporting similarity in terms of the number of similar line items reported by a pair of firms and develop a measure of pairwise financial reporting similarity.

The results show that firms with the same auditors have more similarities in their financial statements, a testament to the information disclosure role of auditors. They also find that financial reporting similarity increases (decreases) when firms switch from having different (the same) auditors to having the same (different) auditors.

This paper sets the background for our research hypothesis in this paper. While the authors above investigate similarity in account presentation, we investigate similarity based on the information disclosed in the Notes to the Financial Statements. Another paper by Hrazdil, Simunic, and Suwanyangyuan (2021) investigates Auditor Choice and the Informativeness of 10-K Reports. Their paper finds that choosing a Big 4 versus a non-Big four auditor contributes to cross-sectional variations in 10-K disclosure volume. They find that abnormally long disclosure is associated with higher audit fees and longer audit report lag, an instance of accounting disclosures responding to non-firm financial outcomes. They state that "*Overall, our findings show that auditors play more than a simple attestation role in the financial reporting process and that the quality of financial reporting in a company's 10-K annual report is a joint product of the effort and decisions of both a company's managers and its auditors"*.

To add to the literature on the information contribution of auditors, the paper by Boritz, Hayes, and Timoshenko (2016) analyzes the SOX 404 reports. SOX 404 requires public reporting companies to disclose the effectiveness of their internal control over financial reporting. They investigated the readability of these reports and associated it with company characteristics, auditor type, opinion type, period, report content, and report length. They find that reports of Big 6 firms are less readable than non-Big six reports, just as auditor reports are less readable than management. Kim, Wang, and Zhang (2019)

associate less readable 10-K reports with higher stock price risk. They argue that managers can successfully hide adverse information by writing complex financial reports.

Blanco et al. (2021) extend the results of Kim, Wang, and Zhang (2019) and show that increased audit effort is associated with lower annual report readability to compensate for a perceived increase in the risk of financial misstatement for United States (US) firms. This again highlights the significant role auditors play in information disclosure to the public.

Literature shows that textual analysis of 10-Ks is indicative of various characteristics and outcomes for firms. Nevertheless, the question remains that since auditors prepare a large portion of the 10-K filings that generally have less readability, the results of previous studies are partly driven by auditors, their internal preparation requirements, changing accounting standards, interpretation of accounting standards, and the extent of substantive testing used in the audit process.

In this paper, the author extends the literature on the role of auditors by investigating the differences between two important sections of the 10-Ks, item 7 and item 8, and associating the readability and similarity of both sections to audit and financial outcomes.

3. Data and empirical measures of annual report readability

3.1 Sample

This paper employs data from the Edgar Database of 10-K reports for companies in the USA. It uses the cleaned files from the University of Notre Dame. The first filed 10-K in this data begins in 1994, but the first reporting year goes back to 1990. The textual information is cross-referenced with the Compustat Database using company CIK

numbers to obtain fundamental information for each firm. The author programmatically extracts items 7 and 8 of the 10-Ks.

The sample is collected after several preprocessing stages: (1) Each 10-K file is parsed through an algorithm that attempts to locate item 7 and item 8. (2) The resulting files are only included in the sample if they are larger than 3 kilobytes. This process drops files with minute information in both sections or refers readers to the appendices of the 10-k for the relevant items. (3) The author then passes each file through the Textstat library to generate several readability scores. (4) The author passes the files through the Simphile and Textacy libraries to generate the similarity scores based on words in each file (excluding stop words and numbers). It is important to note that the similarity score requires data from each subsequent annual report. Thus, the similarity score is calculated if each subsequent year's section was captured in stage 1 above. (5) The author calculates the natural logarithm for large fundamental variables and standardizes all variables except indicator variables to aid interpretation.

The process result is 116,446 and 86,631 observations for MD&A readability and similarity stats, respectively. For item 8 (NTFS), the preprocessing stage above results in 75,835 and 56,604 observations for readability and similarity stats, respectively. The data is merged using company cik and year without dropping instances that do not match. The Compustat data is also merged to create the final dataset on the cik and year variables.

3.2 The readability measures

The author uses several statistics to measure the annual report's readability. They were all calculated using the Textstat library in python. The first is the Flesch Reading Ease Score (flesch_re_). The Flesch Reading Ease gives a text a score between 1 and 100,

with 100 being the highest readability score. Scoring between 70 to 80 is equivalent to school grade level 8. This means text should be reasonably easy for the average adult to read. The FRES test counts the number of words, syllables, and sentences in the text. It then calculates the average number of words per sentence and the average number of syllables per word. The idea is that shorter words and shorter sentences are easier to read. However, there is no lower limit to this score. Some very complicated sentences can have negative scores. The formula is as follows:

$$206.835 - 1.015\left(\frac{\text{total words}}{\text{total sentences}}\right) - 84.6\left(\frac{\text{total syllables}}{\text{total words}}\right)$$

The Flesch-Kincaid (flesh_kincaid_) is similar to the former test but uses different weighing factors. It refers to the years of education generally required to understand a specific text, relevant when the formula results in a number greater than 10. The grade level is calculated with the following formula:

$$0.39\left(\frac{\text{total words}}{\text{total sentences}}\right) + 11.8\left(\frac{\text{total syllables}}{\text{total words}}\right) - 15.59$$

The FRES is an inverse measure of the Flesch-Kincaid score. Thus, we have multiplied the FRES by a negative one to ensure interpretability across our results. The third measure is the Gunning fog index (fog_index_). Like the previous measures, the Fog index estimates a person's years of formal education to understand the text on the first reading. It is calculated as follows:

$$0.4\left[\left(\frac{\text{words}}{\text{sentences}}\right) + 100\left(\frac{\text{complex words}}{\text{words}}\right)\right]$$

Complex words above refer to words consisting of three or more syllables. The fourth measure is the SMOG index (smog_index_), a measure of readability that estimates the years of education needed to understand a piece of writing. It is calculated as follows:

$$\text{grade} = 1.0430\sqrt{\text{number of polysyllables} \times \frac{30}{\text{number of sentences}}} + 3.1291$$

Finally, the Dale–Chall (dale_chall_) readability formula is a readability test that provides a numeric gauge of readers' comprehension difficulty when reading a text. It uses a list of 3000 words that groups of fourth-grade American students could reliably understand, considering any word not on that list to be complicated. It is calculated as follows:

$$0.1579\left(\frac{\text{difficult words}}{\text{words}} \times 100\right) + 0.0496\left(\frac{\text{words}}{\text{sentences}}\right)$$

This work includes other less formal but intuitive readability measures such as word_count, filesize, and reading time. The reading time measure is based on Demberg and Keller's (2008) paper, with an assumption of 14.69ms per character.

3.3 The similarity measures

For the similarity measures, several were used, but two are most consistent with prior literature, the Jaccard and Cosine similarity indexes. The Jaccard Similarity (js_/ tjs_) metric is used to determine the similarity between two text documents. It measures how the two text documents are close to each other, that is, how many common words exist over total words. The mathematical representation of the Jaccard Similarity is:

$$J(doc_1, doc_2) = \frac{doc_1 \cap doc_2}{doc_1 \cup doc_2}$$

Similarly, the Cosine similarity (tcs_) measures the cosine of the angle between two vectors projected in a multi-dimensional space. The vectors are typically non-zero and are within an inner product space. The cosine similarity is mathematically described as the division between the dot product of vectors and the product of each vector's euclidean norms or magnitude.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

Other measures obtained include the Sørensen-Dice index (tsd_), similar to the Jaccard index. The Tversky index (ttv_) is a generalization of Jaccard (alpha=0.5, beta=2.0) and Sørensen-Dice (alpha=0.5, beta=1.0). The Bag approach (tbs_) can be considered an edit distance approximation. The Euclidean Similarity (es_) and Compression Similarity (cs_) were also included.

3.4 Determinants of annual report readability

This section briefly discusses the determinants of annual report readability. These variables are employed as controls in our regression estimation.

Size: Size captures many aspects of a firm's operational and business environment. For instance, the accounting literature has used firm size as a proxy for a firm's political cost (e.g., Watts and Zimmerman, 1986). This variable was also employed by Li (2008). The paper calculates size as an indicator variable if a firm's total asset is greater than the median total assets for each year.

Market-to-book: High market-to-book firms differ from low market-to-book firms in many aspects, including the investment opportunity set and growth potential. Li (2008)

employed this variable as well. Growth firms may have more complex and uncertain business models, culminating in readability differences.

Special items: Firms with significant special items are more likely to experience some unusual events. SI is included as a potential determinant of annual report readability.

Accruals: We include the operating activities net of cash flow as a proxy for the accrual activities of a firm. Increases in cash are presented as positive numbers. Decreases appear as negative numbers. The expectation is that firms with more significant accruals relative to their net income and operating cash flow have more activities that give rise, culminating in differing readability, especially if employed in earnings management.

Goodwill: We include goodwill as a proxy for the acquisition activities of a firm. Firms with merger and acquisition activities and, thus, related goodwill on their balance sheet affect financial statement readability.

In addition, the paper includes year and industry-fixed effects as potential determinants of readability. Large non-negative variables are logged. All resulting variables, excluding indicator variables, are then standardized.

3.5 Summary statistics

Table 1 below presents the summary statistics of the sample. Overall, the annual readability score of the NTFS is worse than those of the MD&A. This result is similar across several scores, such as the Flesch reading ease score, the Smog index, and the Fog index. Similarly, the NTFS has a higher word count (16,796 vs10,608) and reading time (1,305 seconds vs. 815 seconds) measure than the MD&A, indicative on the surface of more information content.

For the similarity stats, the NTFS has more similarities than the MD&A. This implies more instances of information repetition in the NTFS than in the MD&A. The most critical measure is the Jaccard similarity score (0.786 vs. 0.717) and the cosine similarity score (0.877 vs. 0.829).

Table 2 in the appendices presents the correlation matrix for the readability and similarity variables. In the MD&A sections, while the Flesch RE score is highly correlated with the Flesch Kincaid score (0.92), the Fog index (0.72), Smog Index (0.87), it is weakly correlated with the Dale Chall Index (0.28). However, in the NTFS sections, the Flesch RE score is highly correlated with the Flesch Kincaid score (0.99), the Fog index (0.98), and the Dale Chall Index (0.74), and weakly correlated with the Smog Index (-0.06). The correlation between both sections across the measures is low, indicating that readability is inconsistent across sections within 10-K. This is in line with the results of Li (2008), who present a 0.227 correlation coefficient for the MD&A and NTFS.

Both the Jaccard and Cosine similarity stats are highly correlated with each other. However, across both sections, the correlation weakens at 0.48 for the Jaccard index and Cosine similarity at 0.482. This again shows that similarity, like the readability stats, is inconsistent across sections within 10-K.

Furthermore, to add to the observation above, the author performs a paired t-test for readability and similarity measures across both sections. The results, as shown in table 3 below, show that the readability and similarity scores are statistically significantly different between item 7 and item 8 of the 10-Ks.

Table 1: Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
filesize mda	116336	65626.487	67641.423	3000	1346040
wordcount mda	116336	10608.935	10921.124	352	224277
flesch re mda	116,336	-40.26895	7.10109	-75.3	88.71
flesch kincaid mda	116336	12.856	1.998	5.5	58.6
fog index mda	116336	10.773	2.236	4.74	61.18
smog index mda	116336	15.093	1.572	9.6	37.9
dale chall mda	116336	6.999	.938	1.14	16.03
reading time mda	116336	815.264	840.915	36.97	16561.699
filesize ntfs	75761	104765.3	84776.533	3004	1660600
wordcount ntfs	75761	16796.876	13688.144	148	262318
flesch re ntfs	75761	-36.464	22.966	-76.01	715.5
flesch kincaid ntfs	75761	14.427	8.71	5.5	303.6
fog index ntfs	75761	11.884	8.918	4.09	307.74
smog index ntfs	75761	16.103	1.604	0	45.4
dale chall ntfs	75761	6.966	1.421	1.23	43.8
reading time ntfs	75761	1305.877	1054.439	37.25	20701.52
js mda	86539	.676	.166	.007	1
es mda	86539	.99	.007	.921	1
cs mda	86539	.331	.298	.003	.971
tjs mda	86539	.717	.155	.034	1
tsd mda	86539	.825	.121	.065	1
ttv mda	86539	.881	.082	.26	1
tcs mda	86539	.829	.113	.178	1
tbs mda	86539	.742	.168	.007	1
js ntfs	56546	.745	.14	.007	1
es ntfs	56546	.993	.004	.929	1
cs ntfs	56546	.203	.268	.003	.965
tjs ntfs	56546	.786	.121	.014	1
tsd ntfs	56546	.874	.09	.027	1
ttv ntfs	56546	.918	.052	.386	1
tcs ntfs	56546	.877	.083	.118	1
tbs ntfs	56546	.798	.144	.007	1

Please see 3.2 and 3.3 above for the legend of the variables above.

Table 3: Paired T-Test

	Sample	Diff	Pr(| T| > | t|)
READABILITY			
filesize_mda == filesize_ntfs	61,267	-25135.28	0
wordcount_mda == wordcount_ntfs	61,267	-3954.618	0
flesch_re_mda == flesch_re_ntfs	61,267	-3.113187	0
flesch_kin_mda == flesch_kin_ntfs	61,267	-1.329424	0
fog_index_mda == fog_index_ntfs	61,267	-0.9637989	0
smog_index_mda == smog_index_ntfs	61,267	-1.064712	0
dale_chall_mda == dale_chall_ntfs	61,267	-0.0368908	0
reading_time_mda == reading_time_ntfs	61,267	-316.4117	0
SIMILARITY			
tjs_mda	44,515	-0.061	0
tcs_mda	44,515	-0.042	0

Fig 1A and 1B below show the mean of the readability scores of the annual reports for the sample firms over time. Similarly, Fig 1C and 1D plot the mean level of the similarity scores of the annual reports for the sample firms over time. The FRES was not included because it has a larger scale. However, the correlation coefficients already show that it is highly correlated with the Flesch-Kincaid score.

An interesting observation from the first panel is that sharp declines in readability for the NTFS do not correspond with the MDA. For example, in 2018, there was a significant decline in NTFS readability that was not mirrored in the MD&A. Similarly, for the similarity measures, there was a significant decline in similarity for the NTFS in 2019 (there over five accounting standard changes: revenue, employee share-based payment Accounting, cash flows, measurement of inventory, classification of deferred taxes and lease accounting). In contrast, a significant decline in the similarity scores for the MD&A section occurred in 2020, possibly due to the pandemic and new information being introduced.

Fig 1A and 1B: MD&A and NTFS Readability Scores

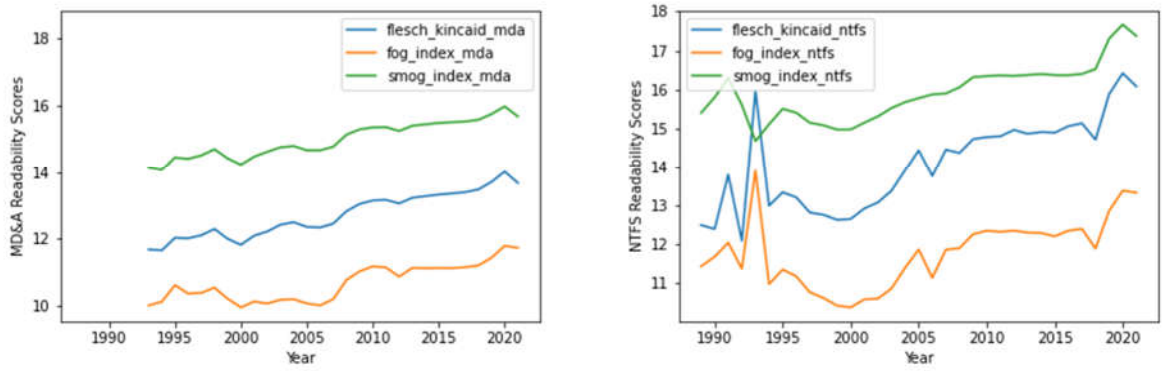
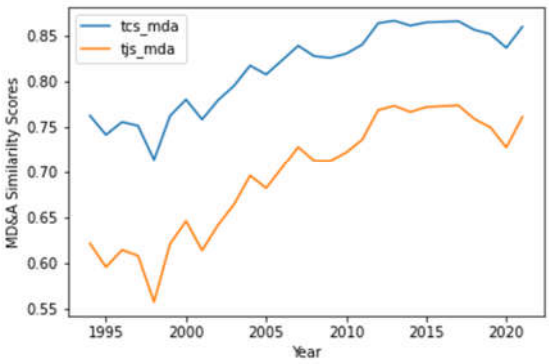
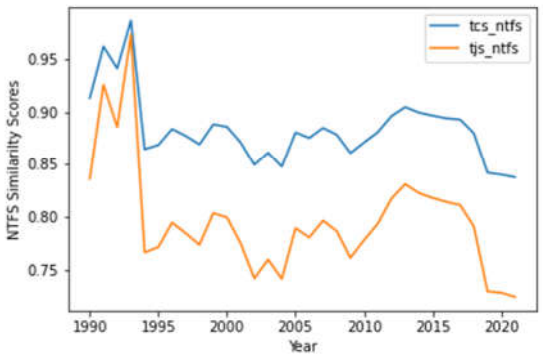


Fig 1A and 1B: MD&A and NTFS Similarity Scores





4. Empirical Results

4.1 MD&A Readability and Audit vs. Firm outcomes.

The author first checks the relationship between the readability score of the MD&A section of the 10-Ks and three audit outcomes (whether or not an auditor is a big four firm, whether the financial statements are qualified and whether the internal control report is qualified) and two financial outcomes (net income and earnings before interest and tax).

Table 4: Determinants of Readability Scores (MD&A Section)

Variable	Flesch RE	Flesch RE	Flesch RE	Flesch RE	Flesch RE	Flesch RE
	(1)	(2)	(3)	(4)	(5)	(6)
big_four	-0.00896					0.0202
qualified		0.0222*				0.00131
qualified_ic			0.0317*			0.0348*
ni				-0.016		-0.0301**
ebit					-0.00273	0.0234**
Size	0.0132	0.0239	0.0228	0.0117	0.0118	0.0386*
Accruals	-0.0107**	-0.0120**	-0.0108**	-0.00634	-0.0101	-0.00838*
Goodwill	-0.000623	-0.00352	-0.0053	0.00313	0.000177	-0.00941
Special Items	-0.00417*	-0.00463**	-0.00379*	0.00303	-0.00428**	0.00966
MTBV	-0.000463	-0.000476	0.0497	-0.000457	-0.000455	0.0644
Observations	65,283	53,179	46,995	65,615	65,565	34,653
Number of Firms	10,420	8,716	7,805	10,458	10,456	5,458
Within R-squared	0.0438	0.0454	0.0535	0.0437	0.044	0.064
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Clustered Error (SIC)	Yes	Yes	Yes	Yes	Yes	Yes
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

The Flesch RE score is multiplied by a negative one to conform with other readability score basis. A higher Flesch Re score is equivalent to increased difficulty comprehending text.

Columns 1 to 5 above present regression results with each audit and financial outcome as the primary independent variable. Column 6 presents a regression model with all relevant audit, financial, and control variables included. All regression models are estimated with robust standard errors clustered at the industry level (SIC classification).

As expected, the audit outcome variables' coefficient is not statistically significant at the 5% confidence interval. In column 6, the financial outcomes variables are significant at the 5% confidence interval. Here, one standard deviation increase in net income decreases the readability score by 0.03 standard deviation, while one standard deviation increase in EBIT increases the readability score by 0.02 units. This result implies that MD&A readability varies in relation to the financial bottom line.

On the other hand, in table 5 below, the NTFS is statistically significantly related to the audit outcomes but not the financial outcomes. Column 1 to 3 below shows that readability declines (an increase in readability score) if an audit firm is a big four firm, if the financial statements are qualified, or if the internal control report is qualified. Here also, but not for the MD&A section, the firm's size is consistently significantly related to the readability of the NTFS; larger firms have worse readability (higher readability score) than smaller firms.

Table 5: Determinants of Readability Scores (NTFS Section)

Variable	Flesch RE (1)	Flesch RE (2)	Flesch RE (3)	Flesch RE (4)	Flesch RE (5)	Flesch RE (6)
big_four	0.0249***					0.0218*
qualified		0.0337***				0.0277**
qualified_ic			0.0148**			0.00745
ni				0.000217		0.000645
ebit					-0.00114	-0.00141
Size	0.0165**	0.0198**	0.0176**	0.0179***	0.0181***	0.0203**
Accruals	-0.00444	-0.00474	-0.00487*	-0.00447*	-0.00418*	-0.00515*
Goodwill	0.00613	0.00888	0.000647	0.0064	0.00661	0.00191
Special Items	-0.00271	-0.00276	-0.0023	-0.00282	-0.00276	-0.00259
MTBV	0.00979	0.00849	-0.0265**	0.0113	0.0113	-0.0252***
Observations	42,215	42,215	30,980	42,485	42,448	24,169
Number of Firms	7,002	7,002	5,197	7,038	7,034	3,838
Within R-squared	0.212	0.212	0.236	0.213	0.212	0.219
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Clustered Error	Yes	Yes	Yes	Yes	Yes	Yes
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

The Flesch RE score is multiplied by a negative one to conform with other readability score basis. A higher Flesch Re score is equivalent to increased difficulty comprehending text.

Tables 6 and 7 below present the result of a similar test above, but with respect to the cosine similarity score of each section of the 10-K. As mentioned above, we expect the scores of the NTFS to be associated with the audit outcomes but not the financial outcomes. Column 6 in table 7 shows that the NTFS similarity score is not significantly related to the financial outcomes but is related to the audit outcomes (except the auditor variable). The relationship with audit outcomes is negative, meaning that when the financial statements or internal control report is qualified, the similarity score of the associated 10-K decreases, evidence of increasing information.

Surprisingly, the MD&A similarity score is also related to the audit outcomes. This means that when the auditors render a negative opinion on the financial statement, the MD&A section changes significantly to introduce more information, possibly to explain and massage user expectations. Contrary to the results in table 7, the MD&A section's similarity score is significantly associated with financial outcomes. The relationship is

positive. This implies that the better the performance of a firm, the less information it discloses in the MD&A section compared with prior years; evidence information declines in the 10-K.

Table 6: Determinants of Similarity Scores (MD&A Section)

Variable	Cosine Similarity (1)	Cosine Similarity (2)	Cosine Similarity (3)	Cosine Similarity (4)	Cosine Similarity (5)	Cosine Similarity (6)
big_four	0.0573**					0.039
qualified		-0.0337***				-0.0301**
qualified_ic			-0.204***			-0.201***
ni				0.0317***		0.00143
ebit					0.0473***	0.0427**
Size	-0.0458**	-0.0305*	-0.0587***	-0.0425**	-0.0427**	-0.0478**
Accruals	0.00206	0.00367	0.00188	-0.00437	-0.00297	-0.00129
Goodwill	-0.0250**	-0.0287**	-0.0164	-0.0302***	-0.0325***	-0.0268**
Special Items	0.000314	0.000289	-0.000445	-0.0143*	0.000727	-0.000355
MTBV	-0.101***	-0.103***	0.132*	-0.101***	-0.101***	0.0867**
Observations	53,363	53,363	39,481	53,610	53,569	30,968
Number of Firms	9,049	9,049	6,912	9,078	9,077	5,121
Within R-squared	0.102	0.102	0.136	0.102	0.1017	0.081
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Clustered Error (SIC)	Yes	Yes	Yes	Yes	Yes	Yes
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Determinants of Similarity Scores (NTFS Section)

Variable	Cosine Similarity (1)	Cosine Similarity (2)	Cosine Similarity (3)	Cosine Similarity (4)	Cosine Similarity (5)	Cosine Similarity (6)
big_four	0.0125					0.0074
qualified		-0.0758***				-0.0657***
qualified_ic			-0.232***			-0.209***
ni				0.00149		-0.00162
ebit					0.0102	0.00933
Size	-0.0375	-0.0261	-0.0394	-0.0344	-0.0345	-0.0299
Accruals	0.00218	-0.00254	0.00284	0.00197	0.000816	-0.00264
Goodwill	-0.0245*	-0.0371***	-0.0138	-0.0246*	-0.0259*	-0.0254***
Special Items	-0.00213	-0.00458*	-0.00284	-0.00269	-0.00213	-0.00448
MTBV	-0.421**	-0.420**	0.00241	-0.428**	-0.428**	0.00175
Observations	34,281	29,923	26,003	34,484	34,454	21,481
Number of Firms	5,817	5,146	4,440	5,837	5,835	3,522
Within R-squared	0.106	0.1118	0.1603	0.106	0.106	0.179
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Clustered Error	Yes	Yes	Yes	Yes	Yes	Yes
Robust SE	Yes	Yes	Yes	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.1

5. Conclusion

Based on the results of our tests, we can conclude that items 7 and 8 of the 10-Ks are significantly different based on the correlation of readability and similarity scores. Similarly, the mean scores for both sections are significantly different, with the NTFS having higher similarity based on Jaccard similarity and cosine similarity indexes but lower readability based on Flesch reading scores, Fog, and Smog indexes. The readability score of the NTFS responds more to audit outcomes than financial outcomes. On the other hand, the readability of the MD&A responds to changes in financial outcomes and not audit outcomes. Since the NTFS is a large portion of the 10-K reports and responds to audit and not financial outcomes, prior and future papers that have used the entire 10-K to measure financial outcomes may have been capturing unrelated metrics.

This paper is still a working document and will include further analysis to test its hypothesis. Such tests include: risk keyword count in the MD&A and NTFS to see the relationship in risk information disclosure between management and auditors. To add to its regression tests, the author intends to collect more data on accounting outcomes, such as audit fees. Similarly, since the results have shown that qualified financial statements have poorer readability, topic modeling will be used to examine if specific issues are discussed in such 10-Ks. There can also be an extension to implement sentiment analysis for item 8 of the 10-Ks to understand if sentiment from audit actors is related to firm performance.

REFERENCES

Biddle, Gary C., Gilles Hilary, and Rodrigo S. Verdi. 2009. "How Does Financial Reporting Quality Relate to Investment Efficiency?" *Journal of Accounting and Economics* 48 (2–3): 112–31. https://doi.org/10.1016/j.jacceco.2009.09.001.

Blanco, Belen, Paul Coram, Sandip Dhole, and Pamela Kent. 2021. "How Do Auditors Respond to Low Annual Report Readability?" *Journal of Accounting and Public Policy* 40 (3): 106769. https://doi.org/10.1016/j.jaccpubpol.2020.106769.

Boritz, J. Efrim, B. Louise Hayes, and Lev Timoshenko. 2016. "Determinants of the Readability of SOX 404 Reports." SSRN Scholarly Paper. Rochester, NY. https://papers.ssrn.com/abstract=3975817.

Brown, Stephen V., and W. Robert Knechel. 2016. "Auditor-Client Compatibility and Audit Firm Selection: AUDITOR-CLIENT COMPATIBILITY AND AUDIT FIRM SELECTION." *Journal of Accounting Research* 54 (3): 725–75. https://doi.org/10.1111/1475-679X.12105.

Cohen, Lauren, Christopher Malloy, and Quoc Nguyen. 2020. "Lazy Prices." *The Journal of Finance* 75 (3): 1371–1415. https://doi.org/10.1111/jofi.12885.

Daines, Robert. 2001. "Does Delaware Law Improve Firm Value?" *Journal of Financial Economics* 62 (3): 525–58. https://doi.org/10.1016/S0304-405X(01)00086-1.

Demberg, Vera, and Frank Keller. 2008. "Data from Eye-Tracking Corpora as Evidence for Theories of Syntactic Processing Complexity." *Cognition* 109 (2): 193–210. https://doi.org/10.1016/j.cognition.2008.07.008.

Francis, Jere R., Matthew L. Pinnuck, and Olena Watanabe. 2014. "Auditor Style and Financial Statement Comparability." *The Accounting Review* 89 (2): 605–33.

Hrazdil, Karel, Dan A. Simunic, and Nattavut Suwanyangyuan. 2021. “Auditor Choice and the Informativeness of 10-K Reports.” *Journal of Accounting, Auditing & Finance*, December, 0148558X2110624. https://doi.org/10.1177/0148558X211062430.

Johnston, Joseph A., and Joseph H. Zhang. 2021. “Auditor Style and Financial Reporting Similarity.” *Journal of Information Systems* 35 (1): 79–99. https://doi.org/10.2308/isys-18-046.

Kim, Chansog (Francis), Ke Wang, and Liandong Zhang. 2019. “Readability of 10-K Reports and Stock Price Crash Risk.” *Contemporary Accounting Research* 36 (2): 1184–1216. https://doi.org/10.1111/1911-3846.12452.

Li, Feng. 2008. “Annual Report Readability, Current Earnings, and Earnings Persistence.” *Journal of Accounting and Economics* 45 (2–3): 221–47. https://doi.org/10.1016/j.jacceco.2008.02.003.

Miller, Brian P. 2010. “The Effects of Reporting Complexity on Small and Large Investor Trading.” *The Accounting Review* 85 (6): 2107–43.

Watts, Ross L., and Jerold L. Zimmerman. 1990. “Positive Accounting Theory: A Ten Year Perspective.” *The Accounting Review* 65 (1): 131–56.

APPENDIX

Table 2: Readability and Similarity Scores Pairwise Correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
(1) filesize_mda	1																	
(2) flesch_re_mda	0.051***	1																
(3) flesch_kincaid_mda	0.034***	0.927***	1															
(4) fog_index_mda	-0.256***	0.727***	0.887***	1														
(5) smog_index_mda	0.012***	0.871***	0.965***	0.896***	1													
(6) dale_chall_mda	-0.713***	0.280***	0.367***	0.688***	0.405***	1												
(7) reading_time_mda	1.000***	0.052***	0.034***	-0.256***	0.012***	-0.712***	1											
(8) filesize_ntfs	0.651***	0.069***	0.038***	-0.210***	0.009**	-0.555***	0.651***	1										
(9) flesch_re_ntfs	0.144***	0.313***	0.323***	0.214***	0.320***	-0.029***	0.144***	-0.008**	1									
(10) flesch_kincaid_ntfs	0.102***	0.278***	0.335***	0.272***	0.335***	0.034***	0.101***	-0.027***	0.993***	1								
(11) fog_index_ntfs	-0.069***	0.237***	0.328***	0.370***	0.339***	0.224***	-0.069***	-0.085***	0.981***	0.996***	1							
(12) smog_index_ntfs	0.082***	0.312***	0.371***	0.317***	0.389***	0.077***	0.082***	0.035***	-0.060***	-0.102***	-0.120***	1						
(13) dale_chall_ntfs	-0.450***	-0.036***	0.003	0.204***	0.026***	0.463***	-0.450***	-0.507***	0.747***	0.775***	0.816***	-0.122***	1					
(14) reading_time_ntfs	0.651***	0.068***	0.036***	-0.212***	0.007*	-0.556***	0.651***	1.000***	-0.008**	-0.028***	-0.086***	0.032***	-0.507***	1				
(15) tjs_mda	0.132***	0.019***	0.018***	-0.040***	0.026***	-0.125***	0.132***	0.078***	0.129***	0.132***	0.101***	0.145***	-0.040***	0.077***	1			
(16) tcs_mda	0.139***	0.007**	0.002	-0.064***	0.007**	-0.149***	0.139***	0.085***	0.122***	0.122***	0.087***	0.132***	-0.049***	0.085***	0.991***	1		
(17) tjs_ntfs	0.039***	-0.101***	-0.106***	-0.115***	-0.100***	-0.079***	0.039***	0.005	0.031***	0.040***	0.042***	-0.033***	0.060***	0.005	0.483***	0.486***	1	
(18) tcs_ntfs	0.037***	-0.103***	-0.109***	-0.123***	-0.106***	-0.086***	0.037***	0.009**	0.022***	0.030***	0.031***	-0.034***	0.044***	0.009**	0.474***	0.482***	0.992***	1

**** p<0.01, ** p<0.05, * p<0.1*