ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Accounting, Auditing and Control

Track: Accounting and Auditing

Textual Analysis on the Extended Audit Report: The Effect on Readability and Standardization for Dutch Firms

Name student: Myrthe Manders

Student ID number: 456189

Supervisor: dr. Yijun Li

Second assessor: dr. Jihun Bae

Final Version: 20 July, 2021

The goal of this paper is to evaluate the decisions made by the IAASB when revising the ISA 700 and introducing ISA 701, using hand-collected data on Dutch firms. As a result of the revision of ISA 700 and the introduction of ISA 701, the readability of the audit report for Dutch firms improved, as measured by the Fog Index, Flesch Index and Flesch-Kincaid Index. In addition, it appears Big Four auditors use more complex English language in their audit report. Through dictionary-based quantitative text analysis, it shows that the proportion of emotion words, thus the level of sentiment, increased as a result of the extension of the audit report, mainly caused by an increase in the proportion of words classified as positive, negative, and uncertain. Finally, the average document similarity measured using the Jaccard Index, is found to be higher for the old audit reports compared to the extended audit reports, suggesting the level of boilerplate text decreases and the level of entity-specific information increases as a result of the new standards. These findings seem to confirm that the objectives of the IAASB by revising ISA 700 and introducing ISA 701 are reached.

Keywords:
ISA 700, ISA 701, audit report, readability, document similarity, Fog Index, Flesch Reading Ease Score, Flesch-Kincaid Readability Score, Dictionary-Based Text Analysis, Jaccard Index

Table of content

1. Introduction

During the financial crisis in 2008, there was a decline in trust in financial reporting and the demand rose for more detailed and transparent audit reports. Not only was the auditor's opinion relevant to users of the financial statements anymore, but they also demanded information on the way the audit was carried out (Quick, 2012). As a reaction to the crisis, the European Commission released a proposal report to improve audit quality and bring back confidence in the financial reporting system (EC, 2010). The audit report should communicate additional information to the shareholders and other interested parties, such as possible risks in the future for the auditee and the audit methodology, according to the European Commission (2010).

The International Auditing and Assurance Standards Board (IAASB) used this Green Paper released by the European Commission to set up an international plan to improve financial reporting, focusing on the audit report. As a result, the IAASB released a new International Standard on Auditing (ISA) which requires the auditor to include the Key Audit Matters (KAM) in the independent auditor's report, with the similar name ISA 701 *Communicating Key Audit Matters in the Independent Auditor's Report* (IAASB, 2015b). In addition, the IAASB revised ISA 700 *Forming an Opinion and Reporting on Financial Statements* (IAASB, 2015a). ISA 700 (Revised) and ISA 701 are effective for audits of financial statements for periods ending on or after December 15, 2016 (IAASB, 2015a; IAASB 2015b).

One of the major changes when ISA 700 was revised, was aimed at improving the readability of the audit report (IAASB, 2015a). For instance, the text in the extended audit report is less standardized or parts of the text are moved to the appendix. The aim of ISA 701 is to decrease the level of standardization in the audit report, to improve the level of transparency and increase the level of entity-specific information (PwC, 2014). Therefore, the audit report should become more valuable and understandable to investors and other users of the financial statement. This leads to the research question: *Does the extension of the audit report through the revision of ISA 700 and implementation of ISA 701 affect the readability and the level of standardization of the audit report for Dutch firms?*

Over the years, numerous measures have been used as proxies for annual report readability. Li (2008) introduced the Fog Index in the accounting literature, which has become a common measure in accounting research to calculate the level of formal education a person needs to understand the text on the first reading. Other common readability measures used in accounting literature are the Flesch Reading Ease Score and Flesch-Kincaid Readability Score, hereafter Flesch Index and Flesch-Kincaid Index respectively (Jones & Shoemaker, 1994). In addition, as one of the goals of the revision of ISA 700 and introduction of ISA 701 was to decrease the level of standardized text and increase the level of entity-specific information in the audit report (PwC, 2014), the expectation rises that the extended audit report contains relatively more words with a specific tone, such as positive or negative, compared to the old audit report. The level of sentiment in the audit report is measured through dictionary-based quantitative text analysis, using the Loughran and McDonald Master Dictionary (2011) containing words classified as positive, negative, uncertain, litigious, constraining, and superfluous. The proportion of these words, also known as emotion words, is expected to be higher in the audit reports adhering to the revised ISA 700 and new ISA 701. Finally, the similarity between the old audit reports is expected to be higher, compared to the similarity between the extended audit reports, which is measured by the Jaccard Index. To apply these research methods, audit reports for Dutch firms adhering to the old standards and following the revised ISA 700 and new ISA 701 are collected by hand.

The results of the regression analysis show that the Fog Index and the Flesch-Kincaid Index are lower when the new standards on the audit report are applied, while the Flesch Index is higher for audit reports following the new standards. As a higher Fog Index and Flesch-Kincaid Index and a lower Flesch Index represent more complex text, this suggests the audit reports written following the new standards contain less complex English language compared to audit reports following the old standards. These results imply that the extension of the audit report improves the readability of audit reports for Dutch firms, ceteris paribus. The regression analyses on the readability indexes also show that, ceteris paribus, Big Four auditors use more complex English language in their audit reports compared to non-Big Four auditors, as the Fog Index and Flesch-Kincaid Index are higher while the Flesch Index is lower for audit reports written by Big Four auditors. Additional Big Four and non-Big Four subsample regression analysis shows that the improvement in readability of the audit report as a result of the revision of ISA 700 and introduction of ISA 701 is larger for audit reports written by non-Big Four auditors compared to audit reports written by Big Four auditors.

In addition, following the dictionary-based text analysis with the Loughran and McDonald Master Dictionary (2011), the audit reports adhering to the revised ISA 700 and new ISA 701, including and excluding stop words, show a significantly higher proportion of emotion words. This suggests that the extension of the audit report increased the level of sentiment in the audit report, making it more likely the level of boilerplate text in the audit report decreased. The increase in the proportion of emotion words in the audit report is mainly caused by the increase in positive and negative words, and words that display uncertainty. Also, the document similarity analysis, with the Jaccard Index as similarity coefficient, shows the average similarity between the old audit reports is 44.8% higher compared to the similarity coefficient of the extended audit reports, also suggesting the level of boilerplate text decreases as a result of the extension of the audit report.

Previous research mainly focuses on the readability of the annual report as a whole, but hardly ever examines the readability of the audit report only. However, the annual report is composed by the company being audited, while the independent auditor's report is written by an auditor. Therefore, this analysis of the audit report solely is an important addition to the existing accounting literature, as the readability of the audit report might be affected by different factors than the annual report as a whole, as the auditor and auditee have different objectives. The goal of the auditor is to give an opinion on the financial statements, while the goal of the company is to present their business most advantageous. Other important additions to existing accounting literature are the application of dictionary-based quantitative text analysis and document similarity analysis, which is not very common yet in accounting literature, especially the usage of the Jaccard Index is quite innovative. Finally, this paper includes a unique sample as all audit reports used in the analyses are hand-collected.

The findings of this research are an important evaluation of the decisions previously made by accounting standard setters. The goal of the revision of ISA 700 and the introduction of ISA 701 is to improve financial reporting, provide more transparent audit reports and improve the readability of the audit report (PwC, 2014), which was demanded after the financial crisis in 2008. This research examines whether the IAASB responded effectively to the demand that rose after the financial crisis, by evaluating the effect of the revision of ISA 700 and the implementation of ISA 701 on the readability and level of sentiment of the audit report. The findings from the dictionary-based quantitative text analysis and document similarity analysis are also important for users of the financial statements, as it provides an insight in the level of boilerplate text used in the independent auditor's report. When the level of standardized text is high, not much entity-specific information is included in the audit report, thus the audit report might not be very useful for, for instance, investors or stock market analysts. However, when the audit report includes more entity-specific information, the audit report might become more relevant for these investors or stock market analysts.

This paper continues with the theoretical framework, elaborating on the relevant changes in the audit report caused by the revision of ISA 700 and the introduction of ISA 701. The theoretical framework also includes a discussion of multiple readability measures, sentiment measurement through dictionary-based text analysis and the introduction of the Jaccard Index. Following, the sample selection and the research design are discussed. Next, the results on the audit report readability, the dictionary-based text analysis and document similarity analysis are presented and elaborated. Finally, a conclusion will be given, with an answer to the research question and suggestions for further research.

2. Theoretical framework

2.1 (Revised) ISA 700

Before defining and assessing the readability of the audit reports, it is important to understand the major changes in ISA 700. This standard focuses on the responsibility of the auditor to form an opinion on the financial statements and how this should be presented in the independent auditor's report (IAASB, 2015a). This is the first objective of the standard, but the focus goes towards the second objective of the standard, which is to express that opinion clearly through a written report (IAASB, 2015a).

The first change in the new audit report is that the opinion section is placed at the beginning of the audit report, instead of in the middle, to make it immediately clear to the user of the financial statements what the auditor's opinion is (PwC, 2014). After the opinion section, the extended audit report continues with a section describing the basis for this opinion. This part consists of standardized text most of the time, as the applicable law and standards and the independence of the auditor are stated. However, new in the audit report is the section on the materiality used during the audit and on what items the performance materiality is based (IAASB, 2015a). Another new part is the explanation on the scope of the group audit, as most of the large companies audited are head of a group of entities. This gives the user of the financial statements an insight in the audit approach used in the other audits of entities of the group (IAASB, 2015a). After the paragraph on the scope, the Key Audit Matters are explained. As this is not part of the revised ISA 700, but included in the new ISA 701, this will be discussed in the next paragraph. The audit report continues with two paragraphs on the responsibilities of the Board of Directors for the financial statements and the auditor's responsibilities for the audit of the financial statements. Especially this last aspect is expanded in the new audit report (IAASB, 2015a). The choice can be made to describe the auditor's responsibilities extensively in the main section of the audit report, but it can also be included in an appendix. The audit report ends with a section on other legal and regulatory requirements. According to the revised ISA 700, the appointment of the auditor should be described in this part to give the user an indication of changes in auditors (IAASB, 2015a).

2.2 ISA 701: Key Audit Matters

The new standard, ISA 701, is an addition to the standard discussed above. ISA 701 focuses on the responsibility of the auditor to communicate Key Audit Matters in the independent auditor's report (IAASB, 2015b). The Key Audit Matters show the most important matters that came to light during the audit or matters that have received significant attention by the auditors (IAASB, 2015b). The section on the Key Audit Matters is often one of the major parts in the extended audit report and should decrease the level of standardized text, as it shows entity-specific information (PwC, 2014).

PwC (2014) analyzed the audit reports of several Dutch listed entities that already included Key Audit Matters in their report before the new standard became effective. The Dutch Association for Public Accountants (NBA) already issued guidelines on the inclusion of the Key Audit Matters in the independent auditor's report before the IAASB (2015b) officially released the new ISA 701 in 2015 (NBA, 2014). PwC (2014) found that a total of 68 Key Audit Matters were included in 17 auditor's reports they analyzed, with matters varying from the valuation of goodwill till derivatives or the IT environment, depending on the nature of the entity. This implies that the inclusion of the Key Audit Matters section in the auditor's report enhances the level of entity-specific information.

The example of the audit report issued by the NBA (2014), adhering to the revised ISA 700 and new ISA 701, is presented in Appendix A. However, ISA 700 (Revised) and ISA 701 need to be effective for audits of financial statements for periods ending on or after December 15, 2016 (IAASB, 2015a). The timeline of the extension of the audit report is shown in Figure 1.

Figure 1

Timeline of the revision of ISA 700 and the introduction of ISA 701

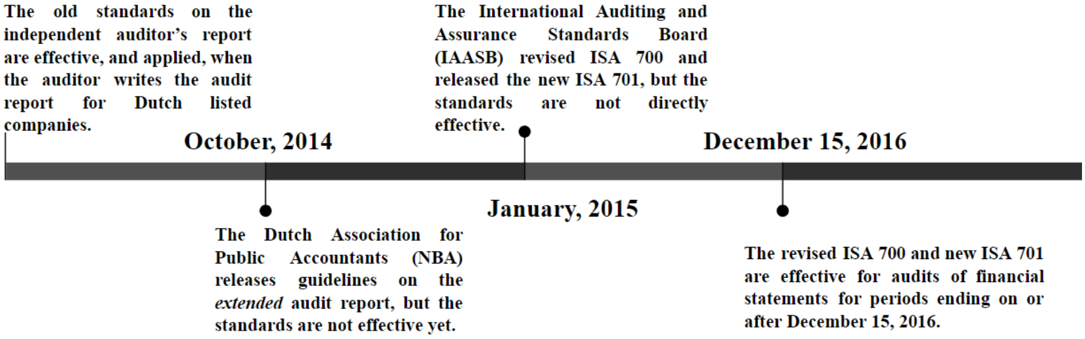

Figure 1 shows a timeline including when the old standards on the audit report were effective, when the NBA released new guidelines on the audit report, and when the revised ISA 700 and new ISA 701 became effective.

2.3 Defining readability

Over the years, not much research focuses solely on audit report readability, however numerous measures have been used as proxies for annual report readability. Li (2008) introduced the Fog Index in the accounting literature, which has become a common measure to calculate the level of formal education a person needs to understand the text on the first reading. Other common readability measures used in accounting literature are the Flesch Reading Ease Score and Flesch-Kincaid Readability Score (Jones & Shoemaker, 1994). The Fog Index and Flesch Reading Ease Score are also used in research by Velte (2018) on audit report readability of UK firms, which is one of the few studies that solely focuses on audit report readability instead of annual report readability.

2.3.1 The Fog Index

The first measure of readability of the independent auditor's report that is used in this paper, is the Fog Index. Before the Fog Index is introduced in accounting literature by Li (2008), the index was traditionally used to measure readability to distinguish grade school textbooks, as the Fog Index equation calculates the number of years of education someone needs to understand the text on the first reading (Loughran & McDonald, 2016). Therefore, a higher Fog Index implies a lower readability of the text. The Fog Index consists of two variables and is calculated as:

$$Fog\ Index = 0.4 * (average\ sentence\ length + percentage\ of\ complex\ words) \quad (1)$$

To measure the readability of the annual report, and the relation between annual report readability and earnings and firm performance, Li (2008) introduces the Fog Index in the accounting literature. Li (2008) finds that companies with lower earnings have annual reports with a higher Fog Index, thus more complex annual reports, and that companies with annual reports with a lower Fog Index, thus easier annual reports, have more persistent positive earnings.

Lehavy, Li and Merkley (2011) examine the readability of companies' written communication on the behavior of sell-side financial analysts, using the Fog Index as readability measure for corporate 10-K filings. They find that companies with 10-K filings that have a higher Fog Index, have more analyst following and more amount of effort incurred to generate the reports. Also, they find that companies with less readable 10-K reports have more informative reports. If this would be the same for the audit reports of Dutch firms, this would counter the objective of the revision of ISA 700 and introduction of ISA 701, to increase the informativeness of the audit report.

Lawrence (2013) also focuses on the readability of 10-K filings and investors, in particular retail investors, again using the Fog Index as a readability measure. Lawrence (2013) finds that retail investors are more likely to invest in firms when they have shorter, and more readable, 10-K filings. In addition, Miller (2010) finds that companies with more readable written documents have more noticeable small investor trading activity just before and after the filing date.

Lo, Ramos and Rogo (2017) explore how the annual report readability varies with earnings management, using the Fog Index to measure readability. In contrast to multiple other studies, Lo et al. (2017) focus on the Management Discussion and Analysis (MD&A) section instead of the annual report as a whole, when calculating the Fog Index. They find that firms that are most likely involved in earnings management to beat prior year's earnings have MD&A sections that are less readable. As the MD&A section mostly consist of only a couple of pages, just as the audit report, this suggests the Fog Index can also be used to calculate short texts such as audit reports.

Loughran and McDonald (2014) also assess 10-K filings' readability, using the Fog Index as one of their readability measures as they state this is the most commonly applied readability measure. However, Loughran and McDonald (2014) find that the 10-K document file size is a more appropriate readability measure compared to the Fog Index. Nevertheless, as the main interest of this paper is the *difference* between the readability before and after the new standards on the audit report, the Fog Index is expected to be an appropriate readability measure analyzing Dutch audit reports. This leads to the first hypothesis:

H1: Firm-years adhering to the revised ISA 700 and new ISA 701 will have audit reports with a lower Fog Index than firm-years with audit reports adhering to the old standards, ceteris paribus.

2.3.2 Flesch Indexes

Alternative readability measure commonly used in accounting literature, which have some similarity with the Fog Index, are the Flesch Reading Ease Score and the Flesch-Kincaid Readability Score (Jones & Shoemaker, 1994). The Flesch Reading Ease Score, hereafter Flesch Index, is calculated as:

$$Flesch\ Index = 206.835 - (1.015 * average\ sentence\ length) - (84.6 * \frac{number\ of\ syllables}{number\ of\ words}) \quad (2)$$

The Flesch Index is inversely related to the Fog Index, meaning that a low Flesch Index represents a complex text while a high Flesch Index means a text is easy to read (Flesch, 1948).

In 1975, the Flesch-Kincaid Readability Score, hereafter Flesch-Kincaid Index, was introduced, which is a derivation of the Flesch Index. The Flesch-Kincaid Index presents a U.S. grade level, thus a lower Flesch-Kincaid Index implicates the text is easier to read (Kincaid et al., 1975). As the Fog Index generates a level of education as well, the Flesch-Kincaid Index and Fog Index are comparable measures of readability. Moreover, when the Flesch-Kincaid Index is larger than 10.0, the index can also be interpreted as the number of years of education required to understand a text, similar to the Fog Index (Kincaid et al., 1975). However, just as with the Flesch Index, the Flesch-Kincaid Index includes the number of syllables of all words and not only the number of words with at least three syllables. The formula used to calculate the Flesch-Kincaid Index is:

$$Flesch - Kincaid\ Index = (0.39 * average\ sentence\ length) + \left(11.8 * \frac{number\ of\ syllables}{number\ of\ words}\right) - 15.59 \quad (3)$$

Even though the Flesch Index and Flesch-Kincaid Index received criticism, they are amongst the most used readability measures in accounting literature (Clatworthy & Jones, 2001). Also, both Flesch Indexes are often used simultaneously with the Fog Index. For instance, Lewis, Parker, Pound and Sutcliffe (1986) use the Fog Index and multiple Flesch Indexes as readability measures for annual reports. However, they only use a small sample including only four years, and they report no change in the complexity of the annual reports over these years. However, Jones (1988) does find a significant downward trend in annual report readability over time, when studying the period 1952-1985.

De Franco, Hope, Vyas and Zhou (2015) use the Fog Index, Flesch Index and Flesch-Kincaid Index to calculate analyst report readability. Thus, the three mentioned reading indexes seem not only appropriate to calculate annual report readability, but also other documents analyzed in accounting literature. De Franco et al. (2015) find a positive relation between the readability of analyst reports, proxied by combining the Fog Index, Flesch Index, and Flesch-Kincaid Index, and the trading volume reactions on these reports.

Velte (2018) is one of the only ones to solely focus on audit report readability, when investigating the relation between the percentage of women on audit committees and the auditors' disclosures on the key audit matters. In addition, Velte (2018) does not only use the Fog Index, but also the Flesch Index to measure audit report readability. Velte (2018) finds that companies with a higher percentage of women on audit committees have more readable Key Audit Matters sections in the audit reports compared to companies with a low percentage of women on the audit committee.

Because this paper focusses on the comparison of the readability of the audit report before and after the revision of ISA 700 and ISA 701, the Flesch Index is expected to be a suitable readability measure. In addition, as the Flesch-Kincaid Index is a derivative of the Flesch Index, the Flesch-Kincaid Index is expected to be an appropriate measure as well. Because the Flesch Index and Flesch-Kincaid Index are inversely related, this requires two separate hypotheses. The goal of the revision of ISA 700 and the adoption of ISA 701 is to improve the readability of the audit report, thus the second and third hypotheses read:

H2: Firm-years adhering to the revised ISA 700 and new ISA 701 will have audit reports with a higher Flesch Index than firm-years with audit reports adhering to the old standards, ceteris paribus.

H3: Firm-years adhering to the revised ISA 700 and new ISA 701 will have audit reports with a lower Flesch-Kincaid Index than firm-years with audit reports adhering to the old standards, ceteris paribus.

The first three hypotheses are stated in the alternative form, as an improvement in audit report readability is expected, because this is one of the objectives of the revision of ISA 700 and the introduction of ISA 701 (PwC, 2014). However, it might be possible the extension of the audit report does not affect the readability of the audit report at all, when measured by the Fog Index, Flesch Index and Flesch-Kincaid Index. In the majority of the audit reports in the sample used in this paper, the old and extended independent auditor's report is written and signed by the same auditor, thus the same person, making it probable that the choice of words is similar in both reports, which would not change the level of complexity of the audit report. Also, this would probably not change the proportion of emotion words and lexical diversity. However, as on the other hand the audit report is expected to be less standardized, the expectation rises that more of the opinion of the auditor will be visible in the audit report, through an increased use of words having a certain sentiment, which does not necessarily change the complexity of the text. The predicted change in sentiment in the audit report is explored with dictionary-based quantitative text analysis, which is described in the next section.

2.4 Dictionary-based quantitative text analysis

The Fog Index and Flesch Indexes mainly focus on an individual's ability to understand a message. However, Bag-of-Words Methods attempt to computationally obtain meaning from a text (Loughran & McDonald, 2016). One common Bag-of-Words Method uses a "dictionary", through which a researcher can count words that are classified as, for instance, positive, negative, or modal, also known as dictionary-based text analysis. The main advantage of using dictionaries to measure tone is that once a dictionary is selected, possible subjectivity of the researcher is avoided (Loughran & McDonald, 2016). In addition, the use of dictionaries that are publicly available enables the replication of the research.

One of the goals of the revision of ISA 700 and introduction of ISA 701 is to decrease the level of standardization of text in the audit report (PwC, 2014). As a result, the expectation arises that more of the professional opinion of the auditor can be seen, when reading the audit report. In that case, the audit report would include more words reflecting an opinion in the extended audit report, compared to the old audit report, such as “confident”, “efficient”, “fraudulent” or “uncertain” (Loughran & McDonald, 2011). In other words, the expectation rises that more sentiment or tone is present in the new audit reports compared to the old audit reports. The sentiment in a text can be measured by counting the words in a text that are, for instance, classified as positive or negative in a dictionary, and dividing the sum by the total number of words in a text. These words that can be classified with a certain sentiment are referred to as emotion words hereafter, and the classification of words can be found in specific dictionaries. Loughran and McDonald (2016) compare multiple dictionaries in their review paper, with each having their own advantages and disadvantages. As a result of the limitations of the Harvard and Diction Sentiment Word Lists, Loughran and McDonald (2011) computed their own dictionary with financial communication in mind, including 354 positive and 2,355 negative words. In addition, they classify words as uncertainty, litigious, constraining, and superfluous. To analyze the fourth hypothesis, the Loughran-McDonald Master Dictionary (2011) will be used:

H4: The proportion of emotion words in the audit reports of firms adhering to the revised ISA 700 and new ISA 701 is higher than the proportion of emotion words in the audit reports of firms adhering to the old standards, ceteris paribus, following the Loughran-McDonald Master Dictionary.

2.5 Document similarity

Finally, as the goal of the revision of ISA 700 and the introduction of ISA 701 is to decrease the level of boilerplate text and increase the level of entity-specific information (PwC, 2014), the expectation rises that the level of similarity between the extended audit reports is lower compared to the similarity between the old audit reports.

Morais and Fialho (2008) assess whether harmonized accounting standards lead to harmonized accounting practices, focusing on IAS 39 measurement requirements in some countries of the European Union. They use the Jaccard similarity coefficient, hereafter Jaccard Index, to determine the level of harmonization between IAS 39 and the financial reporting practice of European listed companies. They find a high level of harmonization between the accounting practices of European firms and IAS 39.

Besides the research performed by Morais and Fialho (2008), not much research in the accounting literature focuses on textual analysis methods that measure textual similarity of annual reports. In addition, hardly any research examines the textual similarity between audit reports. Therefore, it is uncertain now whether the Jaccard Index is an appropriate measure to compare document similarity of the old audit reports with the average document similarity between the extended audit reports. However, as the Jaccard Index is used by Morais and Fialho (2008) to compare accounting practices as a result of a certain standard, this index should be applicable as a measure for document similarity in this paper. As the goal of the revision of ISA 700 and the application of ISA 701 is to decrease boilerplate text, the final directional hypothesis is presented:

H5: The similarity between the audit reports of firms adhering to the revised ISA 700 and new ISA 701 is lower than the similarity between the audit reports of firms adhering to the old standards, ceteris paribus.

3. Research design

3.1 Sample selection

The revised ISA 700 and the new ISA 701 are effective for audits of financial statements for periods ending on or after December 15, 2016 (IAASB, 2015a; IAASB, 2015b), as shown in the timeline in Figure 1. However, numerous companies already included aspects of the new report in the independent auditor's report before ISA 700 (Revised) and ISA 701 were effective (PwC, 2014). For example, PwC already discussed the Key Audit Matters and the application of materiality in the audit of Ahold (2015) over the fiscal year 2014 in the audit report. The auditors of PwC used the guidelines issued by the Dutch Association for Public Accountants (NBA) on the new independent auditor's report, that was released in 2014 (NBA, 2014). Therefore, for the sample selection, it is not appropriate to compare the auditor's reports for financial statements for periods ending before December 15, 2016, with the auditor's reports for financial statements for periods ending on or after December 15, 2016. Per company, it should be examined whether the auditor pre-adopted the guidelines released by the NBA (2014) on the extended audit report.

The sample selection process starts with retrieving all Dutch listed firms with available data between 2013 and 2016 from Compustat Global. The first fiscal year is 2013, because in this year the NBA (2014) did not release the guidelines on the extended auditor's report yet. The last fiscal year included in the sample is 2016, because from this year on the new standards on the auditor's report are effective (IAASB, 2015a; IAASB, 2015b). The sample should only include Dutch firms that have adopted the International Standards on Auditing. ISA is adopted by NBA Netherlands, as issued by the IAASB without modifications, including the effective dates, for the conduct of all financial statement audits (IFAC, 2019). This means that practitioners need to apply ISA 700 (Revised) and ISA 701 in their audit. All firms with domestic standards not in accordance with IFRS or firms following other standards such as U.S. GAAP are removed from the sample. Next, all firms in financial services industries are removed from the sample, following Lo et al. (2017), because they have a different financial and operating structure. Firms with annual reports, audit reports, or both, that are only publicly available in Dutch are removed from the sample as well, because the Fog Index, Flesch Index and Flesch-Kincaid Index are designed for English language (Loughran & McDonald, 2016). Also, companies with annual reports only including the old or the extended audit report are removed, as comparison between the two reports is not possible in that case. In addition, not all firms currently left in the sample are available in I/B/E/S. For those firms, it is not possible to calculate the SUE Score, thus the firms are removed from the sample. Finally, firms with negative book value of equity are removed from the sample, as the natural logarithm of the market-to-book ratio is undefined for negative ratios. The sample selection process is summarized in Table 1.

When the companies are selected for the sample that are appropriate for the regression analysis, dictionary-based quantitative text analysis and additional document similarity analysis, the audit reports of these companies are hand-collected. For each company, two audit reports are included in the final sample; the last audit report following the old standards, and the first audit report following the revised ISA 700 and new ISA 701, or the guidelines released by the NBA (2014) on the extended audit report. Therefore, in the final sample, 178 audit reports are included that are hand-collected.

Table 1

Sample selection

	Sample size
Dutch listed firms in Compustat Global 2013-2016	166
Less: firms using accounting standards not in accordance with IFRS (no IFRS/ISA)	(16)
Less: firms in financial services	(40)
Less: firms with annual reports and/or audit reports only in Dutch	(7)
Less: firms with only the old or extended audit report (publicly) available	(4)
Less: firms not available in I/B/E/S	(7)
Less: firms with negative book value of equity	(3)
Number of firms	89

Table 1 shows the total number of firms initially included in the data sample. Next, it shows characteristics which makes it necessary to exclude firms from the sample. The final sample consists out of 89 Dutch firms.

3.2 Reading indexes

The three reading indexes used to calculate audit report readability are the Fog Index, Flesch Index, and the Flesch-Kincaid Index. First, the Fog Index is calculated as:

$$Fog\ Index = 0.4 * (average\ sentence\ length + percentage\ of\ complex\ words) \qquad (1)$$

The number of complex words is calculated as the number of words of three or more syllables, that are *not* proper nouns, combinations of easy words or hyphenated words, or two-syllable verbs made into three with -ed and -es endings (Li, 2008). When interpreting the Fog Index, the Fog Index presents the number of years of formal education required to understand a text on the first reading (Loughran & McDonald, 2016). The Fog Index is expected to be an appropriate measure for audit report readability, as the index is not expected to be affected by the change in length of the audit report in general, because it focusses on the average sentence length and the percentage of complex words, instead of absolute numbers of words, which can be seen in other measures of readability (Loughran & McDonald, 2014).

One of the major caveats of using the Fog Index as a measure for readability in accounting research is that frequently occurring words, such as financial, company, operations, management, employees, and customers, which are easily understood by investors or other users of the financial statements, are identified as complex words (Ertugrul et al., 2017). However, in this paper, the annual reports before the revision of ISA 700 and the introduction of ISA 701 are compared to the extended audit reports adhering to the new standards. Therefore, it is no problem that words such as financial or company are identified as complex words, as the main interest of this paper is the *difference* between the readability before and after the new standards on the audit report, measured by the Fog Index.

Secondly, the Flesch Index is used to measure audit report readability, which is calculated as:

$$Flesch\ Index = 206.835 - (1.015 * average\ sentence\ length) - (84.6 * \frac{number\ of\ syllables}{number\ of\ words}) \qquad (2)$$

The Flesch Index calculates a score of a text between 0 and 100. When the Flesch Index equals 0, the text is considered practically unreadable, where a Flesch Index of 100 indicates the text is easy to read for any literate person (Flesch, 1948). Nowadays, a text is considered to be comprehensible for the majority of the readers when the Flesch Index is higher than 50 (Stone & Parker, 2013). In contrast to the Fog Index, the Flesch Index does take into account the number of syllables of all words, instead of only including the number of words with three or more syllables. However, just as for the Fog Index, the Flesch Index does not take into account the structure of the text.

The third readability measure is a derivation of the Flesch Index, the Flesch-Kincaid Index. However, because of the different weighting factors for the number of words per sentence and the number of syllables in each index formula, the Flesch Index and Flesch-Kincaid Index are not directly convertible (Jones & Shoemaker, 1994). The Flesch-Kincaid Index is calculated as:

$$Flesch - Kincaid\ Index = (0.39 * average\ sentence\ length) + \left(11.8 * \frac{number\ of\ syllables}{number\ of\ words}\right) - 15.59 \quad (3)$$

The Flesch-Kincaid Index represents a U.S. grade level, thus a low Flesch-Kincaid Index implies a text is easy to understand, while a high Flesch-Kincaid Index represents a complex text. When the Flesch-Kincaid Index is higher than 10.0, it can be interpreted as the number of years required to understand the text, similar to the Fog Index. The lowest grade level score in theory equals -3.40, however this is very uncommon (Kincaid et al, 1975). In addition, the formula does not have an upper bound, but as the readability of audit reports is compared and the score on itself is not the main focus, this does not cause problems.

3.3 Paired sample t-tests

As a result of the revised ISA 700 and the introduced ISA 701, the audit report should become more informative, and the readability should be improved. By using the Fog Index, Flesch Index and Flesch-Kincaid Index as measures of readability, it can be assessed whether the extension of the audit report changes the readability of the report. By comparing the readability indexes of the audit report of a company following the old standards with the audit report of the same company in the first year adhering to the new standards, the effect of the revised ISA 700 and new ISA 701 can be assessed. Through paired sample t-tests, it is analyzed whether the Fog Index, Flesch Index and Flesch-Kincaid Index before and after the new standards differ significantly, testing Hypothesis 1, 2 and 3. In order to perform the paired sample t-tests, the assumption of normality is checked through density plots of the differences per firm-pair for each readability index. Additionally, quantile-quantile plots are presented to verify the normal distribution assumption.

3.4 Regression analysis

After performing the paired sample t-tests, regression analysis is performed to include the control variables when testing the effect of the extension of the audit report on the Fog Index, Flesch Index and Flesch-Kincaid Index. To test Hypothesis 1, 2 and 3, the general regression equation will be used:

$$Readability_{i,t} = \beta_0 + \beta_1 EAR_{i,t} + \Sigma\beta_j ControlsVariables_j + \varepsilon \quad (4)$$

where *EAR* (Extended Audit Report) is a binary variable, taking on the value 1 after the adoption of the revised ISA 700 and new ISA 701, and 0 otherwise. The Fog Index, Flesch Index and Flesch-Kincaid Index are used as measures for readability per firm (i) per year (t). The coefficient β_1 is the coefficient of interest and is expected to be negative for the regression models with the Fog Index and the Flesch-Kincaid Index; the extended audit report should lead to a lower Fog Index and a lower Flesch-Kincaid Index, suggesting the audit report is less complex. On the other hand, the coefficient β_1 is expected to be positive for the model with the Flesch Index, as the Flesch Index is inversely related to the Flesch-Kincaid Index. Thus, the extension of the audit report is expected to lead to a higher Flesch Index, implicating the audit report is easier to understand after adhering to the new standards.

3.5 Control variables

Previous research did not extensively focus solely on audit report readability, but mainly on annual report readability in general. As the audit report is included in the annual report, the audit report is also included in research on the readability of the annual report. Therefore, a number of the control variables used in research on the readability of the annual report are expected to affect the readability of the audit report as well. Annual report readability could also be expected to be correlated with audit report readability, for instance annual reports with a higher Fog Index could be expected to have a higher Fog Index for the audit report as well. However, because it is unlikely there is a relation between the revision of ISA 700 and implementation of ISA 701, and the annual report readability indexes, this is not included as a control variable. In addition, one of the few papers solely measuring audit report readability also does not control for annual report readability (Velte, 2018).

Firstly, Li (2008) finds a negative relation between reported earnings and annual report readability, using the Fog Index as a readability measure. In other words, Li (2008) finds that the Fog Index is higher for firms with lower earnings. For the control variable *Earnings*, Compustat Global is used, and the variable is scaled by the book value of total assets. Following Li (2008), all values below -1 and above +1 are considered outliers and should be removed from the sample, however all values in the final sample fall within the accepted range.

Loughran and McDonald (2014) use the natural logarithm of the firm's size (*log(Size)*) measured through the number of shares outstanding in millions multiplied by the price per share at the end of the fiscal year and the natural logarithm of the market-to-book ratio (*log(MTB)*) to control for the size of and investors' perception of a firm. The natural logarithm of the market-to-book ratio is only defined for positive ratios, therefore firms with a negative market-to-book ratio are removed from the sample. Also, following Loughran and McDonald (2014), the market-to-book ratio is winsorized at the 1% level. The number of shares outstanding and price per share are retrieved from Compustat Security Daily, and the total equity of each company is retrieved from Compustat Global.

In addition, Loughran and McDonald (2014) use the absolute value of the standardized unexpected earnings (*abs(SUE)*) as a control variable when applying alternative readability measures. The SUE Score is calculated as the actual earnings per share minus the analyst consensus earnings per share, divided by the standard deviation. In I/B/E/S, the variable SUE Score is already available, it only needs to be transformed to absolute values as the magnitude of the earnings surprise is of interest and not the sign of the SUE Score (Loughran & McDonald, 2014). I/B/E/S Unadjusted data files are used to avoid rounding issues. Again, following Loughran and McDonald (2014), the variable is winsorized at the 1% level to handle outliers.

Next, what could impact the changes in readability of the audit report is when the audit report is written by a different audit firm compared to the previous audit report adhering to the old standards, as each audit firm is expected to use their own templates for the auditor's report. For instance, in 2013, BDO audited the financial statements of Amsterdam Commodities N.V., but in 2014 PwC became the company's auditor (Amsterdam Commodities N.V., 2014; Amsterdam Commodities N.V., 2015). Not only did the content of the independent auditor's reports change as a result of the new standards, but the layout of the report also changed as well which can be attributed to the change of audit firm. As a result, the changes in the lay-out might influence the Fog Index or Flesch Indexes. Therefore, following Velte (2018) a binary variable (*AuditChange*) is included that takes on the value 1 when the change in audit firm happened simultaneously with the company adhering to the new ISA 701 and revised ISA 700, and 0 otherwise.

One of the few studies that does solely focus on audit report readability includes a binary variable (*Big4*) that takes on the value 1 when the audit report is written by a Big Four auditor, and 0 otherwise (Velte, 2018). As Big Four firms have more industry-specific knowledge and more resources, compared to non-Big Four firms, this is expected to affect the words used in the audit report.

The 48-industry dummies of Fama and French (1997) are also included to control for differences in readability caused by the industry, as some industries are expected to be more complex than others which decreases readability following the measures above (Loughran & McDonald, 2014).

Finally, year-fixed effects are included in the regression analysis, because for the 89 companies in the sample, not all fiscal years are corresponding. For instance, for Akzo Nobel N.V., fiscal years 2013 and 2014 are included, while for Qiagen N.V. fiscal years 2015 and 2016 are included. Therefore, the variable *EAR* does not already capture year-fixed effects, thus year-fixed effects are included in the regression analysis. The definitions of all control variables are described in Appendix B, Table B1.

3.6 Dictionary-based quantitative text analysis

After performing the paired sample t-tests and the regression analysis on the Fog Index, Flesch Index and Flesch-Kincaid Index, additional dictionary-based quantitative text analysis is performed. As already mentioned, the Loughran and McDonald Master Dictionary (2011) will be used for this analysis, which contains 354 words classified as positive, 2,355 negative words, 297 showing uncertainty, 904 litigious words, 184 classified as constraining and 56 superfluous words.

In order to compare the dictionary with each audit report, the texts need modification. All punctuation, numbers and extra whitespaces are removed from the text files, in order to just keep the words of each audit report. Next, all audit reports are duplicated and for one of each pair, stop words, such as “the”, “and”, or “of”, are removed. This creates one sample with all regular audit reports, and another sample with all text files containing the audit reports without stop words.

Before using the dictionary, some basic calculations are performed on the text files. First, the total number of words per text including stop words (*TotalWords*) and excluding stop words (*TotalWordsNSW*) are calculated. Secondly, the total number of unique words per text including stop words (*UniqueWords*) and the number of unique words per text without stop words (*UniqueWordsNSW*) are calculated. Next, the lexical diversity per text can be calculated by dividing the total number of unique words by the total number of words per text. This is done for the sample containing text files with stop words (*LexDiv*) and the sample with the files excluding stop words (*LexDivNSW*).

After the basic calculations on the text are performed, the Loughran and McDonald Master Dictionary (2011) is used to calculate the emotion words per text. As the dictionary does not contain stop words, it does not matter if these calculations are performed on the text files including or excluding stop words. Per text, the total number of positive, negative, uncertain, litigious, constraining, and superfluous words is calculated. These six numbers are added together per audit report, forming the total number of emotion words per text (*EmotionWords*). As this absolute number is not very informative on itself, the number of emotion words is compared to the total number of words per text. The total number of emotion words per text is divided by the total number of the text including stop words (*ProportionEmotion*) and excluding stop words (*ProportionEmotionNSW*). These two variables show what proportion of the text consists of emotion words. Following Hypothesis 4, it is expected that these variables are higher for companies with audit reports adhering to the new standards.

Additionally, the proportion of emotion words is split into the six categories of the sentiments of the Loughran and McDonald Master Dictionary (2011). The proportion of positive words is calculated as the number of positive words divided by the number of emotion words per text (*ProportionPos)*, and the same is done for the negative words (*ProportionNeg*), uncertainty words (*ProportionUnc*), litigious words (*ProportionLit*), constraining words (*ProportionCon*), and superfluous words (*ProportionSup*).

Through paired sample t-tests, it is analyzed whether the proportion of emotion words before and after the new standards are statistically different, testing Hypothesis 4. In addition, paired sample t-tests are performed for the other variables in the dictionary-based text analysis, to evaluate in what sentiment category the possible changes in emotion words are most provoke.

3.7 Jaccard similarity coefficient

An additional textual analysis method is used to assess the similarity between the old audit reports, compared to the similarity between the extended audit reports. The widely used R package textreuse by Muller (2016) is used, which provides multiple functions to measure similarity among documents and to detect passages that have been reused. Of the old and the extended audit reports, pairwise comparison is performed. The mean of the similarity scores of the old reports and the mean of the similarity scores of the extended reports is calculated and compared. As the goal of the revision of ISA 700 and the implementation of ISA 701 is to decrease the level of boilerplate text (PwC, 2014), the expectation is that the average similarity score of the extended audit reports is lower than the average similarity score of the old audit reports, following Hypothesis 5.

The similarity between audit reports is calculated using the Jaccard Index, also known as the Jaccard similarity coefficient, which is a statistic used to measure similarity and diversity in sample sets (Niwattanakul et al., 2013). The Jaccard Index is defined as the size of intersection divided by the size of the union in the sample sets:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A|+|B|-|A \cap B|} \qquad (5)$$

The Jaccard Index can take on values between 0 and 1, where a value of 1 means the similarity is perfect, and 0 indicates there is no similarity present at all (Niwattanakul et al., 2013). When performing all pairwise comparisons, two matrices are formed; one with all Jaccard Indexes between the 89 old reports, and a matrix with all Jaccard Indexes between the 89 new reports. Each matrix will have 89 * 89 = 7,921 Jaccard Indexes. When no similarity is found between two documents, NA will be assigned in the matrix by the textreuse package (Mullen, 2016). To calculate the average Jaccard Index per matrix, all NAs are replaced by zero. Eventually, the means of both matrices are compared to analyze whether the overall document similarity has changed as a result of the revision of ISA 700 and the implementation of ISA 701.

4. Results

The reading indexes for the audit reports before and after adhering to the revised ISA 700 and new ISA 701 are visualized in Figure 2. It shows that the average Fog Index and Flesch-Kincaid Index are lower for the extended audit report, and the average Flesch Index is higher for the extended audit report compared to the old audit report, which gives support for Hypothesis 1, 2 and 3. Also, the distribution of all readability indexes has decreased when the new standards are followed. Figure 2 shows fewer extreme values for the extended audit report (*EAR* = 1) compared to the old audit report (*EAR* = 0), especially fewer extreme high values. This suggests that, following the new standards, there are no extreme complex reports released anymore. Next, the descriptive statistics and Pearson correlation matrix are assessed, and paired sample t-tests and regression analyses are performed, to find further support for Hypothesis 1, 2 and 3.

Figure 2

Reading Indexes before and after the revised ISA 700 and new ISA 701 are applied

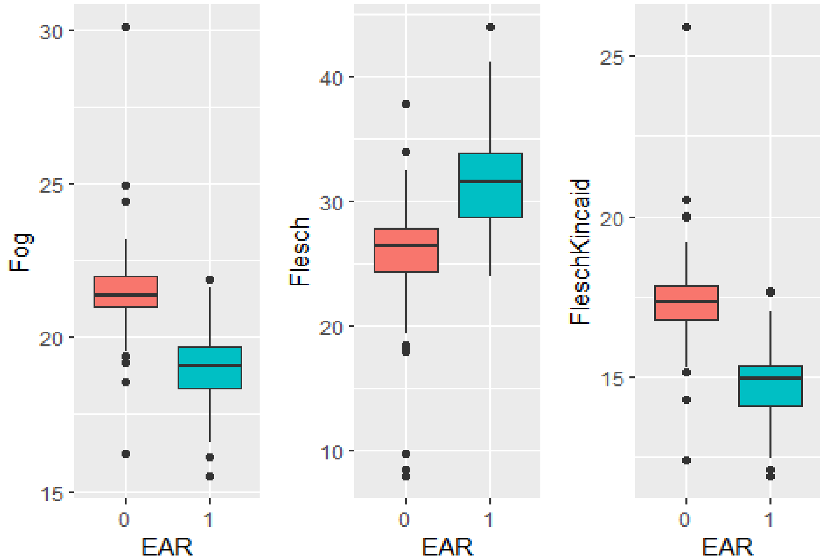

Figure 2 shows boxplots for the Fog Index, Flesch Index and Flesch-Kincaid Index, for the old audit reports (EAR = 0) and the extended audit reports (EAR = 1) separately. The boxplots show the distribution and skewness of the reading indexes through displaying the data quartiles, medians, and potential outliers.

4.1 Descriptive statistics

Table 2 shows the descriptive statistics of the reading indexes and the numeric control variables used in the regression models. In the full sample, shown in Panel A, 178 observations are present, implicating all 89 firms are present in the sample with two observations per firm. This is confirmed in Panel B and C, where Panel B includes all firm-year observations using the old audit report (*EAR* = 0) and Panel C includes all observations adhering to the new standards on the audit report (*EAR* = 1). No missing values are present anymore in the sample, as all missing values are replaced using predictive mean matching. This method is appropriate here, as the number of missing values before the application of predictive mean matching is low (Landerman et al., 1997). In addition, the outliers for the variables *log(MTB)* and *abs(SUE)* are removed using winsorization at the 1% level, following Loughran and McDonald (2014). The variable *Earnings* also does not show extreme values, because the variable is scaled by the book value of total assets, following Li (2008). In addition, if the earnings scaled by total assets exceed +1 or are smaller than -1, Li (2008) removed these outliers from the sample. However, all values of *Earnings* fall within the accepted range here. Finally, the variable *log(Size)* did not show any extreme values, thus it did not need to be winsorized or scaled. Also, Loughran and McDonald (2014) and Lo et al. (2017), who also include the variable *log(Size)* in their data, do not winsorize or scale this variable.

Table 2

Descriptive statistics reading indexes

Panel A: Full sample

Variables	N	Mean	Median	St. Dev.	Min	Q1	Q3	Max
Fog	178	20.522	20.321	1.816	15.492	19.066	21.382	30.075
Flesch	178	28.660	28.233	5.065	7.931	25.972	31.650	43.868
FleschKincaid	178	16.061	16.064	1.839	11.914	14.884	17.332	25.884
Earnings	178	0.016	0.028	0.134	-0.479	-0.015	0.060	0.678
log(Size)	178	6.451	6.129	2.327	1.549	5.024	8.611	10.685
log(MTB)	178	0.564	0.625	0.858	-2.602	0.154	1.087	2.576
abs(SUE)	178	2.484	0.689	4.853	0.000	0.385	1.914	28.168
Panel B: Sample for *EAR* = 0 (old audit report)								
Fog	89	21.486	21.378	1.460	16.230	20.996	21.982	30.075
Flesch	89	25.850	26.400	4.550	7.931	24.314	27.837	37.851
FleschKincaid	89	17.353	17.322	1.450	12.410	16.789	17.824	25.884
Earnings	89	0.025	0.027	0.142	-0.479	-0.007	0.062	0.678
log(Size)	89	6.457	5.992	2.380	1.823	4.654	8.601	10.685
log(MTB)	89	0.541	0.598	0.824	-2.602	0.154	1.090	2.576
abs(SUE)	89	2.449	0.646	4.780	0.000	0.368	1.916	28.168
Panel C: Sample for *EAR* = 1 (extended audit report)								
Fog	89	19.018	19.077	1.191	15.492	18.380	19.712	21.871
Flesch	89	31.471	31.506	3.861	23.985	28.754	33.857	43.898
FleschKincaid	89	14.769	14.943	1.150	11.914	14.109	15.379	17.688
Earnings	89	0.007	0.028	0.126	-0.389	-0.037	0.058	0.381
log(Size)	89	6.445	6.215	2.287	1.549	5.066	8.614	10.575
log(MTB)	89	0.587	0.716	0.895	-2.602	0.176	1.077	2.576
abs(SUE)	89	2.518	0.694	4.953	0.000	0.391	1.851	28.168

Table 2 shows the descriptive statistics of the reading indexes and numeric control variables used in the regression analyses. Panel A shows the descriptive statistics for the full sample, Panel B shows the descriptive statistics for all observations with audit reports adhering to the old standards (EAR = 0), and Panel C shows the descriptive statistics for all observations with audit reports adhering to the revised ISA 700 and the new ISA 701 (EAR = 1). The variables are defined in Appendix B, Table B1.

For the full sample, the mean of the Fog Index is equal to 20.522, which indicates a person needs 20.522 years of formal education to understand the average audit report on the first reading. This is about two years higher than the average Fog Index commonly found for annual report readability, for instance by Li (2008) and Loughran and McDonald (2014), implicating the auditor's report section requires more years of formal education than the annual report on average. The mean of the Flesch Index for the full sample is equal to 28.660. Interpreting the Flesch Reading Ease Score, a Flesch Index between zero and thirty implies a text is very difficult to read and is best understood by university graduates (Flesch, 1948). The mean of the Flesch-Kincaid Index for the full sample is equal to 16.061, implicating the U.S. grade level of the average audit report is equal to 16.061 (Kincaid et al., 1975).

For Panel B and C in Table 2, the sample is split for firm-years adhering to the old standards on the audit report, and firm-years following the revised ISA 700 and new ISA 701. In Panel B, the average Fog Index for the old audit report equals 21.486, implicating 21.486 years of formal education are required to understand the average old audit report on the first reading. However, Panel C shows the mean of the Fog Index is only 19.018, which means only 19.018 years of formal education are required to understand the extended audit report on the first reading. The lower mean for the Fog Index for the extended audit report compared to the old audit report is in line with Hypothesis 1. The mean of the Flesch Index for the old audit report equals 25.850, shown in Panel B. Following the score table of the Flesch Reading Ease Score again, a Flesch Index of 25.850 indicates the text is very difficult to read and is best understood by university graduates (Flesch, 1948). However, the mean of the Flesch Index of the extended audit report, shown in Panel C, equals 31.471. On the Flesch Reading Ease Score scale, a Flesch Index between thirty and fifty the text is difficult to read and is best understood by college graduates (Flesch, 1948). Even though the extended audit report is still considered difficult following the Flesch Reading Ease Score scale, the mean Flesch Index is higher for the audit reports following the revised ISA 700 and new ISA 701, which is in line with Hypothesis 2. Thirdly, the mean of the Flesch-Kincaid Index of the old audit report shown in Panel B equals 17.353, while the mean of the Flesch-Kincaid Index of the extended audit report shown in Panel C equals 14.769. This implies the average U.S. grade level is lower for the audit reports following the revised ISA 700 and new ISA 701 (Kincaid et al., 1975). This is in line with Hypothesis 3.

4.1.1 Fama and French 48-industry means

Besides the reading indexes and the continuous control variables, the data sample includes the Fama and French (1997) 48-industry dummies. Table B2 in Appendix B shows the number of firms from the sample per Fama and French (1997) industry, and the corresponding mean of the Fog Index, Flesch Index and Flesch-Kincaid Index for the old audit report and the extended audit report. Not all industries are present in the data sample of the 89 firms, therefore only the industries that are present in the sample are shown in Table B2.

The industry with the least complex old audit reports, on average, is the printing and publishing industry. This industry shows the lowest average Fog Index and Flesch-Kincaid Index, 18.515 and 14.504 respectively, suggesting 18.515 or 14.504 years of education are required to understand the text on the first reading. The printing and publishing industry also shows the highest mean of the Flesch Index of 33.095, for the old audit reports, which still classifies the average audit report in this industry as difficult to read (Flesch, 1948). After the extension of the audit report, the printing and publishing industry does not show the lowest Fog Index and Flesch-Kincaid Index anymore, and also not the highest Flesch Index. When the new standards on the audit report are followed, the consumer goods industry appears to have the most readable audit reports, with an average Fog Index and Flesch-Kincaid Index of 16.373 and 12.428 respectively, suggesting 16.373 and 12.428 years of education are required to understand the average audit report in the industry on the first reading. It also shows the highest mean of the Flesch Index of 39.513, which classifies the average audit report in the consumer goods industry as difficult to read (Flesch, 1948).

The industry with the most complex audit reports before the extension of the audit report, on average, is the communication industry, with an average Fog Index of 25.536, Flesch Index of 18.315 and Flesch-Kincaid Index of 21.321. Thus, 25.536 or 21.321 years of formal education are required to understand the average audit report on the first reading. Moreover, the Flesch Index of 18.315 classifies the average audit report in the communication industry as very difficult to read, best understood by university graduates (Flesch, 1948). After the revision of ISA 700 and the introduction of ISA 701, the communication industry does not show the highest Fog Index and Flesch-Kincaid Index, and the lowest Flesch Index, on average, but the utilities industry does. The utilities industry shows the highest mean of the Fog Index and the Flesch-Kincaid Index after the extension of the audit report, 21.019 and 16.579 respectively, thus 21.019 or 16.579 years of education would be required to understand the audit report on the first reading. It also shows the lowest mean of the Flesch Index equal to 27.428, classifying the average audit report in the utilities industry as very difficult to read, best understood by university graduates (Flesch, 1948). Thus, there does not seem to be a certain industry that consistently has firms with very complex or readable audit reports, when interpreting Table B2 in Appendix B.

4.2 Pearson correlation matrix

In Table 3, the Pearson correlation matrix of the continuous variables and the variable of interest, *EAR*, is presented to analyze the possible linear relation between these variables. The dependent variables are continuous variables, therefore the binary or categorical variables, such as the dummy variable for audit firm change or the Fama-French 48-industry dummies, are excluded from the correlation matrix. Only

Table 3

Pearson correlation matrix reading indexes

	EAR	*Fog*	*Flesch*	*Flesch Kincaid*	*Earnings*	*log(Size)*	*log(MTB)*	*abs(SUE)*
EAR	1.000							
Fog	-0.681***	1.000						
Flesch	0.557***	-0.921***	1.000					
FleschKincaid	-0.705***	0.993***	-0.915***	1.000				
Earnings	-0.084	0.071	-0.106	0.086	1.000			
log(Size)	0.030	0.117	-0.137*	0.114	0.319***	1.000		
log(MTB)	0.022	-0.031	-0.002	-0.023	0.166**	0.264***	1.000	
abs(SUE)	-0.059	0.002	0.017	-0.011	-0.006	-0.249***	-0.169**	1.000

*Table 3 shows the Pearson correlation coefficients of the variable of interest (EAR), the reading indexes and the numeric control variables. Statistical significance is presented as: * p<0.1; ** p<0.05; *** p<0.01.*

the binary variable *EAR* is included in the correlation matrix, as this variable of interest is expected to have a linear relation with the readability indexes.

The binary variable *EAR*, indicating whether the standards on the extended audit report are followed, is strongly correlated with the Fog Index and the Flesch-Kincaid Index, as it shows statistically significant correlation coefficients of -0.681 and -0.705 respectively. This suggests there is a negative linear relation between the adoption of the new standards and the two reading indexes, implicating the audit report becomes less complex when the new standards are followed. In addition, the correlation coefficient between *EAR* and the Flesch Index is statistically significant as well, equal to 0.557. This suggests a moderate positive linear relation between the Flesch Index and the adoption of the extended audit report, also implicating the audit report becomes more readable when the revised ISA 700 and new ISA 701 are applied.

The three reading indexes are strongly correlated, as expected, as the reading indexes are calculated for the same pieces of text and their formulas consist of similar components. The Fog Index and the Flesch-Kincaid Index both show a low value when a text is easy to read. Therefore, their correlation coefficient is expected to be close to +1, which is the case as the correlation coefficient equals 0.993 and is statistically significant. On the other hand, the Flesch Index shows a high value when a text is easy to understand, thus the Flesch Index is expected to be negatively correlated with the Fog Index and the Flesch-Kincaid Index. The correlation matrix in Table 3 shows a statistically significant correlation coefficient of -0.921 for the Fog Index and the Flesch Index, and a statistically significant correlation coefficient of -0.915 for the Flesch Index and the Flesch-Kincaid Index. This indicates the reading indexes are almost perfectly correlated.

Among the control variables, there is a moderate degree of correlation (0.319) between the earnings and the natural logarithm of the size of a company, which is statistically significant as well. This is in line of expectation, as larger companies are expected to have higher earnings (Loughran & McDonald, 2014). The other statistically significant correlation coefficients only show a low degree of correlation between the control variables, such as the correlation coefficient of 0.166 between the earnings and the natural logarithm of the market-to-book ratio.

4.3 Paired sample t-tests

The descriptive statistics in Table 2 and the visualization of the reading indexes in Figure 2 already imply the readability improved for the audit reports adhering to the revised ISA 700 and new ISA 701, however no statistical significance is proved yet. To analyze whether the Fog Index, Flesch Index and Flesch-Kincaid Index before and after the new standards on the audit report are statistically different, paired sample t-tests are performed with a significance level of 0.05. A paired sample t-test requires the differences between the reading indexes before and after the revision of ISA 700 and introduction of

ISA 701 per company to be normally distributed. According to the central limit theorem, the sampling distribution leans towards a normal distribution if the sample is large enough (N > 30). Even though this is the case, as shown in the descriptive statistics in Table 2, density plots are produced to check the assumption of normality. Figure 3 shows three bell-shaped curves for the three reading indexes, implying the assumption of normality holds.

Figure 3

Density plots differences within pairs per readability index

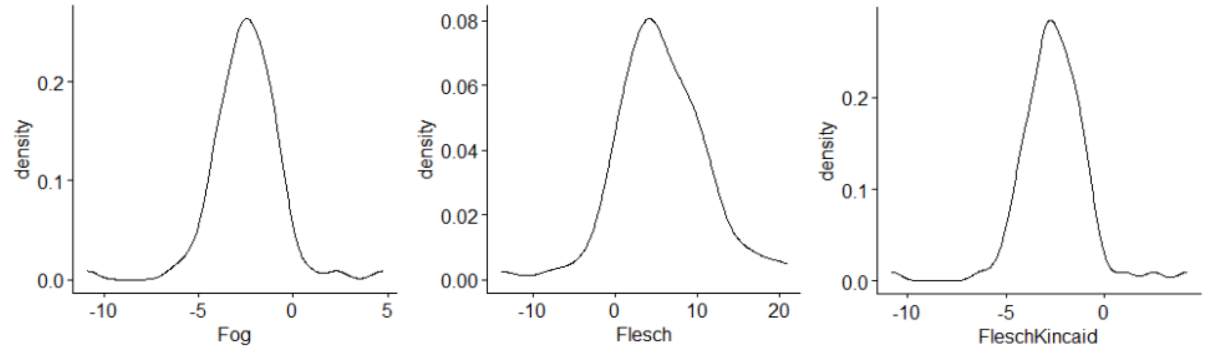

Figure 3 shows the density plots of the differences within pairs per readability index. For the first plot, the difference is calculated as the Fog Index of the extended audit report of firm X, minus the Fog Index of the old audit report of firm X. The same method is followed for all firms and all three reading indexes. A bell-shaped curve suggests the assumption of normality holds.

In addition, quantile-quantile plots are produces to graphically compare two probability distributions per reading index, by plotting the quantiles against each other. When both sets of quantiles have the same distribution, the points should form a line that is approximately straight. Figure 4 shows the quantile-quantile plots for the three reading indexes. Even though the plots for the Fog Index and the Flesch-Kincaid Index show a small number of outliers, normality seems to hold.

Figure 4

Quantile-quantile plots differences within pairs per readability index

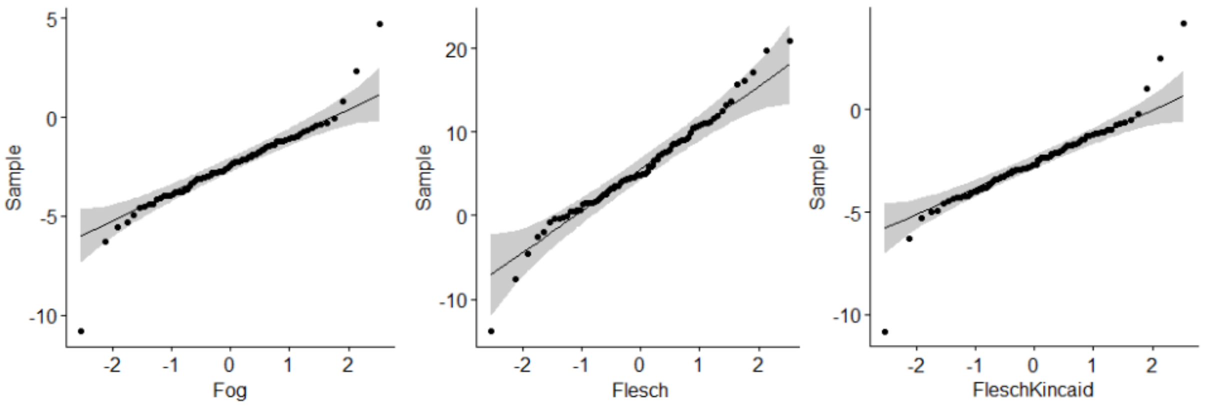

Figure 4 shows the quantile-quantile plots of the differences within pairs per readability index. For the first plot, the difference is calculated as the Fog Index of the extended audit report of firm X, minus the Fog Index of the old audit report of firm X. The same method is followed for all firms and all three reading indexes. When all observations form a line that is approximately straight, it suggests the assumption of normality holds.

After checking for a normal distribution per reading index, the paired sample t-test is performed and presented in Table 4. There is a statistically significant decrease in the Fog Index for the extended audit report (M = -2.468, SD = 1.843) compared to the old audit report, $t(88) = -12.630$, $p < 0.01$. This

Table 4

Paired sample t-test results reading indexes

Reading indexes differences	N	Mean	St. Dev.	[95% Conf. Interval] Lower	Upper	t	df	Sig. (2-tailed)
ΔFog	89	-2.468	1.843	-2.856	-2.079	-12.630	88	0.000***
$\Delta Flesch$	89	5.621	5.475	4.468	6.774	9.687	88	0.000***
$\Delta FleschKincaid$	89	-2.585	1.779	-2.959	-2.210	-13.707	88	0.000***

*Table 4 shows the results of the paired sample t-tests on the reading indexes. The three tests show whether the reading index of the old audit report is statistically different from the reading index of the extended audit report. This is the case when Δ[Reading Index] is statistically different from zero. Statistical significance is presented as: * p<0.1; ** p<0.05; *** p<0.01.*

implicates the audit report adhering to the revised ISA 700 and the new ISA 701 is easier to understand, compared to the old audit report, which supports Hypothesis 1. Also, the Flesch-Kincaid Index shows a decrease when the new standards on the audit report are applied (M = -2.585, SD = 1.779), $t(88)$ = -13.707, p < 0.01. This suggests the extended audit report requires a lower U.S. grade level, meaning the extended audit report is easier to understand, supporting Hypothesis 3. Finally, the Flesch Index shows the inverse effect of the Flesch-Kincaid Index, as the Flesch Index shows a statistically significant increase for the extended audit report compared to the old audit report (M = 5.621, SD = 5.475), $t(88)$ = 9.687, p < 0.01. This result also implies the extended audit report is easier to understand compared to the old audit report, supporting Hypothesis 2.

4.4 Regression analysis

The data sample consists of two-dimensional data, therefore panel data analysis is performed with a significance level of 0.05, using the general linear regression equation:

$$Readability_{i,t} = \beta_0 + \beta_1 EAR_{i,t} + \Sigma\beta_j ControlsVariables_j + \varepsilon \qquad (4)$$

The variable *Readability* is replaced by the Fog Index in the first regression model, the Flesch Index in the second regression model and the Flesch-Kincaid Index in the third regression model. For the group variable *i*, the Global Company Key is used. The time variable *t* refers to two years: the last fiscal year the audit report was issued following the old standards, and the first fiscal year the audit report adheres to the revised ISA 700 and the new ISA 701. The goal of the regression analysis is to assess whether the adoption of the new audit report affects the readability, measured by three different readability indexes, of the audit report. The variable *EAR* takes on the value 0 when the old audit report format is used, and 1 when the new standards are followed. Therefore, the coefficient β_1 is the coefficient of interest, as this coefficient is expected to be different from zero. Besides continuous control variables, fixed-year effects and the Fama and French (1997) 48-industry dummies are included in the regression models.

The results of the regression analysis with the Fog Index as the dependent variable are shown in the first column of Table 5. The results show a statistically significant coefficient for the variable *EAR* (-3.428), which indicates the Fog Index decreases with an absolute value of 3.428 when the new standards on the audit report are followed, ceteris paribus. A decrease of 3.428 in the Fog Index implies that approximately 3.4 years less of formal education are required to understand the text on the first reading, compared to the old audit report, thus the new text is easier to understand. This supports Hypothesis 1. Next, the coefficient of *Big4* (0.726) is statistically significant, however this coefficient shows a positive sign. Interpreting the coefficient of *Big4*, the coefficient implies that the Fog Index for audit reports issued by Big Four firms require on average 0.726 years more of formal education to understand the text on the first reading, compared to audit reports issued by non-Big Four firms, ceteris

Table 5

Regression results on the relation between the extension of the audit report and the Fog Index, Flesch Index and Flesch-Kincaid Index

	(1) Fog	(2) Flesch	(3) FleschKincaid
EAR	-3.428*** (0.000)	8.702*** (0.000)	-3.479*** (0.000)
Earnings	1.145 (0.209)	-4.021 (0.173)	1.281 (0.156)
log(Size)	0.025 (0.714)	-0.188 (0.382)	0.014 (0.828)
log(MTB)	-0.066 (0.602)	0.063 (0.878)	-0.045 (0.721)
abs(SUE)	0.002 (0.749)	-0.012 (0.540)	0.004 (0.556)
AuditChange	-0.317 (0.453)	0.464 (0.734)	-0.180 (0.667)
Big4	0.726* (0.054)	-2.854** (0.020)	0.853** (0.023)
Constant	20.438*** (0.000)	30.181*** (0.000)	16.216*** (0.000)
Observations	178	178	178
R^2	62.6%	49.8%	64.4%
Adj. R^2	52.7%	36.5%	55.0%
F-statistic	6.333*** (df = 34; 143)	3.750*** (df = 34; 143)	6.836*** (df = 34; 143)

*Table 5 shows the regression results using formula (4). The first column shows the regression coefficients and p-values in parentheses using the Fog Index as dependent variable. In the second column, the Flesch Index is used as dependent variable. In the third column, the Flesch-Kincaid Index is used as dependent variable. Statistical significance is presented as: * p<0.1; ** p<0.05; *** p<0.01. All regressions include the 48-industry dummies of Fama and French (1997) and year-fixed effects as well.*

paribus. This would mean Big Four auditors use more complex English language in their audit reports than non-Big Four auditors. The coefficients of the other control variables do not show statistically significant results. Additionally, the regression results on the Fog Index show an R^2 of 62.6%, which means that 62.6% of the variance in the dependent variable is explained by the independent variables. This is higher than the R^2 of the regression analysis on the Fog Index performed by Loughran and McDonald (2014), who find a R^2 of approximately 47%. Even the adjusted R^2 in this regression analysis on the Fog Index, which takes into account the number of independent variables used in the regression model, is higher than the R^2 found by Loughran and McDonald (2014), as the adjusted R^2 for this regression on the Fog Index equals 52.7%, shown in Table 5. The adjusted R^2 of this regression model is also considerably higher than the adjusted R^2 of the well-quoted paper by Li (2008), describing a regression model on the Fog Index, who finds an adjusted R^2 between 6 and 9%. Thus, the data seems to properly fit the regression model shown in the first column of Table 5.

Next, the results of the regression model using the Flesch Index as the dependent variable are shown in the second column of Table 5. Just as for the Fog Index, the coefficient of *EAR* is statistically significant. The coefficient is equal to 8.702, which means that the Flesch Index on average is 8.702 higher for audit reports adhering to the revised ISA 700 and the new ISA 701, compared to the old audit report standards, ceteris paribus. As explained before, the Flesch Index can take on values between 0 and 100, where a Flesch Index of 0 indicates the text is practically unreadable and a Flesch Index of 100 indicates the text is readable for anyone who is literate (Flesch, 1948). Therefore, an increase in the Flesch Index means the text becomes easier to understand. To give an example, a text with a Flesch Index of 45 is classified as difficult to read, and is best understood by college graduates (Flesch, 1948). As a result of the extension of the audit report, following the regression results in Table 5, the new Flesch Index would equal 53.702, which classifies a text as fairly difficult to read instead of difficult to read (Flesch, 1948). Thus, the statistically significant and positive coefficient for the variable *EAR* in the second regression model supports Hypothesis 2. The coefficient for the binary variable *Big4* is statistically significant as well, taking on the value -4.021. This suggests the new audit report has on average a Flesch Index that is 4.021 lower for audit reports issued by a Big Four auditor compared to an audit report issued by a non-Big Four auditor, ceteris paribus. Just as in the first regression model, this suggests that Big Four auditors use more complex English language in their audit reports, compared to non-Big Four auditors. The other control variables do not show any statistically significant coefficients. In addition, the R^2 of the second regression model is equal to 49.8%, which suggests 49.8% of the variance in the Flesch Index is explained by the independent variables. The adjusted R^2 is equal to 36.5%, thus when taking into account the number of independent variables, the variance in the Flesch Index is for 36.5% explained by the independent variables. Both percentages are considerably lower compared to the first regression model on the Fog Index, however the R^2 of the regression model on the Flesch Index and the R^2 found by Loughran and McDonald (2014) of their regression model on the Fog Index are similar, indicating the second regression model properly fits the data as well.

The third regression model uses the Flesch-Kincaid Index as the dependent variable, and the results are presented in the third column of Table 5. Also, for the third regression model, the variable *EAR* shows a statistically significant coefficient, which is equal to -3.479. Thus, the Flesch-Kincaid Index decreases on average by 3.479 when the new standards on the audit report are followed compared to the old standards, ceteris paribus. As a low Flesch-Kincaid Index implicates the text is easy to understand, a decrease in the Flesch-Kincaid Index suggests the readability of the text, so the audit report, is improved. In addition, because all Flesch-Kincaid Indexes in the sample are larger than 10.0, shown in the descriptive statistics in Table 2, the regression coefficient can also be interpreted related to the number of years of education required to understand the text on the first reading (Kincaid et al., 1975). Thus, the regression coefficient of *EAR* of -3.479 suggests the years of education required to understand the audit report on the first reading decreases by 3.479, ceteris paribus, when the new standards on the audit report are followed. This is in line with Hypothesis 3. The coefficient for the binary variable *Big4* is statistically significant as well and is equal to 0.853. This suggests the Flesch-Kincaid Index increases by 0.853 when the audit report is written by a Big Four auditor, compared to a non-Big Four auditor, ceteris paribus. Just as in the first and second regression model, this implicates the complexity of the English language used in the audit report is higher when the report is written by a Big Four auditor. All other control variables in the third regression model do not show statistically significant coefficients. Finally, the third regression model shows the highest R^2 and adjusted R^2 of the three regression models in Table 5, 64.4% and 55.0% respectively. Thus, for the third regression model, 64.4% of the variance in the Flesch-Kincaid Index is explained by the independent variables when not adjusted for the number of independent variables. When taking into account the number of independent variables used in the regression model, 55.0% of the variance in the Flesch-Kincaid Index is explained by the independent variables. So, for all three regression models in Table 5, the data seems to fit the regression models properly. In addition, the F-statistic for all three regression models is statistically significant at 0.01, which implies the regression models fit the data better than models without any independent variables. Conclusively, the regression analysis and the paired sample t-tests show that Hypothesis 1, 2 and 3 cannot be rejected.

4.4.1 Additional Big Four and non-Big Four subsample regression analysis

The only control variable in the previous regression models showing statistically significant coefficients, is the binary variable indicating whether the audit report is written by a Big Four auditor (*Big4*). Therefore, additional subsample regression analysis is performed where the full sample is split in one subsample with all firm-years where the audit report is written by a non-Big Four auditor (*Big4* = 0) and another subsample including all firm-years with audit reports written by Big Four auditors (*Big4* = 1). As this regression analysis is supplemental, the results are presented in Appendix B, Table B3.

Starting with the regression analysis with the Fog Index as readability measure, shown in the first column of Table B3, it appears that the extension of the audit report leads to a larger decrease in complexity of the audit for audit reports written by non-Big Four auditors than by Big Four auditors.

The coefficient of *EAR* of both subsamples is statistically significant, but for the non-Big Four subsample the coefficient is equal to -3.983 while for the Big Four subsample it only equals -3.418. This suggests the readability of the audit report improves more for non-Big Four auditors compared to Big Four auditors, ceteris paribus. This is in line with the findings in the previous regression models, which suggest Big Four auditors use more complex language in their audit reports than non-Big Four auditors.

For the regression analysis with the Flesch Index as dependent variable, presented in the second column of Table B3, similar results are found. Again, of both subsamples the coefficient of *EAR* is statistically significant, but the non-Big Four subsample shows a higher increase in readability compared to the Big Four subsample. The coefficient of *EAR* for the non-Big Four subsample equals 10.230 while the coefficient in the Big Four subsample only equals 8.770. As an increase in the Flesch Index suggests an increase in readability, these results imply the readability for audit reports improves more for audit reports written by non-Big Four auditors than for audit reports written by Big Four auditors, when the new standards are followed.

Finally, the third regression model on the Flesch-Kincaid Index is in line with the other two regression models. The coefficient of *EAR* in both subsamples is statistically significant, but the coefficient is smaller in the non-Big Four subsample than in the Big Four subsample, namely -4.007 and -3.456 respectively. As the Flesch-Kincaid Index has a similar interpretation as the Fog Index, this suggests the decrease in complexity of the audit report is larger for audit reports written by non-Big Four auditors, compared to audit reports written by Big Four auditors, ceteris paribus. However, it is important to note that the non-Big Four subsample only includes 18 observations, thus generalizability of these interpretations might be difficult.

4.5 Dictionary-based quantitative text analysis

In the dictionary-based text analysis, the cause for changes in Fog Index, Flesch Index and Flesch-Kincaid Index is investigated. The sentiment of the audit reports before and after the revised ISA 700 and new ISA 701 is compared, using the Loughran and McDonald Master Dictionary (2011), which contains words classified as positive, negative, uncertain, litigious, constraining, and superfluous. The words containing sentiment are emotion words, and according to Hypothesis 4, the proportion of emotion words is expected to rise as a result of the new standards on the audit report.

4.5.1 Descriptive statistics dictionary-based text analysis

In Table 6, the descriptive statistics of the variables used in the dictionary-based quantitative text analysis are presented. Panel A in Table 6 shows the descriptive statistics of the full sample, with 178 observations for each variable, thus no missing values are present. Firstly, the descriptive statistics show that the audit reports contain many stop words. The audit report in the full sample consists, on average, out of 1,742 words including stop words, but when the stop words are removed from the audit reports, the mean of the total words is only 856. In addition, the proportion of unique words increases when the stop words are removed from the audit reports. The lexical diversity for the full sample including stop words has a mean of 26.7%, while the mean of the lexical diversity for the sample excluding stop words equals 39.6%. Following, the descriptive statistics show that all audit reports contain emotion words, with a minimum of 6 emotion words, and an average of 85 emotion words per audit report. The mean of the proportion of emotion words in the full sample, including stop words, is only 4.6%. However, the mean of the proportion of emotion words in the full sample without the stop words is 9.5%. Most emotion words are negative words in the full sample, as the mean of the proportion of negative words equals 39.4%. However, this can also be caused by the composition of the Loughran and McDonald Master Dictionary (2011), where 2,355 of the 4,150 words are classified as negative.

Next, two subsamples are created. The first subsample includes all audit reports adhering to the old standards on the audit report (*EAR* = 0), with its descriptive statistics presented in Panel B of Table

Table 6

Descriptive statistics dictionary-based text analysis

Variables	N	Mean	Median	St. Dev.	Min	Q1	Q3	Max
Panel A: Full sample								
TotalWords	178	1,742.258	1,841.500	1,024.806	355.000	667.200	5,515.800	3,949.000
TotalWordsNSW	178	855.691	880.500	505.548	187.000	326.000	1,259.200	1,914.000
UniqueWords	178	427.101	480.500	207.506	147.000	196.000	610.500	833.000
UniqueWordsNSW	178	321.062	354.000	168.663	100.000	134.000	464.500	651.000
LexDiv	178	0.267	0.259	0.042	0.150	0.236	0.292	0.414
LexDivNSW	178	0.396	0.396	0.045	0.236	0.371	0.413	0.549
EmotionWords	178	84.736	82.500	56.613	6.000	27.000	126.000	249.000
ProportionEmotion	178	0.046	0.046	0.008	0.017	0.040	0.052	0.071
ProportionEmotionNSW	178	0.095	0.094	0.017	0.032	0.083	0.107	0.141
ProportionPos	178	0.043	0.038	0.017	0.000	0.034	0.047	0.107
ProportionNeg	178	0.394	0.395	0.070	0.143	0.333	0.438	0.582
ProportionUnc	178	0.219	0.222	0.063	0.071	0.154	0.268	0.371
ProportionLit	178	0.124	0.121	0.054	0.000	0.093	0.153	0.571
ProportionCon	178	0.194	0.167	0.080	0.067	0.130	0.286	0.333
ProportionSup	178	0.026	0.028	0.014	0.000	0.015	0.037	0.059
Panel B: Sample for EAR=0 (old audit report)								
TotalWords	89	877.629	667.000	488.728	355.000	644.000	741.000	2,215.000
TotalWordsNSW	89	427.360	326.000	232.383	187.000	316.000	357.000	1,071.000
UniqueWords	89	248.303	196.000	108.973	147.000	190.000	228.000	575.000
UniqueWordsNSW	89	176.090	134.000	86.121	100.000	130.000	158.000	449.000
LexDiv	89	0.297	0.292	0.036	0.228	0.284	0.302	0.414
LexDivNSW	89	0.423	0.411	0.041	0.346	0.401	0.425	0.549
EmotionWords	89	38.820	27.000	24.854	6.000	26.000	36.000	112.000
ProportionEmotion	89	0.043	0.042	0.007	0.017	0.039	0.047	0.063
ProportionEmotionNSW	89	0.088	0.085	0.015	0.032	0.081	0.095	0.140
ProportionPos	89	0.038	0.037	0.013	0.000	0.034	0.038	0.107
ProportionNeg	89	0.364	0.333	0.076	0.143	0.308	0.397	0.582
ProportionUnc	89	0.174	0.154	0.044	0.071	0.148	0.194	0.299
ProportionLit	89	0.136	0.148	0.063	0.000	0.111	0.154	0.571
ProportionCon	89	0.258	0.286	0.062	0.101	0.222	0.308	0.333
ProportionSup	89	0.031	0.037	0.013	0.000	0.027	0.038	0.056
Panel C: Sample for EAR=1 (extended audit report)								
TotalWords	89	2,606.888	2,519.000	601.144	1,280.000	2,172.000	3,031.000	3,949.000
TotalWordsNSW	89	1,284.022	1,262.000	298.251	614.000	1,066.000	1,512.000	1,914.000
UniqueWords	89	605.899	612.000	100.339	345.000	542.000	680.000	833.000
UniqueWordsNSW	89	466.034	465.000	85.378	250.000	408.000	527.000	651.000
LexDiv	89	0.237	0.238	0.022	0.150	0.222	0.249	0.293
LexDivNSW	89	0.368	0.373	0.030	0.236	0.349	0.387	0.443
EmotionWords	89	130.652	126.000	39.552	52.000	106.000	154.000	249.000
ProportionEmotion	89	0.050	0.051	0.008	0.030	0.043	0.054	0.071
ProportionEmotionNSW	89	0.101	0.103	0.016	0.061	0.088	0.111	0.141
ProportionPos	89	0.048	0.043	0.020	0.020	0.032	0.056	0.104
ProportionNeg	89	0.423	0.418	0.048	0.319	0.395	0.448	0.563
ProportionUnc	89	0.264	0.261	0.045	0.137	0.231	0.297	0.371
ProportionLit	89	0.112	0.103	0.040	0.042	0.088	0.124	0.285
ProportionCon	89	0.131	0.133	0.032	0.067	0.108	0.148	0.211
ProportionSup	89	0.021	0.019	0.013	0.000	0.014	0.029	0.059

Table 6 shows the descriptive statistics of the variables used in the dictionary-based quantitative text analysis.

Panel A shows the descriptive statistics for the full sample, Panel B shows the descriptive statistics for all observations with audit reports adhering to the old standards (EAR = 0), and Panel C shows the descriptive statistics for all observations with audit reports adhering to the revised ISA 700 and the new ISA 701 (EAR = 1). The variables are defined in Appendix B, Table B1.

6. The second subsample includes all audit reports following the revised ISA 700 and new ISA 701 (*EAR* = 1), with its descriptive statistics shown in Panel C of Table 6. In line of expectation, the new standards on the audit report cause a longer audit report. The old audit report consists, on average, out of 878 words, while the new audit report has a mean of 2,607 words, including stop words. The extension of the audit report has however decreased the lexical diversity of the audit report, as the mean of the lexical diversity including stop words of the old audit report equals 29.7%, but the mean of the lexical diversity of the new audit report is only 23.7%. Looking at the lexical diversity excluding stop words, there is a decrease as well, from 42.3% to 36.8%. This seems to counter the goal of the extension of the audit report, to decrease the level of standardization in the audit report, as a decrease in lexical diversity suggest a text becomes more repetitive. The proportion of emotion words however appears to have increased by the extension of the audit report. For the audit reports including stop words, the proportion of emotion words has increased from 4.3% to 5.0%, and for the audit reports excluding stop words, there is an increase from 8.8% to 10.1%, which supports Hypothesis 4.

4.5.2 Pearson correlation matrix dictionary-based text analysis

In Table 7, the Pearson correlation coefficients and its statistical significance of the variables used in the dictionary-based text analysis are presented. The Fog Index, Flesch Index and Flesch-Kincaid Index are also included in the correlation matrix to investigate whether the dictionary-based text analysis variables are correlated with the readability indexes. Table 7 suggests there is a relation between all variables included in the dictionary-based text analysis and the readability indexes as all correlation coefficients are statistically significant, except the proportion of litigious words and the proportion of superfluous words, and the Flesch Index. However, as the proportion of emotion words is relevant to find support for Hypothesis 4, this variable will be highlighted.

Starting with the audit reports including stop words, the correlation coefficient between the proportion of emotion words and the Fog Index is equal to -0.229, and the correlation coefficient between the Flesch-Kincaid Index and the proportion of emotion words equals -0.254, indicating a weak negative linear relation. A lower Fog Index and Flesch-Kincaid Index implicates a text is easier to understand on the first reading, thus the correlation coefficients suggest that a larger proportion of emotion words makes the audit reports less complex. However, only looking at the correlation matrix and descriptive statistics, no conclusions on causality can be made. The correlation coefficient between the Flesch Index and the proportion of emotion words is only 0.155, so this does not suggest a linear relation is present.

Moving on to the sample with audit reports excluding stop words, similar results on the correlation coefficients can be found. The correlation coefficient of the Fog Index and the proportion of emotion words equals -0.246 and the correlation coefficient of the Flesch-Kincaid Index and the proportion of emotion words is -0.268, suggesting a weak negative linear relation between these readability indexes and the proportion of emotion words excluding stop words. Again, for the Flesch Index, no clear linear relation seems to be present, as the correlation coefficient for the Flesch Index and the proportion of emotion words is only 0.196.

A variable that is expected to be correlated with the readability indexes is the lexical diversity. When looking at the audit reports including stop words, the correlation coefficient of the lexical diversity and the Fog Index, Flesch Index, and Flesch-Kincaid Index are 0.563, -0.493 and 0.577 respectively, suggesting a moderate linear relation. Focusing on the audit reports excluding stop words, the correlation

Table 7

Pearson correlation matrix dictionary-based text analysis

	Fog	*Flesch*	*FleschKincaid*	*TotalWords*	*TotalWordsNSW*
Fog	1.000				
Flesch	-0.921***	1.000			
FleschKincaid	0.993***	-0.915***	1.000		
TotalWords	-0.578***	0.499***	-0.601***	1.000	
TotalWordsNSW	-0.575***	0.439***	-0.599***	0.999***	1.000
UniqueWords	-0.605***	0.471***	-0.629***	0.986***	0.986***
UniqueWordsNSW	-0.599***	0.462***	-0.622***	0.985***	0.986***
LexDiv	0.563***	-0.493***	0.577***	-0.879***	-0.873***
LexDivNSW	0.452***	-0.386***	0.462***	-0.771***	-0.770***
EmotionWords	-0.518***	0.384***	-0.547***	0.962***	0.962***
ProportionEmotion	-0.229***	0.155*	-0.254***	0.463***	0.462***
ProportionEmotionNSW	-0.246***	0.196***	-0.268***	0.443***	0.436***
ProportionPos	-0.214***	0.202***	-0.202***	0.287***	0.284***
ProportionNeg	-0.361***	0.250***	-0.375***	0.468***	0.458***
ProportionUnc	-0.435***	0.281***	-0.449***	0.747***	0.754***
ProportionLit	0.144*	-0.067	0.144*	-0.264***	-0.261***
ProportionCon	0.575***	-0.425***	0.593***	-0.793***	-0.790***
ProportionSup	0.181**	-0.077	0.195***	-0.515***	-0.513***

	UniqueWords	*UniqueWordsNSW*	*LexDiv*	*LexDivNSW*	*EmotionWords*
UniqueWords	1.000				
UniqueWordsNSW	0.999***	1.000			
LexDiv	-0.839***	-0.833***	1.000		
LexDivNSW	-0.702***	-0.697***	0.956***	1.000	
EmotionWords	0.944***	0.945***	-0.838***	-0.734***	1.000
ProportionEmotion	0.466***	0.467***	-0.379***	-0.284***	0.650***
ProportionEmotionNSW	0.444***	0.442***	-0.378***	-0.268***	0.625***
ProportionPos	0.280***	0.280***	-0.230***	-0.185**	0.231***
ProportionNeg	0.497***	0.482***	-0.431***	-0.324***	0.456***
ProportionUnc	0.767***	0.765***	-0.674***	-0.592***	0.693***
ProportionLit	-0.276***	-0.265***	0.312***	0.302***	-0.221***
ProportionCon	-0.823***	-0.816***	0.679***	0.536***	-0.758***
ProportionSup	-0.517***	-0.519***	0.393***	0.286***	-0.506***

	ProportionEmotion	*ProportionEmotionNSW*	*ProportionPos*	*ProportionNeg*	*ProportionUnc*
ProportionEmotion	1.000				
ProportionEmotionNSW	0.989***	1.000			
ProportionPos	0.095	0.103	1.000		
ProportionNeg	0.335***	0.354***	0.068	1.000	
ProportionUnc	0.329***	0.290***	0.167**	0.376***	1.000
ProportionLit	-0.163**	-0.164**	-0.185**	-0.641***	-0.458***
ProportionCon	-0.413***	-0.400***	-0.263***	-0.680***	-0.778***
ProportionSup	-0.282***	-0.285***	-0.178	-0.397***	-0.387***

	ProportionLit	*ProportionCon*	*ProportionSup*
ProportionLit	1.000		
ProportionCon	0.272***	1.000	
ProportionSup	0.071	0.457***	1.000

Table 7 shows the Pearson correlation coefficients of the reading indexes and the variables used in the dictionary-

*based quantitative text analysis.. Statistical significance is presented as: * p<0.1; ** p<0.05; *** p<0.01.*

coefficient of the Fog Index and the lexical diversity equals 0.452, and the correlation coefficient of the Flesch-Kincaid Index and the lexical diversity is 0.462, suggesting a moderate positive linear relation. The correlation coefficient of the Flesch Index and the lexical diversity equals -0.386. Thus, the audit reports seem to become more complex as the lexical diversity increases. However, just as for the proportion of emotion words, no conclusions can be drawn on causality when only the descriptive statistics and correlation matrix are analyzed.

4.5.3 Paired sample t-tests dictionary-based text analysis

Hypothesis 4 proposes the proportion of emotion words in the audit reports before the extension of the audit report is lower than the proportion of emotion words in the audit reports following the new standards. Support for this hypothesis is gained through paired sample t-tests on all variables introduced in the dictionary-based quantitative text analysis. As the central limit theorem assumes the sampling distribution tends to be normal if the sample size is large enough ($N > 30$), no additional density plots or quantile-quantile plots are created for the additional dictionary-based text analysis.

The results of the paired sample t-tests are presented in Table 8. Remarkably, all fifteen tests are statistically significant. The results and their interpretation are discussed one by one in this section. Firstly, the total number of words in the audit report, including stop words, statistically increased (M = 1,729.258, SD = 763.495) as a result of new standards on the audit report, $t(88) = 21.367$, $p < 0.01$. When looking at the audit reports excluding stop words, the total number of words in the audit report statistically increased as well (M = 856.663, SD = 376.890), as a result of the extension of the audit report, $t(88) = 21.443$, $p < 0.01$. This is in line of expectation, as the new standards on the audit report require the auditor to discuss the Key Audit Matters, materiality, basis for their opinion, scope of the group audit and their appointment, which were no required parts of the audit report previously (IAASB, 2015a; IAASB, 2015b), as shown in the example of the old and new audit report in Appendix A. In addition, the revised ISA 700 requires the auditor to describe the responsibilities of the Board of Directors and the auditor's responsibilities more extensively in the audit report (IAASB, 2015a).

Secondly, the paired sample t-tests in Table 8 show the number of unique words, including stop words, significantly increased (M = 357.596, SD = 144.674) as a result of the new standards, $t(88)$ = 23.318, $p < 0.01$. Also, for the audit reports excluding stop words, the absolute number of unique words statistically increased (M = 289.944, SD = 118.331), $t(88) = 23.120$, $p < 0.01$. Because the total number of words statistically increased, it is also expected that the absolute number of unique words increases. Therefore, the lexical diversity is more informative than the absolute number of unique words, as the lexical diversity is equal to the number of unique words divided by the total number of words.

Starting again with the audit reports including stop words, the lexical diversity statistically decreased (M = -0.060, SD = 0.042), when the new standards on the audit report are followed, $t(88)$ = -13.588, $p < 0.01$. Also, for the audit reports excluding stop words, the lexical diversity statistically decreased (M = -0.055, SD = 0.051), when following the revised ISA 700 and new ISA 701, $t(88)$ = -10.055. These results suggest the text in the new audit reports is more repetitive compared to the old audit reports.

Moving on to the emotion words, the total number of emotion words in the old audit reports is statistically lower (M = 91.831, SD = 47.994) compared to the total number of emotion words in the new audit report, $t(88) = 18.051$, $p < 0.01$. However, just as for the absolute number of unique words, this result is not very informative as the total number of words increased as a result of the extension of the audit report. Therefore, the proportion of emotion words can provide more insight in the level of sentiment in the audit reports.

Table 8

Paired sample t-test results dictionary-based text analysis

Text analysis variables differences	N	Mean	St. Dev.	[95% Conf. Interval] Lower	Upper	t	df	Sig. (2-tailed)
Δ*TotalWords*	89	1,729.258	763.495	1,568.426	1,890.090	21.367	88	0.000***
Δ*TotalWordsNSW*	89	856.663	376.890	777.270	936.056	21.443	88	0.000***
Δ*UniqueWords*	89	357.596	144.674	327.120	388.071	23.318	88	0.000***
Δ*UniqueWordsNSW*	89	289.944	118.311	265.021	314.866	23.120	88	0.000***
Δ*LexDiv*	89	-0.060	0.042	-0.069	-0.052	-13.588	88	0.000***
Δ*LexDivNSW*	89	-0.055	0.051	-0.065	-0.044	-10.05	88	0.000***
Δ*EmotionWords*	89	91.831	47.994	81.721	101.941	18.051	88	0.000***
Δ*ProportionEmotion*	89	0.007	0.010	0.005	0.009	6.485	88	0.000***
Δ*ProportionEmotionNSW*	89	0.013	0.020	0.009	0.017	6.015	88	0.000***
Δ*ProportionPos*	89	0.010	0.023	0.005	0.015	4.034	88	0.000***
Δ*ProportionNeg*	89	0.059	0.082	0.042	0.076	6.809	88	0.000***
Δ*ProportionUnc*	89	0.091	0.065	0.077	0.104	13.227	88	0.000***
Δ*ProportionLit*	89	-0.024	0.073	-0.040	-0.009	-3.165	88	0.002***
Δ*ProportionCon*	89	-0.126	0.067	-0.140	-0.112	-17.777	88	0.000***
Δ*ProportionSup*	89	-0.009	0.017	-0.013	-0.006	-5.214	88	0.000***

*Table 8 shows the results of the paired sample t-tests on the variables used in the dictionary-based quantitative text analysis. The tests show whether the variables used on the old audit reports are statistically different from the variables used on the extended audit report. This is the case when Δ[Variable] is statistically different from zero. Statistical significance is presented as: * p<0.1; ** p<0.05; *** p<0.01.*

For the audit reports including stop words, the proportion of emotion words is statistically higher when adhering to the new standards (M = 0.007, SD = 0.010) compared to the old standards, t(88) = 6.485, p < 0.01. For the audit reports excluding stop words, the proportion of emotion words is statistically higher as well when adhering to the new standards (M = 0.013, SD = 0.020) compared to the old standards, t(88) = 6.015, p < 0.01. These results suggest that the level of sentiment is higher in the new audit reports, as the proportion of emotion words has increased, implicating no evidence is found to reject Hypothesis 4. Next, it is explored which type of emotion words is more common in the extended audit report, or if there is a decrease in a certain category of emotion words.

The first category includes all words classified as positive words. The proportion of positive words statistically increased (M = 0.010, SD = 0.023) when the new standards on the audit report are followed t(88) = 4.034, p < 0.01. Secondly, the proportion of negative words statistically increased as well for the new audit report (M = 0.059, SD = 0.082), t(88) = 6.809. Thirdly, a statistically significant increase (M = 0.091, SD = 0.065) can be found in the proportion of words with a tone of uncertainty when the new standards are applied. For the other three categories, the proportion decreased. Firstly, the proportion of litigious words significantly decreased (M = -0.024, SD = 0.073) for the new audit report t(88) = -3.165, p < 0.01. Secondly, there is a statistically significant decrease (M = -0.126, SD = 0.067) in the proportion of constraining words, t(88) = -17.777, p < 0.01. Finally, the proportion of superfluous words significantly decreased (M = -0.009, SD = 0.017) as a result of the new standards on the audit report, t(88) = -5.214, p < 0.01. Thus, the decrease in emotion words in the audit reports adhering to the revised ISA 700 and new 701 seems mainly caused by a proportionally increase in words with a positive, negative, and uncertain tone.

4.6 Document similarity analysis

The final textual analysis method measures the similarity between the 89 old audit reports and compares this to the similarity between the 89 extended audit reports, using the Jaccard Index as a similarity coefficient. The average Jaccard Index for all old audit reports is 0.143, which implies that there is an average similarity between the audit reports following the old standards of 14.3%. For the extended audit reports, the average Jaccard Index is equal to 0.079, which suggests there is an average similarity of 7.9% between the audit reports following the revised ISA 700 and new ISA 701. This means there is a decrease of approximately 44.8% in the average similarity between the audit reports when the new standards are followed, compared to the old standards on the audit report.

Thus, the objective of the revision of ISA 700 and introducing ISA 701, which is to decrease the level of boilerplate text in the independent auditor's report, seems to be reached, as the similarity between the audit reports decreased by 44.8%. Therefore, Hypothesis 5 cannot be rejected, as the results of the document similarity analysis suggest that the audit report contains less template-text and more entity-specific information when the new standards are followed.

5. Conclusion

As a result of the decline in trust in the financial reporting system, and the increasing demand for more transparent and extended audit reports, the European Commission released their Green Paper on how financial reporting could be improved, in particular the independent auditor's report (EC, 2010). As a reaction, the IAASB revised ISA 700 *Forming an Opinion and Reporting on Financial Statements* and they released ISA 701 *Communicating Key Audit Matters in the Independent Auditor's Report* (IAASB 2015a; IAASB, 2015b). The new and revised standards should improve the auditor's transparency, reduce the level of boilerplate text in the audit report by increasing the level of entity-specific information, and improve the readability of the audit report (PwC, 2014). This led to the research question: *Does the extension of the audit report through the revision of ISA 700 and implementation of ISA 701 affect the readability and the level of standardization of the audit report for Dutch firms?*

The readability measures used to assess the complexity of audit reports are the commonly used Fog Index, Flesch Index and Flesch-Kincaid Index. The paired sample t-tests performed on the readability indexes show that the Fog Index and the Flesch-Kincaid Index on average is lower for the audit reports adhering to the revised ISA 700 and new ISA 701, compared to the old audit report, while the Flesch Index is higher for audit reports following the new standards. This suggests the readability is higher for audit reports following the revised ISA 700 and new ISA 701. The regression analyses performed with the readability indexes as dependent variables show that the application of the revised ISA 700 and new ISA 701 lead to a statistically significant improvement of the readability of the audit report. In addition, the regression analyses show statistically significant coefficients for the binary variable representing whether the auditor works at a Big Four firm. For the Fog Index and the Flesch-Kincaid Index, the coefficient is significantly positive and for the Flesch Index the coefficient is significantly negative, suggesting Big Four auditors use more complex English language in their audit reports, ceteris paribus. Additional regression analysis on non-Big Four and Big Four subsamples suggests the improvement in readability of the audit report as a result of the extension of the audit report is stronger for audit reports written by non-Big Four auditors than Big Four auditors. Finally, the R^2 and adjusted R^2 for all three regression models on the full sample are considerably high, suggesting the data seems to properly fit the regression models. Thus, the regression results support Hypothesis 1, 2 and 3: audit reports adhering to the revised ISA 700 and new ISA 701 have a lower Fog Index, higher Flesch Index, and lower Flesch-Kincaid Index respectively, compared to audit reports following the old standards, suggesting the revision of ISA 700 and implementation of ISA 701 improved the readability of the audit report for Dutch firms.

After assessing the readability indexes, dictionary-based quantitative text analysis is performed, using the Loughran and McDonald Master Dictionary (2011), with sentiment categories including positive, negative, uncertainty, litigious, constraining, and superfluous words. As one of the goals of the revision of ISA 700 and implementation of ISA 701 is to decrease the level of boilerplate text and show more the opinion of the auditor (PwC, 2014), the expectation rises that the new audit report contains more sentiment compared to the old audit report. The level of sentiment is measured by counting emotion words, which are words in the audit report that reoccur in the Loughran and McDonald Master Dictionary (2011). However, the absolute number of emotion words is not informative if the total number of words in the audit report changes as a result of the extension of the audit report. Therefore, Hypothesis 4 focusses on the proportion of emotion words in the audit report before and after the revised ISA 700 and new ISA 701 are applied. The paired sample t-tests in the additional dictionary-based quantitative text analysis show that the proportion of emotion words in the audit reports, including and excluding stop words, is higher for audit reports adhering to the new standards, supporting Hypothesis 4, and suggesting the level of sentiment in the audit report has increased through the revision of ISA 700 and the introduction of ISA 701. This increase in the proportion of emotion words in the audit report seems mainly caused by an increase in the proportion of words with a positive, negative, and uncertain tone. The proportion of words with a litigious, constraining, and superfluous tone significantly decreased for audit reports adhering to the new standards. In addition, the lexical diversity for audit reports, including and excluding stop words, significantly decreased for audit reports following the revised ISA 700 and new ISA 701, suggesting the extended audit report is more repetitive than the old audit report. This seems to counter the objectives of the revision of ISA 700 and introduction of ISA 701, which could be explored further in the future.

Finally, document similarity analysis is performed to explore whether the level of boilerplate text decreases as a result of the extension of the audit report, using the Jaccard Index as similarity measure. The average Jaccard Index of the old audit reports is higher compared to the average Jaccard Index of the extended audit reports, suggesting the document similarity between the audit reports decreases when the revised ISA 700 and new ISA 701 are applied, supporting Hypothesis 5. This implies the level of boilerplate text decreases, and the level of entity-specific information increases.

The application of the Fog Index, Flesch Index and Flesch-Kincaid Index, but also dictionary-based text analysis and document similarity analysis, on the independent auditor's report solely, are important additions to existing accounting literature that mainly focuses on the annual report as a whole, as hardly any research explicitly examines the readability of the audit report, let alone the sentiment analysis or document similarity analysis. However, as the audit report is written by the auditor, while the annual report is composed by the company, the readability of the audit report and the annual report can be completely different and could also be affected by different factors. The main reason for this suspected difference is that the goal of the auditor and the company differ, as the auditor's objective is to express an opinion on the financial statements, while the company wants to present themselves in the best possible way in the annual report.

The results of the regression analyses, dictionary-based text analysis and document similarity analysis are an appropriate evaluation of the decisions made by the IAASB when revising ISA 700 and ISA 701. The findings suggest one of the goals of the revision of ISA 700 and introduction of ISA 701, namely, to improve the readability of the audit report, is achieved. The goal to decrease the level of boilerplate text in the audit report also seems to be reached, as the proportion of emotion words significantly increased for audit reports adhering to the new standards. However, another reason for an increase in the proportion of emotion words in the audit report could be that auditors still use standard templates for the audit report, however these templates include more words that are in the Loughran and McDonald Master Dictionary (2011), percentual, than the old template. But to rule this out, document similarity analysis shows the level of boilerplate text decreased as a result of the extension of the audit report, making it more likely the level of entity-specific information increased.

In further research, more detailed analysis on the content of the audit report could be performed, such as further analyzing the level of similarity of the audit reports of different companies within the same year, by the same audit firm. In this paper, no statistical significance is tested when examining the similarity between the old audit reports, and the similarity between the extended audit reports. In addition, there might be more appropriate similarity coefficients than the Jaccard Index, but this is uncertain as not much textual analysis on document similarity is performed, focusing on annual reports or audit reports. Secondly, the sample could be expanded to make generalization of the results more appropriate. This sample only includes the hand-collected audit reports of Dutch firms in the last year their auditor applies the old standards on the audit report, and the audit report in the first year the revised ISA 700 and new ISA 701 are applied. For robustness, but also to analyze similarity between audit reports for the same company over multiple years, a larger time span could be included in the sample. Naturally, the sample could be expanded through firms in other countries, with the condition that the International Standards on Auditing are followed. Finally, qualitative research could focus on the perception of the extended audit report, compared to the old audit report, by various parties such as investors, lenders, analysts, or standard setters.

References

Ahold. (2015). *Annual Report 2014*. https://www.aholddelhaize.com/media/1388/ahold_ar14_fullreport_interactive.pdf

Amsterdam Commodities N.V.. (2014). *Annual Report 2013*. https://www.acomo.nl/wp-content/uploads/2015/09/Acomo-Annual-Report-2013-final.pdf

Amsterdam Commodities N.V.. (2015) *Annual Report 2014*. https://www.acomo.nl/wp-content/uploads/2015/09/Acomo-Annual-Report-2014.pdf

Clatworthy, M., & Jones, M. J. (2001). The effect of thematic structure on the variability of annual report readability. *Accounting, Auditing & Accountability Journal*, *14*(3), 311-326.

De Franco, G., Hope, O. K., Vyas, D., & Zhou, Y. (2015). Analyst report readability. *Contemporary Accounting Research*, *32*(1), 76-104.

EC. (2010). *Green Paper. Audit Policy: Lessons from the Crisis*. https://www.europarl.europa.eu/meetdocs/2009_2014/documents/com/com_com(2010)0561_/com_com(2010)0561_en.pdf

Ertugrul, M., Lei, J., Qiu, J., & Wan, C. (2017). Annual report readability, tone ambiguity, and the cost of borrowing. *Journal of Financial and Quantitative Analysis*, *52*(2), 811-836.

Fama, E. F., & French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, *43*(2), 153-193.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221.

IAASB. (2015a). International Standard on Auditing (ISA) 700 (Revised), Forming an opinion and reporting on financial statements. IAASB. https://www.iaasb.org/publications/international-standard-auditing-isa-700-revised-forming-opinion-and-reporting-financial-statements-5.

IAASB. (2015b). International Standard on Auditing (ISA) 701 (New), Communicating Key Audit Matters in the independent auditor’s report. IAASB. https://www.iaasb.org/publications/international-standard-auditing-isa-701-new-communicating-key-audit-matters-independent-auditors-4.

IAASB. (2017). Auditor Reporting Standards Implementation – Key Audit Matters. IFAC. https://www.ifac.org/knowledge-gateway/supporting-international-standards/discussion/auditor-reporting-standards-implementation-key-audit-matters.

IFAC. (2019). *Membership: The Netherlands*. https://www.ifac.org/about-ifac/membership/country/netherlands

Jones, M. J. (1988). A longitudinal study of the readability of the chairman's narratives in the corporate reports of a UK company. *Accounting and Business Research*, *18*(72), 297-305.

Jones, M. J., & Shoemaker, P. A. (1994). Accounting narratives: A review of empirical studies of content and readability. *Journal of Accounting Literature*, *13*, 142.

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel*. Naval Technical Training Command Millington TN Research Branch.

Landerman, L. R., Land, K. C., & Pieper, C. F. (1997). An empirical evaluation of the predictive mean matching method for imputing missing values. *Sociological Methods & Research*, *26*(1), 3-33.

Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of Accounting and Economics*, *56*(1), 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*, *86*(3), 1087-1115.

Lewis, N. R., Parker, L. D., Pound, G. D., & Sutcliffe, P. (1986). Accounting report readability: The use of readability techniques. *Accounting and Business Research*, *16*(63), 199-213.

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

Lo, K., Ramos, F., & Rogo, R. (2017). Earnings management and annual report readability. *Journal of Accounting and Economics*, *63*(1), 1-25.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35-65.

Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *The Journal of Finance*, *69*(4), 1643-1671.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(4), 1187-1230.

Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. *The Accounting Review*, *85*(6), 2107-2143.

Morais, A. I., & Fialho, A. (2008). Do harmonised accounting standards lead to harmonised accounting practices? An empirical study of IAS 39 measurement requirements in some European Union countries. *Australian Accounting Review*, *18*(3), 224-236.

Mullen, L. (2016). *Textreuse: Detect Text Reuse and Document Similarity. R package version 0.1.4.900*. https://github.com/ropensci/textreuse

NBA. (2014). *New auditor's report for public interest entities. Featuring additional information about the auditor's activities.* https://www.nba.nl/globalassets/tools-en-voorbeelden/voorbeeldteksten-en-passages/nieuwe-structuur-controleverklaringen-voor-alle-controles/nba_voorlichtingsbrochure_new_auditors_report.pdf

Niwattanakul, S., Singthongchai, J., Naenudorn, E., & Wanapu, S. (2013). Using of Jaccard coefficient for keywords similarity. *Proceedings of the International Multiconference of Engineers and Computer Scientists*, *1*(6), 380-384.

PwC. (2014). *Plain speaking! Benchmarking the new auditor's report among Dutch listed entities.* https://www.pwc.nl/nl/assets/documents/pwc-benchmark-auditors-reporting.pdf

Quick, R. (2012). EC green paper proposals and audit quality. *Accounting in Europe*, *9*(1), 17-38.

Stone, G., & Parker, L. D. (2013). Developing the Flesch reading ease formula for the contemporary accounting communications landscape. *Qualitative Research in Accounting & Management*, *10*(1).

Velte, P. (2018). Does gender diversity in the audit committee influence key audit matters' readability in the audit report? UK evidence. *Corporate Social Responsibility and Environmental Management*, *25*(5), 748-755.

Appendix A

Example of the new independent auditor's report (NBA, 2014), with changes to the old independent auditor's report in italics

INDEPENDENT AUDITOR'S REPORT ON FINANCIAL STATEMENTS
To: The Shareholders and Supervisory Board of ABC N.V.

Report on the Audit of the Financial Statements 201X

Opinion
We have audited the accompanying financial statements 201X of ABC N.V. (the Company), based in (town/city). The financial statements include the consolidated financial statements and the company financial statements.

The consolidated financial statements comprise:
1) the consolidated statement of financial position as at December 31, 201X;
2) the following overviews for 201X: consolidated statements of profit or loss and other comprehensive income, changes in equity and cash flows for the year then ended; and
3) notes to the financial statements, including a summary of the significant accounting policies and other explanatory information.

The company financial statements comprise:
1) the company balance sheet on December 31, 201X;
2) the company profit and loss account for the year 201X; and
3) notes comprising a summary of the accounting policies and other explanatory information.

In our opinion: *(the position of this paragraph changed)*
• The consolidated financial statements give a true and fair view of the financial position of ABC N.V. (name of the Company) on December 31, 201X its financial and its cash flows in the year 201X in accordance with International Financial Reporting Standards as adopted by the European Union (EU-IFRS) and with Part 9 of Book 2 of the Dutch Civil Code.
• The company financial statements give a true and fair view of the financial position of ABC N.V. (name of the Company) as at December 31, 201X and of its financial performance for the year 201X in accordance with Part 9 of Book 2 of the Dutch Civil Code.

Basis for Opinion
We conducted our audit in accordance with Dutch law, which also covers Dutch Standards on Auditing. Our responsibilities under these standards have been further specified in the "Auditor's Responsibilities for the Audit of the Financial Statements" section of our report.

We are independent of ABC N.V. in accordance with the "Verordening inzake de onafhankelijkheid van accountants bij assurance-opdrachten" (ViO) and other relevant independence regulations in the Netherlands. Furthermore we have complied with the "Verordening gedrags- en beroepsregels accountants" (VGBA) and other relevant regulations.

We believe the audit evidence we have obtained is sufficient and appropriate to provide a basis for our opinion.

Materiality
Misstatements may arise due to fraud or error and will be considered material if, individually or in the aggregate, one can reasonably expect them to influence the economic decisions made by users based on the financial statements. The materiality affects the nature, timing and extent of our audit procedures and the evaluation of the effect of identified misstatements on our opinion.

We have used our professional judgement to determine that materiality for the financial statements as a whole was set at EUR X. The materiality is based on ... (% to be filled in for the relevant benchmark e.g. profit, turnover or other criteria). We have also misstatements and/or possible misstatements taken into account that are in our opinion material for qualitative reasons.

We agreed with the Supervisory Board that misstatements in excess of EUR Y, which are identified during the audit, would be reported to them, as would smaller misstatements we believe must be reported on qualitative grounds.

Scope of our group audit
ABC N.V. is at the head of a group of entities. The financial information of this group is included in the consolidated financial statements of ABC N.V.

Because we bear ultimate responsibility for the opinion, we are also responsible for directing, supervising and performing the group audit. And we have considered this responsibility when determining the nature and extent of the audit procedures carried out for group entities. When doing so, the significance and/or risk profile of entities or activities played a key role. On this basis, we selected group entities for which an audit or review had to be carried out on the complete set of financial statements or specific items.

Our group audit mainly concentrated on significant group entities ... [explain what this entails e.g. group entities, countries, activities]. We have performed audit procedures ourselves at group entities aaa and bbb. And used the work of other auditors when auditing entity ccc. We performed review procedures or specific audit procedures at the other group entities.

By implementing the above-mentioned procedures within group entities, together with additional procedures at group level, we have been able to obtain sufficient and appropriate audit evidence about the group's financial information to provide an opinion about the consolidated financial statements.

Key Audit Matters
These key audit matters include matters which, in our professional judgment, were of most significance in our audit of the financial statements. We have communicated the key audit matters to the Supervisory Board. The key audit matters are not a comprehensive reflection of all matters discussed.

Our auditing procedures for these matters were determined as part of our audit on the financial statements as a whole. Our description of individual key matters must thus be seen in this context and not as individual opinions about these matters.

Descriptions for key audit matters contain the following elements:
- *A description of the key audit matter;*
- *A summary of performed audit procedures;*
- *If relevant, key observations relating to key audit matters;*

• If relevant, references to information or notes in the annual report.

Responsibilities of Management and the Supervisory Board for Financial Statements
Management is responsible for:
• the preparation and fair presentation in accordance with EU-IFRS and Part 9 of Book 2 of the Dutch Civil Code, and for the preparation of the management board report in accordance with Part 9 of Book 2 of the Dutch Civil Code, and for
• such internal control as management determines is necessary to enable the preparation of financial statements that are free from material misstatement, whether due to fraud or error.

In preparing the financial statements, management is responsible for assessing the Company's ability to continue as a going concern. Based on the financial reporting frameworks mentioned, management should prepare the financial statements using the going concern basis of accounting unless management either intends to liquidate the Company or to cease operations, or has no realistic alternative but to do so. Management should disclose events and circumstances that may cast significant doubt on the Company's ability to continue as a going concern.

The Supervisory Board is responsible for overseeing the Company's financial reporting process.

Our Responsibilities for the Audit of the Financial Statements
Our objective is to plan and perform the audit assignment in a manner that allows sufficient and appropriate audit evidence to be obtained for our final opinion.

Our audit has been performed with a high, but not absolute, level of assurance, which means we may have not uncovered all errors and fraud. Therefore, the opinion offered in the auditor's report can be seen to provide a reasonable level of assurance.

The text below can be included in the Appendix of the auditor's report. *(the auditor's responsibility's were already present in the old audit report, but not in the appendix and not this extensive)*
We have exercised professional judgment and have maintained professional skepticism throughout the audit, in accordance with Dutch Standards on Auditing, ethical requirements and independence requirements.

Our audit included e.g.:
• Identifying and assessing the risks of material misstatement of the financial statements, whether due to fraud or error, design and perform audit procedures responsive to those risks, and obtain audit evidence that is sufficient and appropriate to provide a basis for our opinion. The risk of not detecting a material misstatement resulting from fraud is higher than for one resulting from error, as fraud may involve collusion, forgery, intentional omissions, misrepresentations, or the override of internal control.

• Obtaining an understanding of internal control relevant to the audit in order to design audit procedures that are appropriate in the circumstances, but not for the purpose of expressing an opinion on the effectiveness of the Company's internal control.

• Evaluating the appropriateness of accounting policies used and the reasonableness of accounting estimates and related disclosures made by management.

• Concluding on the appropriateness of management's use of the going concern basis of accounting, and based on the audit evidence obtained, whether a material uncertainty exists related to events and or conditions that may cast significant doubt on the Company's ability to continue as a going concern.

If we conclude that a material uncertainty exists, we are required to draw attention in our auditor's report to the related disclosures in the financial statements or, if such disclosures are inadequate, to modify our opinion. Our conclusions are based on the audit evidence obtained up to the date of our auditor's report. However, future events or conditions may cause an the Company to cease to continue as a going concern.

• Evaluating the overall presentation, structure and content of the financial statements, including the disclosures, and whether the financial statements represent the underlying transactions and events in a manner that achieves fair presentation.

We communicate with the Supervisory Board regarding, among other matters, the planned scope and timing of the audit and significant audit findings, including any significant findings in internal control that we identify during our audit.

We provide the Supervisory Board with a statement that we have complied with relevant ethical requirements regarding independence, and to communicate with them all relationships and other matters that may reasonably be thought to bear on our independence, and where applicable, related safeguards.

From the matters communicated with the Supervisory Board, we determine those matters that were of most significance in the audit of the financial statements of the current period and are therefore the key audit matters. We describe these matters in our auditor's report unless law or regulation precludes public disclosure about the matter or when, in extremely rare circumstances, we determine that a matter should not be communicated in our report because the adverse consequences of such communication would reasonably be expected to outweigh the public interest benefits of doing so. A precondition is that the Company has not publicly disclosed information about the matter.

Report on other legal and regulatory requirements
Report on the management board report and the other information
Pursuant to legal requirements under Section 2:393 sub 5 at e and f of the Dutch Civil Code (conceming our obligation to report about the management board report and other data), we declare that:
• We have no deficiencies to report as a result of our examination whether the management board report, to the extent we can assess, has been prepared in accordance with Part 9 of Book 2 of this Code, and whether the information as required under Section 2:392 sub 1 at b-h has been annexed.
• Further we report that the management board report, to the extent we can assess, is consistent with the financial statements as required by Section 2:391 sub 4 of the Dutch Civil Code.

Appointment
We were appointed by the Supervisory Board as auditor of ABC N.V. on [dated-mm-yy], as of the audit for year X and have operated as statutory auditor ever since that date.

Place, __________ date __________
__________________ (Name of Audit firm)
__________________ (Name statutory auditor and signature)

Appendix B

Table B1

Variable description overview

Variable	**Variable Description**
Dependent variables	
Fog	The Fog Index is calculated as 0.4 * (average sentence length + percent of complex words). The Fog Index can be interpreted as the years of formal education a person needs to understand a text on the first reading. Thus, a high Fog Index implicates the text is less readable.
Flesch	The Flesch Index is calculated as 206.835 – (1.015 * average sentence length) – (84.6 * (number of syllables/number of words)). A higher Flesch Index implicates the text is easier to read. The Flesch Index can take on values between 0 and 100, with different ranges representing different levels of difficulty. For example, a Flesch Index between 10.0 and 30.0 implicates a text is very difficult to read, best understood by university graduates. On the other hand, a Flesch Index between 60.0 and 70.0 represents a text written in plain English, which can be easily understood by 13- to 15-year-old students.
FleschKincaid	The Flesch-Kincaid Index is calculated as (0.39 * average sentence length) – (11.8 * (number of syllables/number of words)) – 15.59. The Flesch-Kincaid Index presents a U.S. grade level; thus a lower Flesch-Kincaid Index implicates the text is easier to read. When the Flesch-Kincaid Index is larger than 10.0, the score can be interpreted as the years of formal education required to understand a text on the first reading.
Variable of interest	
EAR	This is a binary variable, equal to 0 for the audit reports in the old format, and equal to 1 when the revised ISA 700 and new ISA 701 are applied on the audit report.
Control variables	
Earnings	The net income from Compustat Global for firm i in year t, scaled by the book value of total assets following Li (2008). All values smaller than -1 or larger than +1 should be removed from the sample following Li (2008), however no firms in this sample have earnings scaled by the book value of total assets below -1 or above +1. High Earnings are expected to have a positive effect on the readability of the audit report (thus a lower Fog Index and Flesch-Kincaid Index, and a higher Flesch Index).
log(Size)	The natural logarithm of the size of each firm i in year t is calculated as the natural logarithm of the stock price times shares outstanding (in millions), at the end of the fiscal year, obtained from Compustat Security Daily.
log(MTB)	The natural logarithm of the market-to-book ratio is calculated as the natural logarithm of the market value divided by the book value of each firm i in year t. The market value is calculated as the number of shares outstanding times the price per share, at the end of the fiscal year, retrieved from Compustat Security Daily. The book value is equal to the total equity of firm i in year t, retrieved from Compustat Global. Firms with negative book value are removed, the variable is winsorized at 1% following Loughran and McDonald (2014).
abs(SUE)	This variable shows the absolute value of the standardized unexpected earnings. The standardized unexpected earnings, also known as the SUE Score, is calculated as: (actual earnings per share-average expected earnings per share)/standard deviation. Following Loughran and McDonald (2014), the variable is winsorized at 1%. The SUE score can be retrieved directly from I/B/E/S unadjusted data files.
AuditChange	This is a binary variable, equal to 0 for companies with the same audit firm before and after the revised ISA 700 and new ISA 701 are followed, and 1 otherwise.
Big4	This is a binary variable, equal to 0 for companies who have an audit report written by a non-Big Four auditor, and 1 otherwise. A Big Four auditor is an auditor who works for Deloitte, KPMG, PwC or EY.

FamaFrench	The Fama and French (1997) 48-industry dummies are included to control for industries requiring more complex words, that decrease the level of readability. The SIC code for each firm i is retrieved from COMPUSTAT and converted to one of the 48 Fama and French (1997) industries.
Dictionary-based text analysis variables	
TotalWords	This variable shows the total number of words per audit report, after all interpunction and numbers are removed from the text, but stop words are still included.
TotalWordsNSW	This variable shows the total number of words per audit report, after all interpunction and numbers are removed from the text, and all stop words such as “and” or “the” are removed.
UniqueWords	The total number of unique words per audit report after all interpunction and numbers are removed from the text, but stop words are still included.
UniqueWordsNSW	The total number of unique words per audit report after all interpunction and numbers are removed from the text, and all stop words are removed.
LexDiv	The total number of unique words (*UniqueWords*) divided by the total number of words (*TotalWords*) per audit report.
LexDivNSW	The total number of unique words (*UniqueWordsNSW*) divided by the total number of words (*TotalWordsNSW*) per audit report excluding stop words.
EmotionWords	The total number of emotion words per audit report is counted, using the Loughran and McDonald Master Dictionary (2011), including words classified as positive, negative, uncertain, litigious, constraining, and superfluous. As the dictionary does not contain stop words, it does not matter whether the sample includes audit reports with stop words.
ProportionEmotion	The total number of emotion words (*EmotionWords*) divided by the total number of words (*TotalWords*) per audit report, including stop words.
ProportionEmotionNSW	The total number of emotion words (*EmotionWords*) divided by the total number of words per audit report, excluding stop words (*TotalWordsNSW*).
ProportionPos	The total number of words in an audit report classified as positive, divided by the total number of emotion words (*EmotionWords*).
ProportionNeg	The total number of words in an audit report classified as negative, divided by the total number of emotion words (*EmotionWords*).
ProportionUnc	The total number of words in an audit report classified as uncertainty, divided by the total number of emotion words (*EmotionWords*).
ProportionLit	The total number of words in an audit report classified as litigious, divided by the total number of emotion words (*EmotionWords*).
ProportionCon	The total number of words in an audit report classified as constraining, divided by the total number of emotion words (*EmotionWords*).
ProportionSup	The total number of words in an audit report classified as superfluous, divided by the total number of emotion words (*EmotionWords*).

Table B1 shows the description of all variables used in the regression analyses and the dictionary-based quantitative text analysis. It also presents how some variables are calculated, or from which databases data is retrieved.

Table B2

Fama and French 48-industry averages of the readability indexes

Industry name	Industry number	Number of firms	Old audit report (*EAR* = 0)			Extended audit report (*EAR* = 1)		
			Fog	*Flesch*	*FleschKincaid*	*Fog*	*Flesch*	*FleschKincaid*
Food Products	2	7	21.407	25.598	17.264	18.545	33.184	14.192
Beer and Liquor	4	3	21.647	26.682	17.628	19.487	31.677	15.182
Entertainment	7	2	21.770	25.203	17.590	18.730	32.244	14.305
Printing and Publishing	8	3	18.515	33.095	14.504	19.631	28.189	15.537
Consumer Goods	9	2	21.214	26.928	17.233	16.373	39.513	12.428
Medical Equipment	12	1	22.306	19.750	18.592	19.546	26.934	15.591
Pharmaceutical Products	13	4	20.982	27.489	16.738	19.554	31.335	15.195
Chemicals	14	4	22.476	24.544	18.281	19.807	30.474	15.445
Textiles	16	1	22.235	24.382	18.116	19.081	32.068	14.857
Construction	18	6	21.479	26.105	17.268	19.387	29.326	15.129
Steel Works	19	4	22.005	21.601	17.879	19.872	29.417	15.510
Machinery	21	4	21.770	24.880	17.650	19.160	30.111	14.945
Automobiles and Trucks	23	1	20.694	27.553	17.048	19.449	30.357	15.142
Aircraft	24	1	23.203	22.684	19.191	19.712	30.202	15.376
Precious Metals	27	1	20.946	28.860	16.901	18.532	33.857	14.154
Petroleum and Natural Gas	30	2	21.129	25.957	16.981	19.494	30.830	15.239
Utilities	31	1	21.060	28.263	16.780	21.019	27.428	16.579
Communication	32	2	25.536	18.315	21.321	19.426	30.028	15.200
Personal Services	33	1	22.484	23.584	18.224	18.509	32.374	14.225
Business Services	34	19	20.993	27.324	16.934	18.818	32.045	14.588
Computers	35	1	21.902	26.033	17.601	20.269	27.344	15.853
Electronic Equipment	36	5	21.581	24.432	17.429	18.160	33.492	14.047
Measuring and Control Equipment	37	1	22.083	18.432	17.702	18.395	31.647	14.109
Business Supplies	38	1	21.175	29.940	16.776	18.187	35.766	13.886
Transportation	40	1	22.599	24.690	18.282	20.687	25.292	16.178
Wholesale	41	4	21.762	25.643	17.564	18.621	31.668	14.596
Retail	42	6	21.371	26.042	17.170	18.673	32.336	14.450
Other	48	1	22.099	24.314	17.773	19.291	32.218	14.976
Total number of firms		*89*						

Table B2 shows the Fama and French (1997) 48-industries averages of the reading indexes. The first three columns show the industry name, number, and number of firms included in the sample. The next three columns show the averages of the reading indexes when the old standards on the audit report are followed. The last three columns show the averages of the reading indexes when the standards on the extended audit report are applied.

Table B3

Big Four and non-Big Four subsample regression results on the relation between the extension of the audit report and the Fog Index, Flesch Index and Flesch-Kincaid Index

Panel A: Sample for *Big4* = 0 (non-Big Four auditors)			
	(1) Fog	(2) Flesch	(3) FleschKincaid
EAR	-3.983*** (0.010)	10.230** (0.027)	-4.007**
Earnings	1.755 (0.592)	-7.215 (0.611)	-0.023 (0.991)
log(Size)	-0.261 (0.581)	-0.201 (0.919)	-0.528 (0.162)
log(MTB)	0.446 (0.671)	0.419 (0.925)	1.111 (0.181)
abs(SUE)	-0.052 (0.520)	0.043 (0.897)	-0.025 (0.600)
AuditChange	-1.125 (0.379)	4.790 (0.387)	-1.361 (0.156)
Constant	22.530*** (0.006)	27.774* (0.065)	19.252*** (0.003)
Observations	18	18	18
R^2	99.8%	99.4%	99.9%
Adj. R^2	98.1%	94.8%	99.3%
F-statistic	59.054** (df = 15; 2)	21.666** (df = 15; 2)	161.508*** (df = 15; 2)
Panel B: Sample for *Big4* = 1 (Big Four auditors)			
EAR	-3.418*** (0.000)	8.770*** (0.000)	-3.456*** (0.000)
Earnings	0.504 (0.615)	-1.874 (0.560)	0.711 (0.475)
log(Size)	0.016 (0.827)	-0.173 (0.454)	0.009 (0.899)
log(MTB)	-0.063 (0.635)	0.033 (0.938)	-0.044 (0.738)
abs(SUE)	0.001 (0.832)	-0.010 (0.618)	0.003 (0.616)
AuditChange	-0.061 (0.899)	-0.391 (0.800)	0.087 (0.857)
Constant	21.136*** (0.000)	27.424*** (0.000)	17.009*** (0.000)
Observations	160	160	160
R^2	59.0%	45.4%	60.5%
Adj. R^2	47.0%	29.4%	48.9%
F-statistic	4.923*** (df = 36; 123)	2.843*** (df = 36; 123)	5.226*** (df = 36; 123)

*Table B3 shows the regression results of the additional Big Four and non-Big Four subsample regression analysis. Panel A shows the regression results for the subsample only including observations of firms who are audited by a non-Big Four auditor (Big4 = 0). Panel B shows the regression results for the subsample only including observations of firms who are audited by a Big Four auditor (Big4 = 1). The first column shows the regression coefficients and their p-values in parentheses using the Fog Index as dependent variable. In the second column, the Flesch Index is used as dependent variable. In the third column, the Flesch-Kincaid Index is used as dependent variable. Statistical significance is presented as: * p<0.1; ** p<0.05; *** p<0.01. All regressions include the 48-industry dummies of Fama and French (1997) and year-fixed effects as well.*