DOCTORAL DISSERTATION

Whether Longer Annual Reports Are Better: Evidence From Annual Reports Issued By Foreign Firms Listed On The US Stock Exchange

Graduate School of International Social Sciences

YOKOHAMA NATIONAL UNIVERSITY

NGUYEN THI THUY PHUONG

July 2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 2: The Determinants of the Length of Annual Reports: Evidence from 20-F Filings

Chapter 3: How Plain English Use Affects the Length of Annual Reports: Evidence from the Annual Reports of Foreign Firms listed on the US Stock Exchange

Chapter 4: Disclosure Characteristics of Annual Reports and Information Asymmetry: Evidence from the Annual Reports of Foreign Firms Listed on the US Stock Exchange

Acknowledgments

Doing Ph.D. has never been easy for all students and takes a lot of valuables things from us, therefore, there are no words which fully explain my exciting feelings when writing these acknowledgments. This is really one of great jobs that I have done in my life. In fact, I deeply understand that I was unable to finish this dissertation if I had not received supports from MEXT, my family, my supervisors, and my friends. I would like to take this opportunity to express my gratitude to all my supporters.

Firstly, I would like to thank the Japanese Ministry of Education, Culture, Sports, Science and Technology who granted scholarships for me in 3.5 years. Your financial supports are the prerequisites to make my Ph.D. program possible. I also highly appreciate the supports of members in the MEXT team and the GSISS staff of Yokohama National University who take care of me from the first days when I came to Japan until now. Especially, I am thankful to Mikako Maekawa, my tutor, who help me to settle down my early life in Japan. Your support makes my life in Japan become brighter.

I would like to give my sincere appreciation to my supervisor, Associate Professor Akihisa Kimura, for your knowledge, your patience, and your academic supports. You have inspired me for my Ph.D. years and made it smoothly. Thank you for listening when my feelings went down and for your help in solving my Ph.D. troubles. I hope this is the beginning of our upcoming cooperation in the future. I would also like to express my gratitude to Professor Brian P. Miller for his sharing which is primarily contributed to shorten my Ph.D. time. I am also grateful for the discussion and comments from Professor Satoru Otaka and Professor Toshio Hara at Yokohama National University.

I would like to give special thanks to my dear friends who help me to survive throughout my lonely journey in Japan. Somehow, you make a contribution to my current achievements. Thanks for the delicious meals and memorable dinners you served me, Dang Phuong Linh, Pika Tran, Ho Thao Nguyen. Thank you for your help when I met troubles, Uyen Toly, Jonas Truong, Luu Dang, Mele Visi T, Nguyen Quynh Trang, Binh Le, Hero Nguyen, Ngo Van Toan, Phan Thanh Ngoc, Truong Trong Hieu. I would like to send my sincere thanks to my other friends who are always by my side in any circumstances. I also highly appreciate the programming assistance of Thang Viet, Huynh and Thang Duy, Dang. Thank you so much for your help.

Last but not least, I owe my family a great deal. There are no words to show my appreciation for your supports, especially my mom who helps me to take care of my small family during the time I was not at home. Thanks, my husband, for his patience, and his confidence in me. I would like to apologize to my small kid for all broken-hearted feelings when I did not stay with you during your childhood. I would like to dedicate this dissertation and all my achievements to my beloved family.

Finally, I understand that doing Ph.D. is not the end, it is just a start for the new journeys in my life. I hope all my findings in this dissertation will be continued in the future.

Chapter 1: Introduction

According to efficient market theory, stock prices reflect publicly available information, which is mostly extracted from financial statements, the primary information source on firms for the market as well as for investors, regulators, and other users. The US Securities and Exchange Commission (SEC) understands the crucial role of annual reports and "requires public companies to disclose meaningful financial and other information to the public. This provides a common pool of knowledge for all investors to use to judge for themselves whether to buy, sell, or hold a particular security" (SEC, 2013). Investors, however, do not enjoy all the benefits of this regulation because of a decrease in the readability of annual reports as well as an extremely increase in the length of those reports, which leads to excessively high financial and time costs for extracting information (Bloomfield, 2002). This issue has attracted numerous researchers. Following previous research, this dissertation focuses on the complexity of annual reports issued by foreign firms listed on the US Stock Exchange. This sample set is not popular in previous research relating to the readability of annual reports. By extending the sample set to 20-F rather than 10-K, this dissertation hopes to provide a whole picture of the current tendency in firms' disclosures.

This dissertation follows the guidelines of previous research especially the research of Li (2008) who is the pioneer in solving the readability of a large sample. Packages of Perl language and StyleWriter Professional Software are used in analyzing the complexity of annual reports which are automatically downloaded by Python scripts on the EDGAR website. All data relating to firms' performance and stock markets are extracted from Thomson Reuters Datastreams. This dissertation reports robustness tests, to demonstrate that all my inferences are not sensitive. To some extent, this dissertation is the first research which provides the empirical research for new issues relating to the readability of annual reports such as plain English regulations of SEC or the information asymmetry.

Overall, the remainder of this dissertation is structured as follow. The overview and key findings are summarized in the introduction. The contributions and limitation are demonstrated in the conclusion. The main ideas are discussed more details in the three chapters which bridge the gaps in the literature. The three main chapters proceed as the structure of three research papers. The first chapter presents the tendency of the length of 20-F over time and the determinants of the number of words in 20-F. Chapter 2 mentions how the regulations of SEC on plain English affect the length of 20-F. Chapter 3 demonstrates the associations among the disclosure characteristics of 20-F and the information asymmetry. The summary of key findings displayed in the three main chapters are mentioned in the next sections:

Chapter 2: The Determinants of the Length of Annual Reports: Evidence from 20-F Filings

This chapter examines the length of annual reports for the sample of 20-F forms which are annually filed by foreign firms listed on NYSE and NASDAQ. This sample set is not popular in previous research, moreover, the fact that 20-F forms are issued by firms operating in different

countries with different business environments and different accounting standards owns distinguish characteristics as compared to sample set of 10-K. We record the gradual increase in the length of 20-F over the 10-year period with an average rate of 32% from 2004 to 2013. This tendency of 20-F is more serious than its 10-K. We also realize that operating complexity, poor performance, high volatility, adopted accounting standards, and native language are the key determinants materially affecting the length of annual reports issued by foreign firms in the United Stock Exchange. In which, the adopted accounting standards and native language are the two new determinants of the length of annual reports for the sample set of 20-F.

Chapter 3: How Plain English Use Affects the Length of Annual Reports: Evidence from the Annual Reports of Foreign Firms listed on the US Stock Exchange

Understanding the vital roles of annual reports to users, SEC soon issues the strict regulations of language in firms' disclosures. According to SEC's regulations, firms are forced to use plain English in their disclosures, however, SEC recognizes these regulations are possible to cause the increase in the length of disclosures. This chapter provides the empirical results of the impacts of plain English use on the length of annual reports for a sample of 20-F forms filed annually by foreign firms listed on NYSE and NASDAQ. We find that foreign firms adopt plain English in their disclosures over time with shorter sentences, less jargon, and more readable writing styles; however, the length of their 20-F forms significantly increases from 2004 to 2013. This chapter also tests the effects of four regulated elements of plain English such as the average words per sentence, jargon, passive, and writing styles on the length of 20-F. We recognize that the impacts of plain English use on the length of annual reports are mixed. The length and plain English adherence level in 20-F forms show statistically significant differences between accounting standards as well as between English-speaking countries and other countries.

Chapter 4: Disclosure Characteristics of Annual Reports and Information Asymmetry: Evidence from the Annual Reports of Foreign Firms Listed on the US Stock Exchange

Overall, there are three main characteristics which are recently concerned by researchers: the level of disclosure (how much is disclosed), the tone (what do you say), and the transparency (how information is communicated). This chapter concentrates on the associations among two most currently interested disclosure characteristics of annual reports and the information asymmetry. It illustrates that the level of disclosures, measured by the number of words in annual reports, is negatively associated with the information asymmetry between informed and uninformed investors during the window days. In contrast, our research shows that the improvement in readability or transparency of annual reports does not significantly affect the information asymmetry among investors. Loss firms own higher information asymmetry, although, the level of disclosures of annual reports issued by loss firms have stronger impacts on the improvement of information asymmetry as compared to profit firms. Annual reports following

IFRS contain more information that improves the information asymmetry as compared to annual reports adopting US-GAAP.

In summary, this dissertation provides some interesting key findings to academic research and practice. It also has some new positive ideas about the current tendency of the length of annual reports as compared to previous research.

Chapter 2: The Determinants of the Length of Annual Reports: Evidence from 20-F Filings

ABSTRACT

This chapter examines the length of annual reports for the sample of 20-F forms which are annually filed by foreign firms listed on NYSE and NASDAQ. We record the gradual increase in the length of 20-F over the 10-year period with an average rate of 32% from 2004 to 2013. We also realize that operating complexity, poor performance, high volatility, adopted accounting standards and native language are the key determinants materially affecting the length of annual reports issued by foreign firms in the United Stock Exchange.

Keywords: 20-F, length of annual reports, determinants

1. Introduction

According to efficient market theory, stock prices reflect publicly available information, while annual reports are considered as the main source of information. Understanding the crucial role of annual reports to users, SEC and FASB have promulgated a series of rules and instructions for firms to file the annual reports. However, the extant research records the substantial increase in the complexity of annual reports which causes high financial and time cost for investors to analyze (Bloomfield, 2002).

In detail, previous research, as well as reports of professional organizations, has reported a dramatic increase in the number of words in 10-K (KPMG, 2012; Cazier and Pfeiffer, 2016). They showed that operating complexity, redundancy, and firm residual disclosures cause the difference in the length of 10-K. Recently, Li (2017) recorded the rapid increase in the repetitive disclosures of 10-K; however, Li said that repetitive disclosures do not mean less informative like SEC's suggestion. Most of the other research assumes that the increase in the length of annual reports and the decline in the readability of annual reports undoubtedly create negative impacts on users along with stock markets (Li, 2008; Miller, 2010; You and Zhang, 2008, Lee, 2012; Lehavy et al, 2011). This requires SEC and other standard setters to spend more efforts to understand the length of annual reports and what determinants of length of annual reports are.

Continuing the previous research relating to the length of annual reports, our research uses the sample set of 20-F rather than 10-K to test the length of annual reports and its respective determinants. Our findings show some interesting contributions to practice and academic research. In this research, we choose the sample set of the annual reports of foreign firms listed on the US Stock Exchange, which is not a popular selection in previous research. Understanding the length of 20-F forms is expected to widen the knowledge on the length of annual reports. Our research adds one more piece to the whole picture of the significant increase in the length of annual reports occurring not only in 10-K but also in 20-F. The increase in the length of 20-F is more serious. In

addition to operating complexity and firm performance, we also record three more new and crucial determinants of length of annual reports, namely volatility, accounting standards, and language. Realizing new determinants help users and standard setters to deeply understand the behaviors of firms in issuing their annual reports.

The article is organized as follows. The following section displays the change in the length of annual reports as well as the impacts of the increase in the length of annual reports on stock markets and users. Section 3 presents the sample selection and shows how to measure the length of annual reports. Section 4 identifies what determinants of 20-F length are and predicted signs. Section 5 presents our findings and discussion. Finally, Section 6 gives some concluded remarks.

2. Literature Review

Each firm has frequent disclosures through the fiscal year with the hope of updating more timely information to investors; among of firms' disclosures, the annual reports contain more price-relevant information (You and Zhang, 2009). According to SEC's requirements, firms must annually file form 10-K for the US firms and form 20-F for the foreign firms listed on the US Stock Exchange, respectively. Those forms are strictly regulated by SEC to protect investors; however, the length of annual reports has significantly increased over time with overloaded information and redundancy (Cazier and Pfeiffer, 2016; Li, 2017).

2.1 The significant increase in the length of annual reports

To make sufficient, accurate information available to investors for making decisions, SEC "requires public companies to disclose meaningful financial and other information to the public. This provides a common pool of knowledge for all investors to use to judge for themselves whether to buy, sell or hold a particular security" (SEC, 2013). Besides, SEC, as well as FASB, force all firms to strictly obey the mandatory rules specifically to the financial statements, related notes, and footnotes. They also concern about the language used in disclosures of firms. They compel the use of plain English in communication to users which emphasizes to "clarity, not brevity" and leads to "increase the length of particular sections" (SEC, 1998). These suggestions of SEC mean that the longer annual reports are, the more information annual reports provide.

A significant increase in the length of annual reports has been recorded by researchers and institutional organizations. Miller (2010) reports that the number of words in 10-K increased from 25,515 in 1995 to 40,579 in 2006. Similarly, KPMG (2012) records a 16% increase in the length of annual reports and a 28% increase in the length of notes in the period of 2004 to 2010. More recently, Cazier and Pfeiffer (2016) present the average number of words in 10-K during the period from 2003 to 2012 at 55,335.

Unfortunately, the rapid increase in the length of annual reports is assumed not to provide more information like SEC's expectation. Li (2017) reports the percentage of information in notes repeated in MD&A gradually rises from 19% in 1995 to over 26% in 2013. Li also presents a

higher percentage of repetitive disclosures in the significant accounting policy, litigation and revenue notes. The reasons for the increase in the repetition of disclosures are assumed as the appearance of special events such as new CEO, issued equity or missed prior year's earnings benchmark. In addition, Cazier and Pfeiffer (2016) provide evidence for reasons of the significant increase in the length of 10-K. They showed that the operating complexity, redundant disclosures between SEC and FASB, and residual disclosures of firms cause such increase of 10-K.

The increase in the length of annual reports due to the repetitive disclosures has different views. According to SEC, the repetition in disclosures of annual reports does not provide more information to investors and makes investors lose the concentration in other material information (SEC,2003). In contrast, the succession hypothesis has the different view about the repetition of communication. This hypothesis supposes the repetition is informative since investors do not read all sections of annual reports and spend around ten to thirty minutes for reviewing the annual reports (SEC, 2008; Arnold, Bedard, Phillips and Sutton, 2010), therefore, the repetitive disclosures in different sections of annual reports help investors not to miss the important information. Additionally, the repetitive disclosures in annual reports improve the information processing ability of investors since the repetition enhance the understanding of investors when information is displayed in different ways (Li, 2017).

2.2 Negative impacts of the increase in the length of annual reports.

The dramatic increase in the length of annual reports has attracted the interest of researchers under controversial views. On the one hand, some research considers the length of annual reports as the readability of annual reports in the sense that the increase in the length of annual reports means the decrease in the readability of annual reports, thereby leading to negative impacts on stock markets and users. On the other hand, other research supposes the increase in the number of words in annual reports improves the readability of these reports. In fact, readability of annual reports is extremely important to external users and managers. For investors and shareholders, readability of annual reports prevents them not only from omitting important information of firms, but also from misunderstanding the current business performance of firms. Most importantly, readability of annual reports helps them save time and cost for acquiring information. Lower readability of annual reports causes investors to spend more money to buy analyst services and time for further search (Lehavy, Li, Merkley, 2011). In terms of managers, they must follow SEC's regulations about how to disclose information which requires annual reports of firms to "be readable and easy to understand to all shareholders" (SEC, 1998a). Additionally, the previous research shows that the way of disclosure that managers strategically decide can reduce the unfavorable views on negative economic consequences of investors to firms and prevent transparency.

Li (2008) reports that firms with poor performance provide the longer annual reports with the hope to hide adverse information to investors, hence the annual reports with more words

prevent small investors from trading (Lawrence, 2013; Miller, 2010). Some research also records the more underreaction of the stock market when annual reports become longer (Callen, Khan, & Lu, 2010; Lee, 2012; You & Zhang, 2009). The increase in the length of annual reports also causes greater dispersion, lower accuracy, greater uncertainty in analysts' earnings forecasts (Lehavy et al, 2011).

However, some research records that the increase in the length of annual reports improves the readability of annual reports which is highly appreciated by investors, such as Cheung and Lau (2016) and Lundholm et al (2014). They show that the longer annual reports due to different expressions upgrade the readability of annual reports. Other research also provides the evidence that firms with better adhering plain English improve the readability of their annual reports in regardless of the significant increase in the length of annual reports (Chueng and Lau, 2016).

3. Sample selection

In contrast to existing research, we use the annual reports of foreign firms listed on the US Stock Exchange. As SEC's requirements, foreign firms listed on the US Stock Exchange, normally called "foreign private issuers", trade less than 50% of their shares on the US Stock Exchange, and annually fill the 20-F forms rather than 10-K like the US firms. We use Python to automatically collect all the 20-F filings in the period from 2004 to 2013 available on EDGAR (Electronic Data Gathering, Analysis, and Retrieval) in the period of 2005 to 2014 and find 7,588 filings. We remove all filings in the finance and insurance industries and filings with fiscal year on the different date of December 31st. We only choose filings of firms listed on NYSE or NASDAQ which they have RIC codes (i.e., .N or .O) available on Thomson Reuters Datastream. We eliminate all filings which have missing data. We also accept filings in text files which are compatible with using Perl language to measure the number of words. Finally, our sample has 1,527 observations (see Table 1).

[INSERT TABLE 2.1 HERE]

To identify the length of annual reports, we need to clean the raw text files downloaded from EDGAR which contains both images, tables, figures. We follow the guidelines of Bonsall et al (2017), Li (2008), and Loughran and McDonald (2014a, 2014b) for cleaning (Appendix 2.1). We use the Lingua::EN:: Fathom package and other packages of Perl language to calculate the number of words of filings after cleaning. Perl language is proved as the better way for analyzing the text of the large sample sets as compared to different solutions (Li, 2008).

4. Determinants of 20-F length

Cazier and Pfeiffer (2016) shown that there are three determinants of 10-K length which are operating complexity, redundancy, and residual disclosure. Applying findings of Cazier and Pfeiffer (2016), Li (2008), Lundholm et al (2014) into our setting for 20-F filings, we use the specification in Equation (1) to identify the determinants of 20-F length.

$$NUMWORDS_{it} = \beta_0 + \beta_1 SIZE_{it} + \beta_2 DEBT\_RATIO_{it} + \beta_3 LOSS_{it} + \beta_4 PTB_{it} + \beta_5 VOLATILITY_{it} + \beta_6 IFRS_{it} + \beta_7 ENGLISH_{it} \quad (1)$$

Overall, we identify there are seven determinants in five groups explaining the difference in the length of 20-F filings. The definitions and expected signs of variables are summarized in Appendix 2.2.

4.1 Operating Complexity

Based on suggestion of Cazier and Pfeiffer (2016), we assume that firms with more transactions and more complicated operations have more disclosures at the end of fiscal year since they have more explanations about their business activities to investors under disclosure regulations of SEC and FASB, therefore, it is difficult for those firms to shorten their annual reports. To measure the operating complexity of firms, we use two different measurements:

We use the size as the proxy of operating complexity since firms with bigger size mean more business activities or more frequently complicated transactions or more operating complexity. In fact, there are several ways to measure the operating complexity of firms such as size, age, business or geographic segments but SIZE is the most popular proxy for operating complexity that is used in most of the prior research (Li, 2008; Cazier, Pfeiffer, 2016). More importantly, Size is easy to get reliable data while others are so difficult to acquire in Thomson Reuters Datastream. SIZE is identified by the natural logarithm of the market value of equity at the end of fiscal year.

Information on debts of firms is extremely sensitive to different users and necessary to disclose more details under regulations. We expect annual reports of firms with higher debt ratio are longer. DEBT_RATIO is equal to total liabilities scaled by total assets at the same year.

4.2 Poor Performance

According to Li (2008), firms with poor performance often provide longer annual reports with fewer self-referential, more causation words, more future tense verbs to hide the adverse information as compared to reports of firms performing well. Therefore, we use LOSS as an indicator variable set equal to 1 when firms have positive earnings, and 0 otherwise. Additionally, price-to-book is also the other indicator of the performance of firms. PTB is equal to the market price per share divided into book value per share at the end of fiscal year. We expect firms with loss and lower price to the book have longer annual reports to explain more details about business activities or obfuscate the disadvantage information.

4.3 Uncertainty

The uncertainty in stock returns is expected as another critical determinant of the length of annual reports. Under the efficient market theory, stock prices reflect the available information to markets, therefore firms with more information cause the volatility in stock prices. Meanwhile, annual reports are considered as the official resources of information to investors. It means that if

stock prices of firms become more volatile, the annual reports are projected to become longer with more information or more explanations. VOLATILITY is equal to the daily standard deviation of adjusted stock returns during a year. We expect there is a positive association between the volatility of stock returns and the length of annual reports.

4.4 Accounting Standard Systems

To improve the comparability of annual reports issued by firms globally, the harmonization and convergence of global accounting systems have occurred in recent years. Since 2007, SEC has recognized that US-GAAP and IFRS are converging sufficiently. Therefore, it allows foreign firms to eliminate a reconciliation to US-GAAP if they adhere to IFRS. The SEC has also started to consider allowing the US firms to file their annual reports by adopting IFRS (Topic Gateway Series, 2008). Kim et al. (2012) find that such SEC's elimination does not cause the negative impacts on stock markets and users. Firms adhering to IFRS do not increase more disclosures or the frequency of disclosures.

However, there have been some material differences existed between US-GAAP and IFRS which demand to have more declarations from firms (PwC, 2016). Therefore, we suppose annual reports of foreign firms under IFRS are assumed to be longer than annual reports following US-GAAP since these firms want to increase their comparability with the US-firms.

SEC allows foreign firms to choose IFRS or US-GAAP for their annual reports provided that firms must declare which accounting standard system is applied to their annual reports at the beginning of 20-F. Based on firms' announcement, we sort our filings into two different groups, namely IFRS and US-GAAP. IFRS is an indicator of which accounting standard is followed by firms and equal to 1 if firms adopt IFRS, and 0 otherwise.

4.5 Language

Lundholm et al. (2014) find that foreign firms in non-English-speaking countries issue the longer MD&A section than foreign firms located in English-speaking countries. Following Lundholm et al. (2014), we expect the length of the annual reports issued by foreign firms located in English-speaking countries is longer than its of annual reports provided by foreign firms in non-English speaking countries, implying that the advantage in the language of native speakers can shorten their annual reports.

Based on firms' declaration on 20-F, we identify the executive locations of firms. We follow Lundholm et al. (2014) for sorting filings into English native speaking countries and others. ENGLISH is equal to 1 if firms located in English speaking countries, and vice versa equal to 0. The list of countries and their language are mentioned in Appendix 2.3.

5. Findings and Discussion

5.1 How does the length of annual reports change over time?

The previous research, as well as reports of professional organizations, have recorded the significant increase in the length of annual reports. Our research has the similar result with a dramatic increase in the length of 20-F at the average rate of 32% during 10 years from 2004 to 2013 which is equivalent to 19,508 words rising.

[INSERT FIGURE 2.1 HERE]

On average, each 20-F filing contains 61,908 words in 2004. Surprisingly, the number of words in 20-F has gradually risen over 10 years and reached to 81,416 words in 2013. Figure 1 describes the change in the length of annual reports over 10 years. Such increase in the length of annual reports does not mean to cause the decline in the readability of 20-F over time (not tabular), the change in the writing styles is considered as the reason for this tendency. Foreign firms become better adoption to the suggestions of SEC in using plain English for communication. In detail, they write more sentences but shorter sentences to improve the readability of 20-F over time.

5.2 Descriptive statistics

Table 2.2 illustrates the descriptive statistics of all variables in Equation (1). The average number of words in 20-F form is 73,097 words, more than double those in 10-K which is reported at around 37,000 words in the previous research (Li, 2008; Bonsall et al., 2017). However, extremely long annual reports do not mean to be more informative since they contain more repetitive disclosures (Li, 2017; Cazier and Pfeiffer, 2016). The difference in the length of 20-F and 10-K is explained by the fact that foreign firms located in different countries, different legal and business environment so that they have to disclose more information to the US investors (Lundholm et al, 2014). Foreign firms with more explanation hope to increase the comparability and reliability of their annual reports to the US investors who underweight the equity of foreign firms (French & Poterba, 1991).

[INSERT TABLE 2.2 HERE]

The average size of firms in our sample is more than 12,425 thousand dollars equivalent to the natural logarithm of 7.07. The mean of debt ratio is around 23% which means that around 23% of total assets are financed by debts. The mean of LOSS indicates that there are around 74% observations in our sample owning positive earnings. Price to book value and Volatility of stock returns are 2.55 and 0.0314 on average, respectively. The number of filings in our sample adopting IFRS is approximately equal to the number of filings adhering US-GAAP. There are around 42% of annual reports filed by English native speakers.

Table 2.3 shows the Pearson correlation between variables in Equation (1). There is a significant correlation between NUMWORDS and independent variables. The length of annual reports has the positive correlation with the size of firms, debt ratio, accounting standards, in contrast, the negative correlation between the length of 20-F with the loss, price-to-book value, volatility, and English. Although other variables are correlated with each other, the correlation

magnitudes are not large. Especially, SIZE shows the negative correlation with the volatility. Our result is consistent with previous research (Fu, Kraft, and Zhang, 2012; Amiram, Owens and Rozenbaum, 2016). This negative correlation is explained by leverage effect and time-varying risk premia (Chueng and Ng, 1992; Duffee, 1995; Perez-Quirross and Timmermann, 2000).

[INSERT TABLE 2.3 HERE]

5.3 Empirical results

Table 2.4 summarizes the regression results of Equation (1). Column 1 shows the regression results of Equation without industry fixed effects and year fixed effects. Column 2 and Column 3 presents the coefficients of Equation (1) when either industry fixed effects or year fixed effects are included in the regression. Column 4 illustrates the results when these both fixed effects are included in the regression. Overall, the coefficients and significance among models are not so different. Adjusted R square is improved over 4 models; however, the degree of freedom is declined when we include fixed effects into regressions.

[INSERT TABLE 2.4 HERE]

As shown in Table 2.4, the length of annual reports is positively associated with the operating complexity. The coefficients of SIZE and DEBT_RATIO are 4,825 with t-statistic of 14.06 and 20,150 with t-value of 6.58 respectively. The positive coefficients of size and debt ratio mean that firms with bigger size and/or higher debt ratio issue longer annual reports than other firms, ceteris paribus.

In contrast, the number of words in 20-F is negatively correlated with the performance of firms. This result is consistent with the results of Li (2008) and Aymen, Mhamed, and Badreddine (2016). They assume firms with negative earnings issue longer annual reports. The coefficients of LOSS and PTB are negative but significant at 99% confidence level. Firms getting a loss or lower price-to-book value provide longer reports since they want to obfuscate the adverse information to investors.

The regression result also supposes for the fact that daily stock returns of firms are more volatile, firms disclose more information in their annual reports which leads to longer annual reports. The coefficient of VOLATILITY is 169,469 with t-statistic at 3.95. Other words, the difference in the length of annual reports results from the variation in the volatility of stock returns.

Interestingly, filings following IFRS are extremely longer than filings under US-GAAP. The coefficient of IFRS is positive and significant at 99% confidence. Foreign firms listed on the US Stock Exchange adopting IFRS have to provide more explanation the difference between IFRS and US-GAAP to investors which lengthen their annual reports. Another reason is that IFRS is principle-based and have a broad guideline, while US-GAAP is rule-based with a list of detailed rules. Principle-based accounting requiring more disclosures than rule-based accounting, which leads to a substantial increase in the length of annual reports issued by firms adopting IFRS. In

addition, more disclosures are expected to attract more US investors who are familiar with US-GAAP.

Table 2.4 also shows that annual reports of foreign firms located in English speaking countries are significantly shorter than those of firms in non-English speaking countries. ENGLISH has a negative coefficient. This result assumes that native speakers with advantage language are likely to shorten their annual reports under the same formats and same regulations, however, the length of sentences in those reports are longer than reports issued by firms in non-English speaking countries. Our result is consistent with Lundholm et al. (2014) who find that MD&A of 20-F in non-English speaking countries are longer than MD&A of 20-F in English speaking countries. However, the increase in the length of 20-F in non-English speaking countries does not worsen the readability of those reports (not tabular).

Our results do not significantly change when including industry fixed effects and/or year fixed effects, however, the significant level of IFRS decrease from 99% to 95% when we insert industry fixed effects into models. As concluded, the results over different models remain. Adjusted R squared of our models are from 26% to 33%. Other words, our independent variables can explain 26% to 33% the change of dependent variable.

Overall, the empirical results support for the expected signs mentioned above. In conclusion, the length of 20-F becomes longer when firms become bigger, higher debt-ratio, higher volatility. Conversely, firms with positive earnings and higher P/B shorten their annual reports. Additionally, adopted accounting standards and native language substantially affect the length of annual reports.

6. Conclusion

Annual reports are considered as the official communication channel of firms to investors, therefore, SEC, as well as FASB, seriously concern the format and contents of annual reports. They set a series of regulations and guidelines for firms to annually file forms. The more complicated business environment and various regulations make annual reports become more difficult-to-read for investors. The previous research recognizes the dramatic increase in the length of annual reports, moreover, the repetitive disclosures also rise rapidly. Similarly, our research records the length of 20-F forms issued by foreign firms listed on the US stock exchange rapidly increase over 10 years by 32% which is more serious than 10-K forms. Foreign firms increase the length of their reports due to shortening the sentences which is followed the SEC's suggestions.

In addition to operating complexity and firm performance, we realize that the volatility, accounting standards that firms apply to their annual reports and issuers' native language crucially determine the length of annual reports. This result is expected to assist the users as well as standard setters in understanding which factors determine the length of annual reports. From that, they can adjust their behaviors to adapt to the complicated tendency of current annual reports.

Despite our interesting findings, our research has still contained some open issues that need to deeply be investigated. Our research concentrates on how the number of words in 20-F change over time and the determinants of the length of annual reports, but we do not test whether longer annual reports are better/informative or not. Also, we have not yet concerned about the repetitive disclosures in 20-F forms which are shown as the serious problem in 10-K. We leave these issues to future work.

APPENDICES

APPENDIX 2.1: CLEANING THE RAW TEXT FILES OF THE 20-F FORMS

To clean the raw text files downloaded from EDGAR, we follow the instructions of and Bonsall *et al.* (2017), Li (2008), and Loughran and McDonald (2014a, 2014b). We use the packages File::Slurp; HTML::format text; and HTML::TreeBuilder; HTML::Entities;Text::Unidecode to clean the raw 20-F files, following these steps:

1. Remove format design in the raw files.
2. Remove all content between the <XBRL> and </XBRL> tags.
3. Remove all tables with more than 15% numeric characters.
4. Remove all markup tags.
5. Remove other textual expressions such as newline and underscored characters.

APPENDIX 2.2: VARIABLE DEFINITIONS AND EXPECTED SIGNS

Variable Name	Expected Sign	Definition
NUMWORDS	N/A	A number of words acquired by Lingua::EN:: Fathom package of Perl language for 20-F after cleaning.
SIZE	(+)	Natural logarithm of the market value of equity at the end of fiscal year.
DEBT_RATIO	(+)	Total liability scaled by total assets
LOSS	(-)	An indicator variable which is equal to 1 if firms have positive net income, 0 otherwise.
PTB	(-)	Market price to book value per share
VOLATILITY	(+)	Daily standard deviation of stock returns during a year.
IFRS	(+)	An indicator variable which is equal to 1 if firms adopt IFRS, 0 if firms adhere to US-GAAP
ENGLISH	(-)	An indicator variable which is equal to 1 if firms locate in English speaking countries, 0 if firms' residence in non-English speaking countries.

APPENDIX 2.3: LIST OF ENGLISH-SPEAKING AND NON-ENGLISH-SPEAKING COUNTRIES

Based on the declaration in the first part of the 20-F forms, the executive locations of firms are identified. The classification of Lundholm *et al.* (2014) is used to sort filings of foreign firms into the group of firms locating in English speaking countries and the group of firms in non-English speaking countries.

Country	***English***	***Country***	***English***
Anguilla	1	*Liberia*	1
Argentina	0	*Luxembourg*	0
Australia	1	*Mexico*	0
Austria	0	*Monaco*	0
Bahamas	1	*Netherlands*	0
Belgium	0	*New Zealand*	1
Brazil	0	*Norway*	0
Bulgari	0	*Panama*	0
Canada	1	*Papua New Guinea*	1
Chile	0	*Peru*	0
Columbia	0	*Philippine*	1
Denmark	0	*Portugal*	0
Finland	0	*Russia*	0
France	0	*Singapore*	1
Germany	0	*South Africa*	1
Greece	0	*Spain*	0
Hong Kong	1	*Sweden*	0
India	1	*Switzerland*	0
Indonesia	0	*Taiwan*	0
Ireland	1	*Turkey*	0
Israel	1	*United Kingdom*	1
Italy	0	*Uruguay*	0
Japan	0	*Venezuela*	0
Korea	0		

TABLE 2.1: Sample selection

We use Python to collect all the 20-F filings from 2004 to 2013 from EDGAR's website. We remove all filings in the finance and insurance sectors and filings with fiscal year on the different date of December 31st from our sample. We only choose filings of firms having RIC codes (ie. .N or .O) available on Thomson Reuters and filings which have no missing data on Thomson Reuters.

Total of observations collected from EDGAR	7,588
Number of firms	1,475
Number of firms whose fiscal years end on December 31	966
Number of firms whose RIC codes available on Thomson Reuters	264
Number of observations without missing data	**1,527**
In which:	
Agriculture, forestry, and fishing	2
Mining	118
Construction	5
Manufacturing	758
Transportation, communication, electric, gas, and sanitary services	385
Wholesale trade	23
Retail trade	13
Services	223

TABLE 2.2: Descriptive Statistics

	N	Mean	Std. Dev	Q1	Median	Q3
NUMWORDS	1,527	73,097	28,417	54,989	68,639	84,510
SIZE	1,527	7,07	2.52	5.20	7.00	9.13
DEBT_RATIO	1,527	0.23	0.21	0.04	0.19	0.35
LOSS	1,527	0.74	0.44	0	1	1
PTB	1,527	2.55	13.09	1	1.66	2.62
VOLATILITY	1,527	0.0314	0.02	0.0192	0.0269	0.0393
IFRS	1,527	0.5029	0.52	0	0	1
ENGLISH	1,527	0.4119	0.49	0	0	1

TABLE 2.3: Pearson correlation between variables

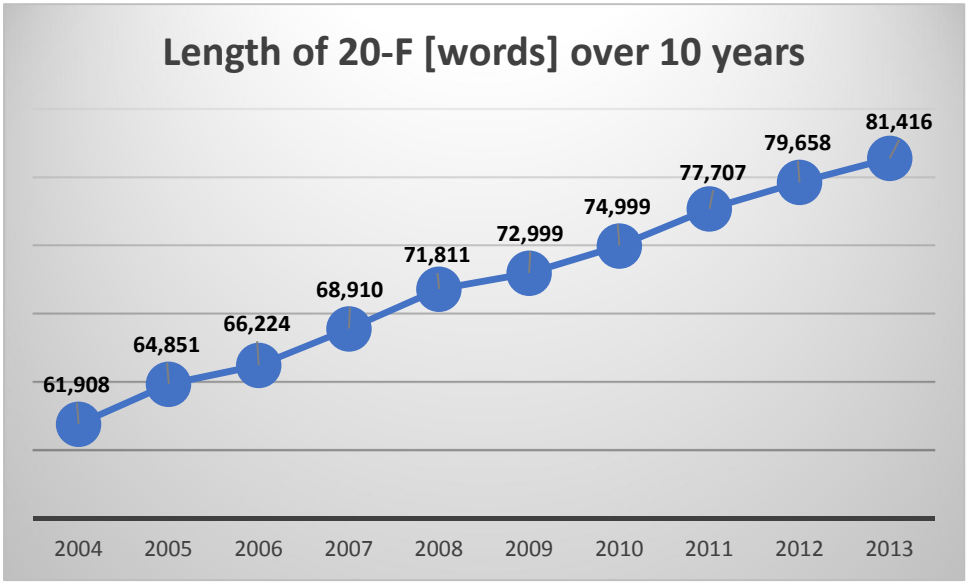
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Numwords (1)	1							
Size (2)	0.39	1						
Debt_Ratio (3)	0.20	0.03	1					
Loss (4)	-0.04	0.36	-0.06	1				
PTB (5)	-0.07	0.01	-0.01	-0.04	1			
Volatility (6)	-0.10	-0.54	0.03	0.02	-0.32	1		
IFRS (7)	0.29	0.5	0.04	0.01	0.20	-0.23	1	
English (8)	-0.26	-0.26	-0.15	-0.02	-0.14	0.08	-0.24	1

TABLE 2.4: Regression results

Independent variables	**Dependent variable**			
	(1)	**(2)**	**(3)**	**(4)**
Intercept	39,453***	84,037***	23,924***	63,512***
	(11.15)	(4.91)	(5.48)	(3.74)
Size	4,825***	4,270***	5,189***	4,613***
	(14.06)	(12.04)	(14.59)	(12.46)
Debt_Ratio	20,150***	19,712***	17,255***	16,355***
	(6.58)	(5.71)	(5.83)	(4.82)
Loss	-12,633***	-12,319***	-11,340***	-11,075***
	(-8.11)	(-8.03)	(-7.40)	(-7.33)
PTB	-180.47***	-172***	-149***	-141**
	(-3.78)	(-3.70)	(-3.18)	(-3.08)
Volatility	169,496***	150,284***	265,998***	242,761***
	(3.95)	(3.51)	(5.35)	(4.87)
IFRS	5,604***	4034**	6,253***	4,855**
	(3.99)	(2.65)	(4.52)	(3.23)
English	-8,143***	-9,068***	-7,183***	-8,009***
	(-6.06)	(-6.75)	(-5.50)	(-6.08)
Industry fixed effects	No	Yes	No	Yes
Year fixed effects	No	No	Yes	Yes
Adjusted R2	0.26	0.30	0.30	0.33
Observations	1,527	1,527	1,527	1,527

Figure 2.1: The length of reports from 2004 to 2013.

Figure 1.1 describes the length of 20-F over the 10 years. The length of annual reports is based on analyzing the text of the 20-F forms downloaded from EDGAR by using the Perl language.



Chapter 3: How Plain English Use Affects the Length of Annual Reports: Evidence from the Annual Reports of Foreign Firms listed on the US Stock Exchange

ABSTRACT

This chapter examines the impacts of plain English use on the length of annual reports for a sample of 20-F forms filed annually by foreign firms listed on NYSE and NASDAQ. We find that foreign firms adopt plain English in their disclosures over time with shorter sentences, less jargon, and more readable writing styles; however, the length of their 20-F forms significantly increases from 2004 to 2013. We also recognize that the impacts of plain English use on the length of annual reports are mixed. The length and plain English adherence level in 20-F forms show statistically significant differences between accounting standards as well as between English-speaking countries and other countries.

Keywords: 20-F, annual report length, plain English.

1. Introduction

Investors, researchers, regulators, and standard setters are paying attention to the recent significant increase in the length and complexity of annual reports, which contain more price-relevant information compared with other disclosures by firms (You and Zhang, 2009). The length of annual reports climbed by 16% for 10-K forms from 2004 to 2010 and by 32% for 20-F forms from 2004 to 2013 (KPMG, 2012; see also Chapter 2). Such a rapid increase prevents investors from accessing information at an acceptable cost and causes a series of deficiencies for the stock market. For instance, longer annual reports exacerbate the underreaction of stock markets, prevent small investors from trading, and increase dispersion and uncertainty (Lehavy and Merkley, 2011; Miller, 2010; You and Zhang, 2009). More seriously, an increase in the length of annual reports is accompanied by an increase in the repetition of disclosures (Cazier and Pfeiffer, 2016; Dyer et al., 2017; Li, 2017). Further, such repetition of disclosures becomes more serious over time (Li, 2017), leading to greater worries for regulators and standard setters since it increases the cost and time for investors when reading and analyzing information but provides no more detail. Moreover, repetition can distract investors and lead to them omitting important information (Ernst & Young, 2014; KPMG, 2012; SEC, 2003, 2014)[1].

To reduce the complexity of annual reports, the Securities and Exchange Commission (SEC) regulates firms to adopt plain English in their disclosures. The SEC also provides guidelines

[1] However, Li (2017) argues that repetition in annual reports is informative according to the succession hypothesis. Succession theory supposes that repetitive disclosures prevent investors from missing valuable information since it is mentioned in different sections of annual reports. Repeating information also enhances the understanding of investors when it is displayed in different ways. Hence, Li (2017) suggests that repetition improves the informativeness of annual reports.

for firms on how to use plain English in disclosures; however, it admits that adopting plain English can increase the length of annual reports (SEC, 1998a). In such a case, the increase in the length of annual reports due to the use of plain English does not disclose additional information to investors but can increase the time and cost for analysis, thereby reducing the effectiveness of such reports. By contrast, such an increase can improve the readability of annual reports thanks to using plain English in disclosures.

Our research uses a sample of annual reports issued by foreign firms listed on the US stock exchange (termed 20-F forms) to record the impacts of plain English use on the length of annual reports. To measure plain English in these 20-F forms, we use the Lingua::EN::Fathom package of the Perl language and StyleWriter Professional software. As the first research identifying the impacts of plain English on the length of annual reports, we contribute to practice and academic research in three main ways.

Firstly, we deeply understand how issuers adhere to the SEC's regulations on plain English. Overall, foreign firms better comply with the SEC's plain English guidelines in their disclosures over the period from 2004 to 2013. They write shorter sentences, use less jargon, and adopt friendlier writing styles that improve the readability of 20-F forms; however, this significantly increases the length of their annual reports. In other words, the greater length of annual reports issued by foreign firms is also due to plain English use. However, as mentioned above, such an increase does not offer additional information to investors.

Secondly, our research recognizes that the impacts of plain English use on the length of annual reports are mixed. For instance, writing short sentences and adopting friendly writing styles can increase the length of annual reports, whereas using the active voice and less jargon can reduce the length of annual reports. Hence, applying the SEC's regulations on plain English use has mixed effects on the length of annual reports. Our research provides empirical evidence of the types of plain English that lengthen annual reports for regulators. To reduce costs, standard setters should aim to review and modify the guidelines on plain English to improve the readability of annual reports but not significantly increase their length.

Finally, we find significant differences in how foreign firms in various countries disclose information. Firms in English-speaking countries use shorter sentences, less jargon, and less readable writing styles to shorten their annual reports compared with firms in other countries. Similarly, annual reports following different accounting standards show significant differences in length and plain English adherence level. Most importantly, language and accounting standards can affect the relationship between plain English use and the length of annual reports. In other words, the extent to which the elements of plain English affect the length of annual reports depends on the language and accounting standards used; Our findings thus explain the impacts of plain English on the length of annual reports under different accounting standards and in different languages.

The remainder of this article is organized as follows. The following section presents the literature review on the length of annual reports and research on the textual analysis of such reports. Section 3 presents the sample selection. Section 4 illustrates how to measure elements of plain English and the research design. Section 5 shows our findings and discussion. Section 6 provides additional tests. Finally, section 7 concludes.

2. Literature Review

To protect investors, the SEC requests that firms "disclose meaningful financial and other information to the public. This provides a common pool of knowledge for all investors to use to judge for themselves whether to buy, sell or hold a particular security" (SEC, 2013). This regulation encourages firms to disclose timely information to investors throughout the fiscal year. However, investors cannot enjoy all the benefits from firms' disclosures if it is too costly to extract and analyze useful information from extremely long and complex reports (Bloomfield, 2002). Surprisingly, most evidence from professional organizations and researchers shows a significant increase in the complexity of firms' disclosures, especially annual reports, which are strictly regulated by the SEC and the Financial Accounting Standards Board (FASB).

2.1 Negative impacts on the US stock market

All firms listed on the US stock market must file mandatory annual reports. US firms must file 10-K forms, whereas foreign firms that have less than 50% shares traded on the US stock exchange issue 20-F forms annually. Annual reports are considered to be the official information channel through which firms communicate with investors. In fact, they contain more price-relevant information than other disclosures by firms (You and Zhang, 2009). Given the crucial role of such reports, the SEC and FASB strictly regulate their format and content to make them easily understandable for average investors.

SEC (1969) forces firms to avoid unnecessarily long and complex reports and/or verbose writing. However, despite the regulations and concerns of the SEC, annual reports have become severely long over time (Li, 2008; see also Chapter 2). Miller (2010) reports that the average number of words in 10-K forms increased from 25,515 in 1995 to 40,579 in 2006. Similarly, there were 16% and 28% increases in the length of annual reports and in the notes of 10-K forms from 2004 to 2010, respectively (KPMG, 2012). More recently, the average number of words in 10-K forms from 2003 to 2012 was recorded as 55,335 (Cazier and Pfeiffer, 2016). A similar but more serious trend is recorded for the annual reports of foreign firms listed on the US stock market. As mentioned in chapter 2, we find a 32% rise in the length of 20-F forms from 2004 to 2013, equivalent to 19,508 additional words. Fair value, internal controls, and risk factor disclosures are the three sections recorded as most increasing the length of annual reports (Dyer et al., 2017).

The dramatic increase in the length of annual reports can be explained by operating complexity, redundancy between the SEC and FASB in disclosure requirements, and residual firm

disclosures (Cazier and Pfeiffer, 2016; Dyer et al., 2017). In addition, the rise in the number of words in annual reports results from firm performance and stock return volatility (Li, 2008; Lundholm et al., 2014; see also Chapter 2). Accounting standards and language also explain the rise in the length of the annual reports of foreign firms in the United States (see Chapter 2). Because the significant increase in the length of annual reports is more likely to be determined by managerial discretion than the complexity of firms, such an increase can cause obfuscation and cautiousness (Cazier and Pfeiffer, 2016; Li, 2008; Li 2017).

The increase in the length of annual reports is also accompanied by an increase in repetition. Li (2017) reports that the proportion of information repeated in notes and in MD&A rose from 19% in 1995 to over 26% in 2013. Such repetition in annual reports does not raise informativeness to investors and may even cause investors to overlook other material information (SEC, 2003, 2014). Similarly, KPMG (2012) and Ernst & Young (2014) argue that repetition in disclosures is redundant and unnecessary as it makes it difficult for investors to identify which information is important.

The increase in the length of annual reports has received criticism from users, standard setters, and researchers. Levitt (1997) was the first to voice his concern about the negative impacts of overloading information in the annual reports, stating that "more disclosure does not always mean better disclosure." Some research finds a greater underreaction of the stock market when annual reports become longer (Callen et al., 2010; Lee, 2012; You and Zhang, 2009). Longer annual reports take more time and are costlier for investors to understand since they must employ professional analysts or increase online searching, which prevents small investors from trading the stocks of those firms issuing complex annual reports (Guay et al., 2016; Lehavy et al., 2011; Lawrence, 2013; Miller, 2010). The increase in the length of annual reports also causes greater dispersion, lower accuracy, and greater uncertainty in analysts' earnings forecasts (Lehavy et al., 2011).

2.2 Textual characteristics in annual reports and effects on stock markets

To protect the average investor from misunderstanding or misleading information extracted from annual reports, the SEC strictly mandates firms to use plain English in communication with outside users. In detail, SEC (1998a) requires that companies filing registration statements under the Securities Act of 1993 must:

> "Write the forepart of these registration statements in plain English;
>
> Write the remaining portions of these registration statements in a clear, understandable manner; and
>
> Design these registration statements to be visually inviting and easy to read."

Under Rule 421(d), it also recommends that firms use "short sentences; definite and concrete everyday language; active voice; tabular presentation; no legal jargon; [and] no multiple

negatives" in their disclosures. It suggests avoiding "legal and highly technical business terminology" and "legalistic or complex presentations" (see Rule 421(b)). Further, firms' disclosures need to show "clarity, not brevity" and firms have to trade off writing clarity and disclosure quantity (SEC, 1998a).

The SEC's regulations on plain English have significantly affected the behaviors of firms. In particular, firms have improved the stylistic features of their filings to comply with the SEC's regulations such as shorter sentences, more personal pronouns, or less legal terminology. Plain English in annual reports is also affected by the effectiveness of corporate governance (Loughran and McDonald, 2014b). Similarly, an improvement in using boilerplate in the annual reports of 42 countries from 1998 to 2011 was recorded. The lower usage of boilerplate is explained by adopting IFRS and US-GAAP (Lang and Stice-Lawrence, 2015).

However, current research shows some negative viewpoints about the textual characteristics of annual reports. In 10-K forms from 1996 to 2013, Dyer et al. (2017) record a significant increase in length, stickiness, and redundancy as well as greater boilerplate text and a decrease in readability. In terms of the tone used in annual reports, Li (2010) points out the negative tone[2] in MD&A disclosures. Loughran and McDonald (2015) also state that the tones used in 10-K forms are mostly negative regardless of which dictionary[3] is used to measure.

These textual characteristics affect the stock market. Li (2008) argues that the linguistic features of annual reports correlate with firm performance. The frequency of causation words, positive emotion words, and future tense verbs is associated with earnings persistence. This result is confirmed by Li (2010), who applies machine learning to test the tone and content of 10-K forms. He recognizes that firms with good performance tend to use more positive forward-looking statements in their MD&A. Hence, the tone used in MD&A can help predict future performance. Additionally, the frequency of words has a correlation with stock returns and volume during filing days. Therefore, a positive tone is correlated with higher returns and higher volumes during window periods (De Franco et al., 2015; Henry and Leone, 2016; Loughran and McDonald, 2011). Identifying the impacts of the tone of annual reports on stock markets has attracted various researchers to introduce specific dictionaries. For example, Loughran and McDonald's dictionary and Henry's wordlist have replaced well-known dictionaries such as Diction and Harvard GI Word Lists.

3. Sample Selection

We use the annual reports issued by foreign firms listed on the US stock exchange (i.e., 20-F forms). Such foreign firms have different characteristics to US firms since they work in

[2] Negative tones mean using more pessimistic or uncertain words in annual reports.

[3] Loughran and McDonald (2014b) use Diction and the Loughran and McDonald dictionary to identify the tone of 10-K forms. Diction is the most popular dictionary used in the business context. Loughran and McDonald's dictionary is statistically identified by the frequency of words in the entire 10-K universe.

various business and legal environments; therefore, the annual reports of foreign firms listed on the US stock exchange can be longer, have more numbers, be more/less readable, and be affected by language and accounting standards (Lundholm et al., 2014; see also Chapter 2). We thus choose 20-F forms to explain the length–readability nexus in annual reports.

We use Python to auto-collect all the annual reports of foreign firms from EDGAR (Electronic Data Gathering, Analysis, and Retrieval) and obtain 7,588 filings from 2004 to 2013 (available on EDGAR website in the period from 2005 to 2014). We eliminate all filings from the finance and insurance industries and filings whose fiscal years end on dates other than 31 December. We include filings of firms listed on the NASDAQ and NYSE that have RIC codes available in Thomson Reuters Datastream. We also include filings in a text file that suits analysis using the Perl language. All filings with missing data are removed from our sample. Our final sample has 1,522 observations (Table 3.1).

[INSERT TABLE 3.1 HERE]

We follow Bonsall et al. (2017), Li (2008), and Loughran and McDonald (2014a, 2014b) to clean the raw text files downloaded from EDGAR. Appendix 3.1 presents the details of cleaning the 20-F forms. We use the Lingua::EN:: Fathom package of the Perl language to identify the length of annual reports, Fog Index, average words per sentence, and percentage of complex words. These measures are popular in previous research on the readability of annual reports (Lee, 2012; Lehavy et al., 2011; Lee, 2008; Lundholm et.al, 2014). Following the suggestion of Miller (2010) and Bonsall et al. (2017), we use StyleWriter Professional Edition to record the Bog Index, Passive Index, Sentence Bog, length of annual reports, Pep Index, and Jargon Index. However, we do not use all the variables in the Lingua::EN:: Fathom package and StyleWriter Professional Edition software for our analysis since some have the same results and are irrelevant to our research.

4. Research Design

Using plain English can increase the length of annual reports, as described earlier. Additionally, some research finds a negative correlation between the length of annual reports and Fog Index, which measures average words per sentence and the percentage of complex words (Cheung and Lau, 2016; Lundholm et al., 2014). This fact supports that the increase in the length of annual reports owing to the use of plain English (e.g., shorter sentences and less complex words) can improve the readability of such reports. In other words, using plain English can lower the Fog Index and thus can affect the length of annual reports. Moreover, some research supposes that the increase in the length of annual reports is not more informative for investors (Cazier and Pfeiffer, 2016; Dyer et al., 2017; KPMG, 2011; SEC, 2003). The above evidence allows us to suppose that using plain English in disclosures can explain the variation in the length of annual reports. If our suggestion is acceptable, we can question the opinions of the SEC about the rise in the length of annual reports.

Based on the elements of plain English suggested by the SEC, we use the specification in Equation (1) to identify the impacts of plain English on the length of 20-F forms:

$$\text{LN\_NUMWORDS} = \beta_0 + \beta_1 AVE\_SEN_{it} + \beta_2 PASSIVE_{it} + \beta_3 JARGON_{it} + \beta_4 PEP_{it} + \beta_5 LN\_SIZE_{it} + \beta_6 DEBT\_RATIO_{it} + \beta_7 PTB_{it} + \beta_8 LOSS_{it} + \beta_9 VOLATILITY_{it} \quad (1)$$

Overall, we measure the impacts of four elements of plain English, namely average words per sentence, the Passive Index, the Jargon Index, and the Pep Index, on the number of words in 20-F forms. We also include five control variables in our regression: firm size, the debt ratio, price-to-book value, loss, and volatility. Appendix 3.2 summarizes the definitions and expected signs of the variables.

4.1 Elements of plain English

The SEC lists a number of problems that firms should avoid such as using the passive voice, weak or hidden verbs, superfluous words, legal and financial jargon, numerous defined terms, abstract words, unnecessary details, lengthy sentences, and an unreadable design and layout (SEC, 1998b). Following the SEC's suggestion, we use thus four variables to measure plain English in 20-F forms (see Appendix 3.3 for further details).

Firstly, we use average words per sentence (AVER_SEN) to measure sentence length. Following the SEC's suggestion, firms should use shorter sentences in their annual reports; therefore, if average words per sentence are higher, the less plain English is applied. This variable is acquired from the Lingua::EN:: Fathom package of the Perl language[4]. As SEC (1998a) mentions: "writing a disclosure in plain English can sometimes increase the length of particular sections…", thus, shorter sentences are expected to lengthen annual reports.

Secondly, the Jargon Index of StyleWriter is calculated based on the use of abbreviations and acronyms, jargon, difficult words, unusual words, and unknown words. This index therefore measures some of the SEC's plain English problems relating to jargon such as legal and financial jargon, numerous defined terms, and abstract words. Compared with the percentage of complex words identified by the Perl language, the Jargon Index is superior since its percentage of complex words is based on syllables rather than difficult words. The latter index in the Fog Index has received criticism from researchers (Bonsall et al., 2017; Loughran, and McDonald, 2014a). The expected sign of the Jargon Index on the length of annual reports is controversial. First, a higher Jargon Index makes annual reports longer since firms need to provide more explanation with jargon. For instance, when annual reports contain unusual words or abbreviations such as SEK and

[4] StyleWriter software has Sentence Bog; however, this index is identified by squaring average sentence length and then scaling to a standard long sentence limit of 35 words per sentence. Compared with average words per sentence in the Perl package, Sentence Bog does not directly mention sentence length.

AED (currency terms), firms need to explain what they mean. Otherwise, using abbreviations can shorten reports since they replace long compound nouns with shorter words.

Thirdly, StyleWriter software counts the number of passive verbs and then divides them by the number of sentences and multiplies the result by 100 to generate the Passive Index. Passive sentences are always longer than active ones; therefore, the higher the Passive Index, the longer annual reports are. The SEC recommends that firms use the active voice for their disclosures; hence, the higher the Passive Index, the lower is the compliance with the SEC's regulations.

Finally, the Pep Index identified by StyleWriter software measures writing styles that make reports more readable and thus enjoyable for readers. It includes lively verbs, interesting adverbs or adjectives, interesting nouns, people's names, conversational words, personal pronouns, and direct questions. A higher Pep Index means using more plain English. StyleWriter software awards credit points for categories of the Pep Index. Appendix 3.3 presents further details of these credit points. Annual reports with a higher Pep Index are projected to be longer but more readable.

4.2 Other variables

LN_NUMWORDS: We take the natural logarithm of the number of words in 20-F forms acquired by the Lingua::EN:: Fathom package of the Perl language. This is the most popular method in previous research (De Franco et al., 2015; Lehavy et al., 2011; Li, 2008; You and Zhang, 2009)[5]. Therefore, we decide to use the data acquired by the Perl language.

LN_SIZE: Firm size affects the length of annual reports since larger firms carry out more business activities, conduct more complicated transactions, have more business segments, and need more disclosures (Cazier and Pfeiffer, 2016; Li, 2008). Size is identified by the natural logarithm of the market value of equity at the end of the fiscal year.

DEBT_RATIO: Information on the debts of firms is extremely sensitive to different users and it is necessary to disclose more details under regulations. We expect the annual reports of firms with a higher debt ratio to be longer. This variable represents total liabilities scaled by total assets in the same year (Cazier and Pfeiffer, 2016).

LOSS: This indicator variable is equal to 1 when firms have positive earnings and 0 otherwise. Firms with positive earnings are expected to publish shorter annual reports than firms with negative profits (Li, 2008).

PTB: This variable is identified by market value per share divided by book value per share at the end of the fiscal year. We expect that firms with a lower price to book value have longer annual reports to allow them to explain more details about their business activities or hide disadvantageous information (Li, 2008).

[5] StyleWriter software also provides the number of words in annual reports; however, this is not significantly different to the figures acquired by the Perl package (around 2%, not tabulated).

VOLATILITY: According to efficient market theory, stock prices reflect the information available to markets; therefore, firms with more information have volatile stock prices. If the stock prices of firms become more volatile, their annual reports are projected to become longer because of the presence of more information or more explanation. VOLATILITY thus represents the daily standard deviation of adjusted stock returns during a year. A positive association between the volatility of stock returns and length of annual reports is expected (Lundholm et al., 2014; see also Chapter 2).

5. Results

5.1 How plain English use changes over time

Figure 1 shows the changes in the length and readability[6] of annual reports as well as in the elements of plain English over the 10-year study period. Overall, the length of annual reports increases rapidly at an average of 32% from 2004 to 2013 (see also Chapter 2). Similarly, the Bog Index rises from 79.78 to 82.92 in these 10 years. Noticeably, the rate of change in the Bog Index between two continuous years slightly declines. By contrast, the Fog Index increases in the first seven years before slightly declining in the last three years. The improvement in the Fog Index is explained by the shortening of the sentences in the annual reports of foreign firms[7]. Indeed, the pattern of average words per sentence is the same as the pattern of the Fog Index. Average words per sentence is 21.68 in 2004, reaching 26.09 in 2010 before reducing to 24.29 in 2013.

[INSERT FIGURE 3.1 HERE]

In terms of jargon, foreign firms have better adapted to using plain English with a gradual decline in the Jargon Index in the last seven years, although there is a slight increase from 3.48 to 3.56 in the first three years. Panel F of Figure 1 also shows that the passive voice is more popular in the annual reports of foreign firms, as shown by the increase in the Passive Index identified by StyleWriter. The Pep Index is the only element of plain English that fluctuates over the 10 years. This increases slightly in the first four years to a peak of 10.12 in 2007, and then declines gradually to 9.97 in 2012, before it increases again to 10.05 in 2013.

Overall, there is an improvement in adopting plain English in the annual disclosures of foreign firms from 2004 to 2013 with shorter sentences, less jargon, and friendlier writing styles. This enhances the readability of annual reports regardless of the significant increase in their length. This result is consistent with the findings of Loughran and McDonald (2014b) and Lang and Stice-Lawrence (2015).

---

[6] We use the Fog Index and Bog Index to measure the readability of annual reports. The Fog Index, the most popular measurement of readability in previous research, is measured by average words per sentence and the percentage of complex words. The Bog Index, as a measurement using plain English to identify readability, is calculated by the Word Bog, Sentence Bog, and Pep Index. The Bog Index is used by Bonsall et al. (2017) and Miller (2010).

[7] The percentage of complex words in the Fog Index is stable over the 10 years at around 25%.

5.2 Descriptive analysis

Table 3.2 presents the descriptive statistics of all the variables in Equation (1). The natural logarithm of the number of words in a 20-F form is 11.12, equivalent to 73,097 words. The annual reports issued by foreign firms listed on the US stock exchange are far longer than the 10-K forms filed by US firms (Lundholm et al., 2014). The difference in the length of 20-F and 10-K forms is explained by the fact that foreign firms are located in different countries under different legal and business environments, meaning that they have to disclose more information and offer more explanation to enhance the comparability and reliability of their annual reports to US investors who underweight the equity of foreign firms (French and Poterba, 1991).

[INSERT TABLE 3.2 HERE]

The average size of the firms in our sample is more than 12,425 thousand dollars, equivalent to a natural logarithm of 7.07. The mean debt ratio is around 23%, which means that 23% of total assets are financed by debt. The mean of LOSS indicates that around 74% of the observations in our sample have positive earnings. The average price-to-book value and volatility of stock returns are 2.55 and 0.0314, respectively.

The number of words per sentence in 20-F forms is around 24.45 words, which is higher than the optimum length of 15–20 words per sentence (Cutts, 2013). The Passive Index is 35.41 over the 10-year period, while the Jargon Index and Pep Index are 3.47 and 9.98, respectively. Overall, the annual reports issued by foreign firms listed on the US stock exchange violate the suggestions of the SEC on plain English, as they are extremely long with long sentences, high passive voice use, and frequently used jargon. Therefore, 20-F forms are considered to be difficult for readers to understand, showing a high Fog Index of 19.86 and a high Bog Index of 81.93.

Table 3.3 presents the Pearson correlation between the variables in Equation (1). There is a significant correlation between LN_WORDS and the independent variables. The length of annual reports has a positive correlation with firm size, the debt ratio, the Passive Index, the Jargon Index, and the Pep Index, whereas it is negatively correlated with the price-to-book value, loss, volatility, and average words per sentence. Although other variables also show correlations, the correlation magnitude is not large.

[INSERT TABLE 3.3 HERE]

5.3 Plain English and the length of annual reports

Table 3.4 summarizes the regression results of Equation (1). The impacts of plain English on the length of the annual reports issued by foreign firms listed on the US stock exchange are displayed in Column (1) of Table 3.4 without the control variables. Column (2) shows the intercept with the coefficients of all the variables in Equation (1). Columns (3) and (4) present the regression results of Equation (1) with industry fixed effects and year fixed effects, respectively. Column (5) presents the results when both these fixed effects are included in the regression. Overall, the

coefficients and significance among the models are similar. The adjusted R-squared improves over the four models; however, the degrees of freedom decline when we include fixed effects in the regressions.

[INSERT TABLE 3.4 HERE]

As shown in Column (1) of Table 3.4, the length of annual reports is positively associated with the Passive Index, Jargon Index and Pep Index, whereas it is negatively correlated with average words per sentence. All the coefficients are significant at the 99% confidence level. The coefficient of the Jargon Index is the highest. The adjusted R-squared is over 7%, showing that the four elements of plain English explain 7% of the variation in the length of 20-F forms. However, the effects of plain English on the length of annual reports are still ambiguous. For 20-F forms, the Passive Index and Pep Index increase over time, but sentences become shorter and use slightly less jargon. Overall, three out of the four elements of plain English (more passive voice, a higher Pep Index, and shorter sentences) increase the number of words in 20-F forms. Indeed, the increase in the length of annual reports due to plain English use does not provide more information to investors.

When adding the control variables into the regression, the impacts of all four elements of plain English on the length of annual reports do not change. Annual reports become longer if firms use more jargon, more passive voice, friendlier writing styles, and shorter sentences in their annual reports, ceteris paribus. All the coefficients are still statistically significant at the 1% level. Interestingly, the adjusted R-squared substantially improves to 20% when adding the control variables into the regression. The significant improvement in the adjusted R-squared demonstrates that the explanatory power of our models is enhanced.

As shown in Columns (3)–(5), the regression results do not change significantly when including industry fixed effects and/or year fixed effects; however, the significance level of the price-to-book value (PTB) decreases from 99% to 95% when we add year fixed effects into the models. The adjusted R-squared of our models ranges from 20% to 27%, showing that our independent variables explain 20–27% of the change in the dependent variable. This supposes that plain English use makes annual reports longer and thus more readable if not more informative.

Table 3.4 shows that the length of annual reports is positively associated with operating complexity measured by firm size and the debt ratio of firms. The coefficients of size and the debt ratio are 0.06 with a t-value of 12.48 and 0.3765 with a t-statistic of 7.893, respectively. The positive coefficients of size and the debt ratio suppose that larger firms and/or firms with a higher debt ratio issue longer annual reports than other firms, ceteris paribus. By contrast, the number of words in 20-F forms declines when firms perform badly. This result is consistent with the results of Aymen et al. (2016) and Li (2008). The negative coefficients of LOSS and PTB assume that firms making a loss and those with a lower price-to-book value provide longer reports since they want to obfuscate adverse information from investors. The regression result also supposes that as the daily stock returns of firms are more volatile, firms disclose more information in their annual

reports, which leads to longer reports. The coefficient of VOLATILITY is 2.914 with a t-statistic of 4.329. In other words, the difference in the length of annual reports results from the variation in the volatility of stock returns.

Table 3.5 also shows the regression results of testing the impacts of plain English use on the length of annual reports; however, we use the number of words acquired by Lingua::EN:: Fathom rather than the natural logarithm of the number of words in the 20-F forms. Column (1) shows the impacts of the four elements of plain English on the number of words in annual reports, Column (2) includes the control variables in the regression, and Columns (3)–(5) display the results with fixed effects.

[INSERT TABLE 3.5 HERE]

As shown in Table 3.5, all the coefficients are extremely high since the dependent variable changes. Interestingly, the positive and negative associations between the variables remain compared with the results in Table 3.4. However, the intercept becomes insignificant when adding industry fixed effects into the regressions. The significance levels of average words per sentence and the price-to-book value reduce from 99% to 95% when we change the dependent variable. Replacing the natural logarithm of the number of words by the number of words in the 20-F forms enhances the adjusted R-squared to over 30%. In conclusion, the sensitivity test shows the same results for the impact of plain English use on the length of annual reports.

Overall, the empirical results support our expected signs. Although plain English use affects the length of annual reports, the patterns of the elements of plain English are mixed. As a consequence, the change in the length of annual reports due to the use of plain English or repetition does not improve the informativeness of annual reports but rather increases the time and cost of users (Cazier and Pfeiffer, 2016; KPMG, 2012; SEC, 2003, 2014). However, better compliance with plain English guidelines can improve the readability of annual reports and enhance the understanding of investors. Improvement in the readability of annual reports is expected to have some positive effects on the stock market. Standard setters should therefore consider some of the elements of plain English, especially those raising the length of annual reports. That said, researchers should be cautions when using the length of annual reports to measure their readability. Longer annual reports do not mean they are less readable (Bonsall et al., 2017; Cheung and Lau, 2016; Loughran and McDonald, 2014a).

6. Additional Tests

6.1 Plain English and language

Our sample comprises annual reports filed by foreign firms listed on the US stock exchange. Those firms are located in various countries around the world, some of which the native language is English, while other firms are situated in non-English-speaking countries. The SEC requires all foreign firms to file their annual reports in English. We suppose that language has an effect on the

length of annual reports and on the four elements of plain English. Based on the declaration of firms in their 20-F filings, we identify their locations to understand the important role of language in plain English use and the length of annual reports. Following the classification of Lundholm et al. (2014), we divide our observations into filings of foreign firms located in English-speaking countries and filings of firms in other countries. Appendix 3.4 provides further details of our classification.

[INSERT TABLE 3.6 HERE]

As shown in Table 3.6, firms in countries speaking English file shorter annual reports on average (64,235) words than the annual reports of firms in non-English-speaking countries (79,305 words). Our finding is consistent with that of Lundholm et al. (2014), who show that English plays an important role in deciding the length and readability of annual reports, especially for a sample of foreign firms. Annual reports filed by native English speakers contain longer sentences (25.43 vs. 23.77), use more passive voice (36.70 vs. 34.56), and have less readable styles (9.84 vs. 10.08) than the reports of firms in non-English-speaking countries. Those elements help the 20-F forms of foreign firms in English-speaking countries become shorter; however, they are less friendly to readers since these reports ignore the SEC's guidelines on plain English. Foreign firms in non-English-speaking countries use more jargon than firms in English-speaking countries. Non-English-speaking countries contain a lot of dialects and unusual words for which users need an explanation. All the differences in the length of annual reports and elements of plain English are significant at the 99% confidence level for both the t-test and the Wilcoxon test.

Given these statistically significant differences between the filings issued by native and non-native English speakers, we suppose that interactions between language and the elements of plain English exist. We therefore retest the effects of plain English on the length of annual reports by adding language interactions into Equation (1):

$$\text{LN\_NUMWORDS} = \beta_0 + \beta_1 AVE\_SEN_{it} + \beta_2 PASSIVE_{it} + \beta_3 JARGON_{it} + \beta_4 PEP_{it} + \beta_5 LN\_SIZE_{it} + \beta_6 DEBT\_RATIO_{it} + \beta_7 PTB_{it} + \beta_8 \text{LOSS}_{it} + \beta_9 VOLATILITY_{it} + \beta_{10} ENGLISH_{it} + \beta_{11} ENGLISH_{it} * AVE\_SEN_{it} + \beta_{12} ENGLISH_{it} * JARGON_{it} + \beta_{13} ENGLISH_{it} * PEP_{it} + \beta_{14} ENGLISH_{it} * PASSIVE_{it} \quad (2)$$

Table 3.7 presents the regression results of Equation (2). Interestingly, the coefficient of average words per sentence becomes insignificant in countries speaking other languages; however, the average length of sentences negatively affects the number of words in the 20-F forms issued by English speakers at the 99% confidence level. This finding means that the impact of average words per sentence on the length of annual reports depends on the language of issuers. Similarly, there is a different effect of the Pep Index on the length of annual reports when the native languages of firms are different. The coefficient of the Pep Index is 0.0578 when firms are located in non-English-speaking countries, showing the significant positive association between friendlier writing styles and the length of annual reports in non-English-speaking countries. Nevertheless, the F-test

for the sum of the coefficients of the Pep Index and PEP*English is significant[8], showing that the Pep Index has no impact on the number of words in the 20-F forms issued by native English speakers. Jargon shows the opposite pattern. The significant positive association between the use of jargon and length of annual reports does not depend on the language of issuers. The coefficients of the Passive Index and Passive*English are positive and significant at 90% and 95%, respectively, which shows that the Passive Index has a stronger impact on the length of annual reports in English-speaking countries than in non-English-speaking countries.

[INSERT TABLE 3.7 HERE]

Columns (2)–(4) of Table 3.7 show the regression results of Equation (2) when adding both interaction terms and fixed effects. In general, the impacts of the elements of plain English on the length of annual reports are the same as the results without adding fixed effects, except the coefficient of the Passive Index becomes insignificant when adding year fixed effects. The insignificance of the Passive Index's coefficient means that using the passive voice does not impact on the length of annual reports in non-English-speaking countries when adding year fixed effects; however, the Passive Index still affects the number of words in the 20-F forms issued by native English speakers.

Overall, language thus plays a significant role in the variation of the length of annual reports as well as the plain English applied in them. The impacts of average words per sentence, the Passive Index, and the Pep Index on the length of annual reports depend on the language of issuers, while there is no difference in the impacts of jargon on the length of annual reports between firms in English-speaking countries and firms in non-English-speaking countries. As a consequence, the language of issuers can change the impacts of the elements of plain English on the length of annual reports, suggesting that standard setters should include language in their guidelines on plain English.

6.2 Plain English and accounting standards

Foreign firms listed on the US stock exchange operate in different geographic locations and thus follow different accounting standards. The SEC allows foreign firms to choose IFRS or US-GAAP for their annual reports provided that firms declare which accounting standard is applied at the beginning of the 20-F form. Based on firms' announcements, we sort our filings into two groups, namely IFRS and US-GAAP.

Global accounting systems have converged in recent years to improve the comparability and reliability of annual reports. Since 2007, the SEC has recognized that US-GAAP and IFRS are sufficiently harmonious, thereby allowing foreign firms to eliminate a reconciliation to US-GAAP if foreign firms adhere to IFRS. Kim et al. (2012) find that the SEC's elimination does not cause

[8] The F-probability is 0.836.

negative impacts on stock markets or users. Firms adhering to IFRS do not disclose more or more frequently.

Table 3.8 shows that annual reports following IFRS are far longer than annual reports under US-GAAP (81,250 vs. 64,905 words). Some material differences between US-GAAP and IFRS demand an additional explanation from firms and this leads to the increase in the number of words in 20-F forms under IFRS (PwC, 2016). Additionally, IFRS is principle-based accounting and has a broad guideline, while US-GAAP is rule-based with a list of detailed rules. Principle-based accounting requires more disclosures than rule-based accounting, which lengthens the annual reports adhering to IFRS compared with annual reports complying with US-GAAP. Our result is thus consistent with the results of Lang and Stice-Lawrence (2015).

[INSERT TABLE 3.8 HERE]

Table 3.8 also shows the differences in the four elements of plain English between these two accounting standard groups. The Wilcoxon test and t-test results show that all these differences are statistically significant at the 1% level. Firms in the IFRS group write shorter sentences and use more readable writing styles but more aggressively abuse the passive voice and jargon in their annual reports compared with firms in the US-GAAP group. Indeed, more than 70% of the firms in the IFRS group are located in non-English-speaking countries, implying that firms tend to use conversational language with short sentences, direct questions, and simple sentence structures. More importantly, all four elements of plain English in the IFRS group contribute to increasing the length of annual reports. Consequently, the increase in the length of annual reports following IFRS somehow results from the plain English used in those reports, which do not contain more information.

We run Equation (3) to test the impacts of plain English on the length of annual reports under the interaction of IFRS:

$$\text{LN\_NUMWORDS} = \beta_0 + \beta_1 AVE\_SEN_{it} + \beta_2 PASSIVE_{it} + \beta_3 JARGON_{it} + \beta_4 PEP_{it} + \beta_5 LN\_SIZE_{it} + \beta_6 DEBT\_RATIO_{it} + \beta_7 PTB_{it} + \beta_8 LOSS_{it} + \beta_9 VOLATILITY_{it} + \beta_{10} IFRS_{it} + \beta_{11} IFRS_{it} * AVE\_SEN_{it} + \beta_{12} IFRS_{it} * JARGON_{it} + \beta_{13} IFRS_{it} * PEP_{it} + \beta_{14} IFRS_{it} * PASSIVE_{it} \quad (3)$$

Column (1) of Table 3.9 shows the results of the regression without fixed effects. Columns (2)–(4) present the coefficients and t-values of Equation (3) with fixed effects. Overall, the effects of plain English on the length of annual reports remain significant at the 95% and 99% confidence levels when adding the interaction terms; however, those effects have different patterns for filings following IFRS and filings adopting US-GAAP.

[INSERT TABLE 3.9 HERE]

Hence, regardless of the accounting standards adopted, the significantly positive impacts of the Pep Index and Passive Index on the length of annual reports remain. In other words, the insignificance of the coefficients of the interaction terms shows that the impacts of the Passive

Index and Pep Index on the length of annual reports do not depend on the prevailing accounting standards. By contrast, the effects of the Jargon Index on the length of annual reports depends on the adopted accounting standards. For filings following US-GAAP, the coefficient of the Jargon Index is higher than that for the group of filings adopting IFRS; however, both coefficients are positive. To summarize, the positive effect of jargon use on the length of annual reports is stronger for filings under US-GAAP than those under IFRS. The F-test[9] for the sum of the Jargon Index's coefficient and the Jargon–IFRS's coefficient is statistically significant at the 99% confidence level. Most importantly, average words per sentence have different effects on the length of annual reports under these two accounting standards. The number of words per sentence has a positive impact on the length of annual reports that follow US-GAAP. Hence, longer sentences lead to longer reports. Conversely, the length of the 20-F forms under IFRS is negatively associated with the number of words per sentence[10]. In IFRS filings, the shorter the sentences, the longer the reports are. Surprisingly, the coefficient of IFRS is significant at the 99% confidence level when adding this variable into the regression; hence, accounting standards determine the variation in the length of annual reports (see also Chapter 2).

When adding fixed effects into the regression, the effect of average words per sentence on the length of annual reports that adhere to US-GAAP changes the significance level. Average words per sentence become insignificant when adding year fixed effects into the model; however, they are still highly significant when adding industry fixed effects. When both year and industry fixed effects are included in the regression, the impact of the number of words per sentence on the length of annual report becomes significant at the 90% confidence level. The other results remain when adding fixed effects into the model; however, the magnitude of the coefficients changes insignificantly.

In summary, the prevailing accounting standards serve as a determinant of the length of annual reports. The impacts of average words per sentence and jargon on the length of annual reports depend on the accounting standards that filings follow, while the Passive Index and Pep Index have no interactions with accounting standards.

7. Conclusion

Under efficient market theory, firms strive to distribute more timely and reliable information to investors by issuing more disclosures, among which annual reports are the primary information source for investors, regulators, and other users. Since 1998, the SEC has strictly mandated the format and content of those reports. Firms are strictly forced to meet all mandatory regulations and guidelines when issuing these important reports. The SEC requests firms to write their reports in plain English and in a clear and understandable manner; however, this requirement does influence the length of annual reports. Most previous research shows the significant increase

[9] The F-probability is 0.0011.
[10] The F-probability for the sum of the coefficients of AVER_SEN and AVER_SEN:IFRS at zero is -2.2e-16.

in the length of annual reports over time. Such an increase has attracted the attention of regulators, standard setters, and researchers since it causes deficiencies in the stock market and requires more time and costs to read and analyze the information. On the contrary, the increase in the length of annual reports does not improve informativeness since the repetition in those reports rapidly climbs. Our findings support the view that increasing the length of annual reports does not provide more information to investors since we are the first researchers to show that the SEC's regulations on plain English have different impacts on the length of annual reports. In other words, the increase in the length of annual reports that may result from compliance with the SEC's plain English suggestion does not provide more information to investors; however, investors must spend more time and cost reading such reports.

Our research also shows significant differences in the length and plain English adherence level between IFRS and US-GAAP filings. Language also plays an important role in deciding the length and writing styles of annual reports. Firms in English-speaking countries seem to adopt plain English to a lesser extent than firms in other countries. Interestingly, lower compliance to plain English use can shorten the annual reports of these firms. In addition, language and accounting standards affect the association between the elements of plain English and length of annual reports.

Despite our interesting findings, our research has open issues that need to be investigated more in depth in the future. Although we show that the length of annual reports is impacted by plain English, we do not test whether those impacts cause stock market reactions. Indeed, those impacts on the stock market are still ambiguous because they do not provide more information yet still affect the readability of annual reports, which improves investors' understanding. The readability of annual reports thus influences the stock market. We leave more investigation into this to future research.

APPENDICES

APPENDIX 3.1: CLEANING THE RAW TEXT FILES OF THE 20-F FORMS

To clean the raw text files downloaded from EDGAR, we follow the instructions of Bonsall *et al.* (2017), Li (2008), and Loughran and McDonald (2014a, 2014b). We use the packages File::Slurp, HTML::format text, and HTML::TreeBuilder; HTML::Entities;Text::Unidecode to clean the raw 20-F files, following these steps:

a. Remove the format design from the raw files.

b. Remove all content between the <XBRL> and </XBRL> tags.

c. Remove all tables with more than 15% numeric characters.

d. Remove all markup tags.

e. Remove other textual expressions such as newline and underscored characters.

APPENDIX 3.2: VARIABLE DEFINITIONS AND EXPECTED SIGNS

Variable Name	Expected Sign	Definition
LN_NUMWORDS	N/A	Natural logarithm of the number of words acquired by the Lingua::EN:: Fathom package of the Perl language for 20-F forms after cleaning.
AVE_SEN	(-)	Average words per sentence identified by the Lingua::EN:: Fathom package of the Perl language for 20-F forms after cleaning.
PASSIVE	(+)	Passive Index acquired by StyleWriter Professional Edition.
JARGON	(+/-)	Jargon Index acquired by StyleWriter Professional Edition.
PEP	(+)	Pep Index acquired by StyleWriter Professional Edition.
LN_SIZE	(+)	Natural logarithm of the market value of equity at the end of the fiscal year.
DEBT_RATIO	(+)	Total liability scaled by total assets.
PTB	(-)	Market price to book value per share.
LOSS	(-)	An indicator variable equal to 1 if firms have positive net income and 0 otherwise.
VOLATILITY	(+)	Daily standard deviation of stock returns during a year.
ENGLISH	N/A	An indicator variable equal to 1 if firms are located in English-speaking countries and 0 otherwise.
IFRS	N/A	An indicator variable equal to 1 if firms' annual reports adopt IFRS and 0 if firms' annual reports comply with US-GAAP.

APPENDIX 3.3: PLAIN ENGLISH WRITING PROBLEMS AND THE PEP INDEX

The SEC's Plain English Problems	Variables Measuring Plain English
Passive voice	Passive Index
Weak/hidden verbs	N/A
Superfluous words	N/A
Legal and financial jargon	Jargon Index
Numerous defined terms	Jargon Index
Abstract words	Jargon Index
Unnecessary details	N/A
Long sentences	Average words per sentence
Unreadable design and layout	Pep Index

PEP INDEX

The Pep Index measures elements that make reports more readable and enjoyable for readers. It is defined and acquired by StyleWriter Professional Edition. The Pep Index has credit points for different forms of Pep:

Elements of Pep	Credit Point
Name of people and place	0.1
Interesting adverbs, adjectives, nouns, and verbs	0.2
Contractions, personal pronouns, conversational words, and phrasal verbs	0.5
Direct questions	1.0

APPENDIX 3.4: LIST OF ENGLISH-SPEAKING AND NON-ENGLISH-SPEAKING COUNTRIES

Based on the declaration in the first part of the 20-F forms, the executive locations of firms are identified. The classification of Lundholm et al. (2014) is used to sort the filings of foreign firms into the group of firms in English-speaking countries and the group of firms in non-English-speaking countries.

Country	English speaking	Country	English speaking
Argentina	0	Japan	0
Australia	1	Korea	0
Bahamas	1	Luxembourg	0
Belgium	0	Mexico	0
Brazil	0	Monaco	0
Canada	1	Netherlands	0
Chile	0	Norway	0
Columbia	0	Panama	0
Denmark	0	Papua New Guinea	1
Finland	0	Peru	0
France	0	Portugal	0
Germany	0	Singapore	1
Greece	0	South Africa	1
Hong Kong	1	Switzerland	0
Indonesia	0	Taiwan	0
Ireland	1	United Kingdom	1
Israel	1	Uruguay	0
Italy	0		

TABLE 3.1: Sample selection

We use Python to collect all the 20-F filings from 2004 to 2013 from EDGAR's website. We remove from our sample all filings in the finance and insurance sectors and filings with a fiscal year not on December 31 . We only choose filings of firms that have RIC codes (i.e., .N or .O) available in Thomson Reuters Datastream and filings that have no missing data.

Total observations collected from EDGAR	7,588
Number of firms	1,475
Number of firms whose fiscal years end on December 31	966
Number of firms with RIC codes available in Thomson Reuters	264
Number of observations without missing data	**1,522**
In which:	
Agriculture, forestry, and fishing	2
Mining	118
Manufacturing	758
Transportation, communication, electric, gas, and sanitary services	385
Wholesale trade	23
Retail trade	13
Services	223

TABLE 3.2: Descriptive statistics

Table 3.2 presents the descriptive analysis of all the variables in Equation (1). The number of words in annual reports and average words per sentence are acquired from the Lingua::EN:: Fathom package of the Perl language. The Passive Index, Jargon Index, and Pep Index are identified by StyleWriter Professional Edition. Size, volatility, price-to-book value, loss, and the debt ratio are extracted from Thomson Reuters Datastream based on firms' RIC codes.

	N	**Mean**	**Std. Dev.**	**Q1**	**Median**	**Q3**
LN_WORDS	1,522	11.12	0.42	10.92	11.14	11.35
LN_SIZE	1,522	7.07	2.53	5.2	7.0	9.13
DEBT_RATIO	1,522	0.23	0.21	0.04	0.19	0.35
LOSS	1,522	0.74	0.44	0	1	1
PTB	1,522	2.55	13.09	1	1.66	2.62
VOLATILITY	1,522	0.0314	0.02	0.0192	0.0269	0.0393
AVE_SEN	1,522	24.44	5.03	21.20	23.49	26.64
PASSIVE	1,522	35.41	4.78	32	36	38
JARGON	1,522	3.47	0.88	2.8	3.3	4.0
PEP	1,522	9.98	0.95	9.0	10.0	10.0
FOG INDEX	1,522	19.86	2.05	18.56	19.53	20.81
BOG INDEX	1,522	81.93	5.82	78	82	86

TABLE 3.3: Pearson correlations between variables

This table shows the Pearson correlations between the variables in Equation (1) including the natural logarithm of the number of words in 20-F forms, natural logarithm of size, debt ratio, volatility, price-to-book value, loss, average words per sentence, Passive Index, Jargon Index, and Pep Index.

	(1)	***(2)***	***(3)***	***(4)***	***(5)***	***(6)***	***(7)***	***(8)***	***(9)***	***(10)***
LN_WORDS (1)	1									
LN_SIZE (2)	0.39	1								
DEBT_RATIO (3)	0.20	0.03	1							
PTB (4)	-0.07	0.004	-0.01	1						
LOSS (5)	-0.04	0.36	-0.06	-0.04	1					
VOLATILITY (6)	-0.10	-0.54	0.03	0.02	-0.32	1				
AVE_SEN (7)	-0.04	-0.06	-0.03	-0.03	-0.03	0.13	1			
PASSIVE (8)	0.06	-0.07	0.02	-0.06	-0.07	-0.01	0.19	1		
JARGON (9)	0.17	0.17	-0.10	0.002	0.005	-0.02	-0.004	-0.28	1	
PEP (10)	0.09	0.02	0.10	-0.02	0.03	0.02	0.005	-0.41	0.003	1

TABLE 3.4: Regression results

This table reports the regression results for the impacts of plain English use on the length of annual reports issued by foreign firms listed on the US stock exchange. We use the natural logarithm of the number of words as the dependent variable. The regressions include 1,522 observations during 2004 to 2013. The regressions in Columns (3)–(5) add year fixed effects and/or firm fixed effects. All OLS regressions show the intercept, coefficients, and t-statistics.

Independent variables	**Dependent variable (LN_WORDS)**				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Intercept	9.669***	9.414***	9.422***	10.14***	10.10***
	(49.679)	(50.51)	(50.580)	(31.146)	(31.493)
AVER_SEN	-0.011***	-0.010***	-0.012***	-0.010***	-0.012***
	(-5.089)	(-5.02)	(-5.966)	(-5.065)	(-6.119)
PASSIVE	0.018***	0.016***	0.014***	0.014***	0.012***
	(6.959)	(6.82)	(5.836)	(6.092)	(4.960)
JARGON	0.113***	0.091***	0.089***	0.109***	0.107***
	(9.001)	(7.66)	(7.623)	(9.192)	(9.22)
PEP	0.069***	0.055***	0.048***	0.046***	0.038***
	(5.662)	(4.83)	(4.222)	(4.086)	(3.421)
LN_SIZE		0.060***	0.064***	0.050***	0.053***
		(12.48)	(12.847)	(9.747)	(9.921)
DEBT_RATIO		0.375***	0.339***	0.318***	0.269***
		(7.893)	(7.217)	(5.912)	(5.056)
PTB		-0.003***	-0.002**	-0.002***	-0.002**
		(-3.46)	(-2.95)	(-3.35)	(-2.809)
LOSS		-0.121***	-0.107***	-0.120***	-0.105***
		(-5.001)	(-4.46)	(-5.048)	(-4.461)
VOLATILITY		2.915***	3.964***	2.797***	3.765***
		(4.329)	(5.10)	(4.186)	(4.855)
Industry fixed effects	No	No	No	Yes	Yes
Year fixed effects	No	No	Yes	No	Yes
Adjusted R2	0.072	0.2037	0.2331	0.2433	0.2737
Observations	1,522	1,522	1,522	1,522	1,522

TABLE 3.5: Sensitivity results

This table reports the regression results for the impacts of plain English use on the length of annual reports issued by foreign firms listed on the US stock exchange. We use the number of words as the dependent variable. The regressions include 1,522 observations during 2004 to 2013. The regressions in Columns (3)–(5) add year fixed effects and/or firm fixed effects. All OLS regressions show the intercept, coefficients, and t-statistics.

Independent variables	**Dependent variable (NUM_WORDS)**				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Intercept	-38984**	-64158***	-63766***	-5731	-9437.2
	(-2.98)	(-5.393)	(-5.37)	(-0.276)	(-0.46)
AVER_SEN	-439.9**	-338.5**	-446.3***	-343**	-459.3***
	(-3.08)	(-2.66)	(-3.5)	(-2.74)	(-3.65)
PASSIVE	1227.4***	1145.9***	955.4***	987.9***	799.0***
	(7.1)	(7.49)	(6.29)	(6.51)	(5.32)
JARGON	7371.2***	5082.8***	4948.7***	6383.9***	6223.8***
	(8.77)	(6.7)	(6.66)	(8.45)	(8.42)
PEP	5379***	4462.3***	3964.5***	3770.7***	3255.3***
	(6.55)	(6.12)	(5.53)	(5.27)	(4.63)
LN_SIZE		5580.6***	5963.3***	4744***	5077***
		(18.13)	(18.71)	(14.44)	(14.85)
DEBT_RATIO		23075***	20329.2***	21524***	18004***
		(7.59)	(6.80)	(6.26)	(5.32)
PTB		-144.6**	-120.6**	-139.1***	-113.4*
		(-3.05)	(-2.58)	(-3.01)	(-2.50)
LOSS		-10595***	-9486.3***	-100.98***	-9080***
		(-6.86)	(-6.21)	(-6.65)	(-6.07)
VOLATILITY		200832***	300318***	179183***	270074***
		(4.68)	(6.06)	(4.2)	(5.47)
Industry fixed effects	No	No	No	Yes	Yes
Year fixed effects	No	No	Yes	No	Yes
Adjusted R2	0.070	0.2775	0.3076	0.3134	0.3450
Observations	1,522	1,522	1,522	1,522	1,522

TABLE 3.6. Plain English and the length of annual reports issued by firms in English-speaking and non-English-speaking countries

Based on the declaration of firms, we identify the executive locations. We follow Lundholm *et al*. (2014) to sort filings into those issued by firms in English-speaking countries and those issued by firms in non-English-speaking countries. The length of 20-F forms, words per sentence, Passive Index, Jargon Index, and Pep Index are acquired by the Perl language and StyleWriter Professional Edition.

	N	*Length of 20-F*		*Words per sentence*		*Passive Index*		*Jargon Index*		*Pep Index*	
		Mean	*Std. Dev.*	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev*
Non_English	*893*	*79,305*	*27,324*	*23.77*	*4.57*	*34.56*	*4.57*	*3.62*	*0.93*	*10.08*	*1.00*
English	*629*	*64,235*	*27,604*	*25.43*	*5.47*	*36.7*	*4.84*	*3.25*	*0.74*	*9.84*	*0.85*
Difference		*15,070*		*-1.66*		*-2.14*		*0.37*		*0.24*	
t-test		*<2.2e-16****		*3.591e-10****		*<2.2e-16****		*<2.2e-16****		*5.327e-07****	
U-test		*<2.2e-16****		*6.148e-08****		*<2.2e-16****		*2.343e-14****		*2.841e-06***	

*Note: *** denotes the 1% significance level; ** denotes the 5% significance level.*

TABLE 3.7: Regression results with interactions

This table reports the regression results for the impacts of plain English use on the length of annual reports issued by foreign firms listed on the US stock exchange by adding the interactions of language. All OLS regressions show the intercept, coefficients, and t-statistics.

LN_NUMWORDS = $\beta_0 + \beta_1 AVE\_SEN_{it} + \beta_2 PASSIVE_{it} + \beta_3 JARGON_{it} + \beta_4 PEP_{it} + \beta_5 LN\_SIZE_{it} + \beta_6 DEBT\_RATIO_{it} + \beta_7 PTB_{it} + \beta_8 LOSS_{it} + \beta_9 VOLATILITY_{it} + \beta_{10} ENGLISH_{it} + \beta_{11} ENGLISH_{it} * AVE\_SEN_{it} + \beta_{12} ENGLISH_{it} * JARGON_{it} + \beta_{13} ENGLISH_{it} * PEP_{it} + \beta_{14} ENGLISH_{it} * PASSIVE_{it}$

Independent Variables	**Dependent Variable (LN_WORDS)**			
	(1)	**(2)**	**(3)**	**(4)**
Intercept	9.6954***	10.279***	9.6613***	10.192***
	(41.461)	(29.707)	(41.786)	(29.985)
AVER_SEN	0.004	0.0017	0.00356	0.0013
	(1.44)	(0.648)	(1.351)	(0.482)
PASSIVE	0.005*	0.0051*	0.00228	0.0017
	(1.74)	(1.648)	(0.734)	(0.552)
JARGON	0.053***	0.0707***	0.0485***	0.0675***
	(3.815)	(5.096)	(3.583)	(4.965)
PEP	0.0578***	0.0516***	0.0522***	0.0462***
	(4.29)	(3.875)	(3.958)	(3.546)
LN_SIZE	0.056***	0.0468***	0.0620***	0.0518***
	(12.034)	(9.291)	(12.834)	(9.885)
DEBT_RATIO	0.338***	0.3265***	0.3032***	0.2754***
	(7.326)	(6.327)	(6.707)	(5.425)
PTB	-0.0023***	-0.0023***	-0.002***	-0.0019***
	(-3.421)	(-3.382)	(-2.873)	(-2.808)
LOSS	-0.1435***	-0.1411***	-0.1262***	-0.1234***
	(-6.251)	(-6.224)	(-5.581)	(-5.521)
VOLATILITY	2.588***	2.3599***	3.876***	3.5533***
	(4.050)	(3.685)	(5.265)	(4.788)
ENGLISH	-0.14	-0.1272	-0.0344	0.0026
	(-0.402)	(-0.364)	(-0.100)	(0.008)
Aver_sen:English	-0.0312***	-0.0261***	-0.0351***	-0.0300***
	(-8.100)	(-6.717)	(-9.217)	(-7.821)
Passive:English	0.0337***	0.0305***	0.0349***	0.0316***
	(7.214)	(6.601)	(7.630)	(6.995)
Jargon:English	0.0205	0.018	0.0296	0.0243
	(0.857)	(0.759)	(1.261)	(1.049)
Pep:English	-0.0536**	-0.0553**	-0.0602***	-0.0626***
	(-2.370)	(-2.482)	(-2.722)	(-2.871)
Industry fixed effects	No	Yes	No	Yes
Year fixed effects	No	No	Yes	Yes
Adjusted R2	0.2929	0.3184	0.3255	0.35

TABLE 3.8. Plain English and the length of annual reports adopting IFRS and US-GAAP

Based on firms' declaration in their 20-F reports, we sort filings into IFRS and US-GAAP. We use the packages of the Perl language and StyleWriter to identify the number of words and elements of plain English.

	N	*Length of 20-F*		*Words per sentence*		*Passive Index*		*Jargon Index*		*Pep Index*	
		Mean	*Std. Dev.*	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev.*
IFRS	*753*	*81,250*	*33,242*	*24.13*	*5.13*	*36.01*	*4.94*	*3.56*	*0.89*	*10.05*	*1.02*
US-GAAP	*769*	*64,905*	*19,503*	*24.75*	*4.91*	*34.83*	*4.55*	*3.38*	*0.86*	*9.91*	*0.87*
Difference		*16,345*		*-0.62*		*1.18*		*0.18*		*0.14*	
t-test		*4.383e-15****		*0.0168***		*1.656e-6****		*0.00011****		*0.0037****	
U-test		*<2.2e-16****		*0.010****		*0.0002****		*8.99e-5****		*0.02***	

*Note: *** denotes the 1% significance level; ** denotes the 5% significance level.*

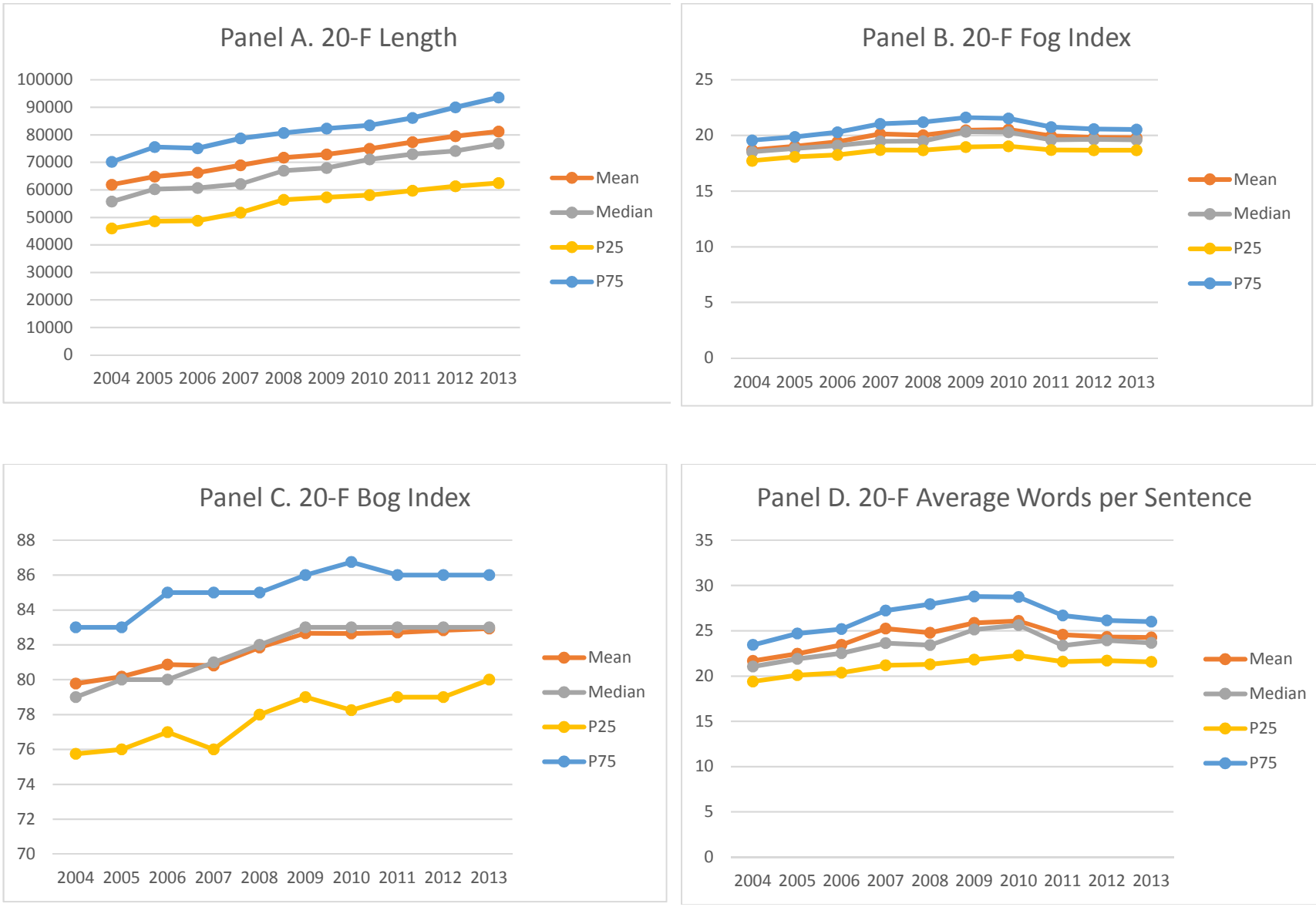
TABLE 3.9: Regression results with interactions

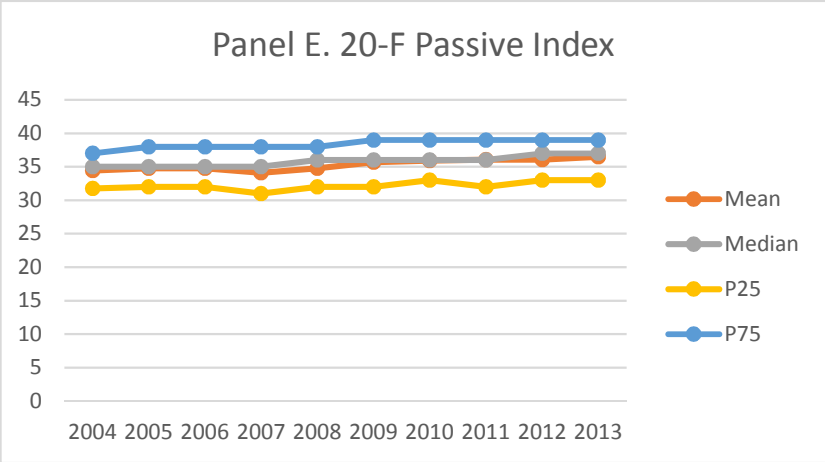
This table reports the regression results for the impacts of plain English use on the length of annual reports issued by foreign firms listed on the US stock exchange by adding interactions of accounting standards. All OLS regressions show the intercept, coefficients, and t-statistics.

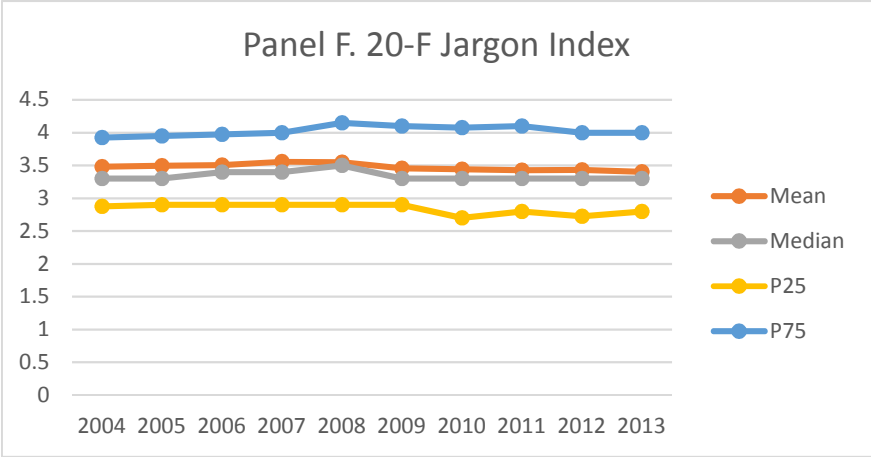
LN_NUMWORDS = $\beta_0 + \beta_1 AVE\_SEN_{it} + \beta_2 PASSIVE_{it} + \beta_3 JARGON_{it} + \beta_4 PEP_{it} + \beta_5 LN\_SIZE_{it} + \beta_6 DEBT\_RATIO_{it} + \beta_7 PTB_{it} + \beta_8 LOSS_{it} + \beta_9 VOLATILITY_{it} + \beta_{10} IFRS_{it} + \beta_{11} IFRS_{it} * AVE\_SEN_{it} + \beta_{12} IFRS_{it} * JARGON_{it} + \beta_{13} IFRS_{it} * PEP_{it} + \beta_{14} IFRS_{it} * PASSIVE_{it}$

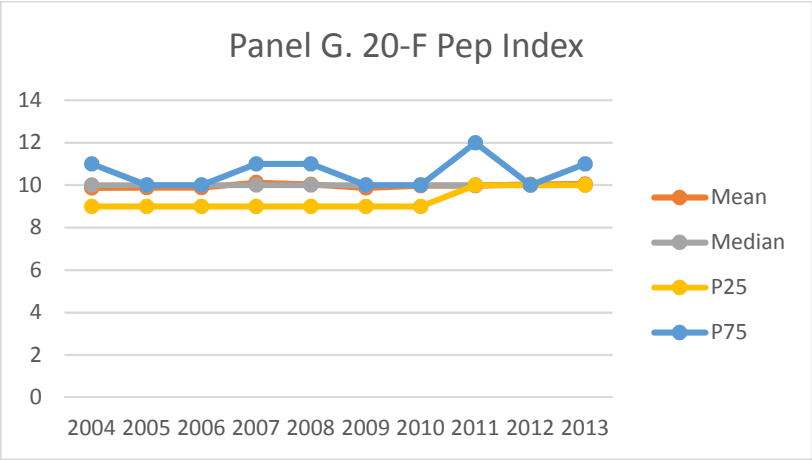
Independent Variables	**Dependent Variable (LN_WORDS)**			
	(1)	**(2)**	**(3)**	**(4)**
Intercept	8.8096***	9.4075***	8.8974***	9.4493***
	(30.292)	(24.743)	(30.691)	(25.141)
AVER_SEN	0.0063**	0.00734***	0.00406	0.00495*
	(2.215)	(2.606)	(1.423)	(1.762)
PASSIVE	0.0141***	0.013***	0.0111***	0.0098***
	(3.656)	(3.424)	(2.912)	(2.596)
JARGON	0.1412***	0.1503***	0.13091***	0.14018***
	(7.804)	(8.480)	(7.331)	(8.028)
PEP	0.0648***	0.0598***	0.0559***	0.05051***
	(3.723)	(3.523)	(3.247)	(3.016)
LN_SIZE	0.0599***	0.0521***	0.0622***	0.05402***
	(11.179)	(9.512)	(12.834)	(9.395)
DEBT_RATIO	0.4268***	0.3547***	0.38653***	0.30601***
	(9.021)	(6.715)	(8.247)	(5.854)
PTB	-0.003***	-0.0024***	-0.0022***	-0.00199***
	(-3.501)	(-3.344)	(-3.041)	(-2.858)
LOSS	-0.1192***	-0.1165***	-0.10747***	-0.1042***
	(-5.002)	(-4.991)	(-4.558)	(-4.515)
VOLATILITY	2.910***	2.8808***	3.9373***	3.868***
	(4.408)	(4.410)	(5.142)	(5.075)
IFRS	1.0484***	1.1247***	0.9988***	1.0619***
	(2.830)	(3.113)	(2.739)	(2.990)
Aver_sen:IFRS	-0.0305***	-0.0325***	-0.02931***	-0.0313***
	(-7.806)	(-8.488)	(-7.613)	(-8.307)
Passive:IFRS	0.0047	0.00351	0.00458	0.00358
	(0.95)	(0.723)	(0.940)	(0.751)
Jargon:IFRS	-0.091***	-0.0744***	-0.0775***	-0.061***
	(-3.828)	(-3.173)	(-3.303)	(-2.638)
Pep:IFRS	-0.015	-0.0209	-0.0153	-0.0207
	(-0.655)	(-0.929)	(-0.678)	(-0.937)
Industry fixed effects	No	Yes	No	Yes
Year fixed effects	No	No	Yes	Yes
Adjusted R2	0.2433	0.284	0.2684	0.3098

Figure 3.1 Trends in plain English over time




Panel E. 20-F Passive Index
45
40
35
30
25
20
15
10
5
0
2004 2005 2006 2007 2008 2009 2010 2011 2012 2013
Mean
Median
P25
P75


Panel F. 20-F Jargon Index
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
2004 2005 2006 2007 2008 2009 2010 2011 2012 2013
Mean
Median
P25
P75


Panel G. 20-F Pep Index
14
12
10
8
6
4
2
0
2004 2005 2006 2007 2008 2009 2010 2011 2012 2013
Mean
Median
P25
P75

Chapter 4: Disclosure Characteristics of Annual Reports and Information Asymmetry: Evidence from the Annual Reports of Foreign Firms Listed on the US Stock Exchange

ABSTRACT

This chapter examines the associations among two of the most interesting disclosure characteristics of annual reports and information asymmetry. It illustrates that the level of disclosures, measured by the number of words in annual reports, is negatively associated with information asymmetry between informed and uninformed investors during windows. By contrast, our research shows that an improvement in the readability or transparency of annual reports does not significantly affect information asymmetry among investors. Loss-making firms have higher information asymmetry, although the level of the disclosures of annual reports issued by loss-making firms has stronger impacts on the improvement in information asymmetry compared with profit-making firms. Annual reports following IFRS contain more information that improves information asymmetry compared with annual reports adopting US-GAAP.

Keywords: Length, Fog Index, Bog Index, annual reports, information asymmetry

1. Introduction

In recent years, the disclosure characteristics of annual reports have attracted the attention of researchers, especially the level of disclosure as well as the tone and transparency of such reports (Li, 2010). Research shows that annual reports are becoming longer, more complex, and with a more negative tone over time, negatively impacting on users and the stock market (Li, 2008; You & Zhang, 2009; Miller, 2010; Callen, Khan, & Lu, 2013; Lehavy, Li, & Merkley, 2011; Lee, 2012; Lawrence, 2013; Lundholm, Rogo, & Zhang, 2014; De Franco, Hope, Vyas, & Zhou, 2015). However, research that directly focuses on the associations among the disclosure characteristics of annual reports and information asymmetry is lacking. To bridge this gap, our research examines the relationships among the two most noticeable disclosure characteristics of annual reports and information asymmetry. We investigate the extent to which information asymmetry is affected by how much is disclosed and how information is communicated using a sample of annual reports issued by foreign firms listed on the US stock exchange.

In theory, an increase in disclosure should reduce information asymmetry by reducing private information in trading and increasing the liquidity of stocks; however, some research still finds a negative association between the disclosure of annual reports and information asymmetry (Verrecchia, 1983; Diamond & Verrecchia, 1991; Heflin, Shaw, & Wild, 2005; Leuz & Verrecchia, 2000; Petersen & Plenborg, 2006; Brown & Hillegeist, 2007; Botosan, 1997), whereas other researchers find a positive relationship (Kim & Verrecchia, 1994; McNichols & Trueman, 1994). The reason for this positive relationship is the difference in the information-processing ability of

investors (Kim & Verrecchia, 1994). Previous research focuses on the amount of disclosure or mixture of quantity and quality of information via ratings. However, these measurements can be subjective because of analysts' ratings, unavailability, and high costs (Healy, Hutton, & Palepu, 1999). More importantly, these ratings combine both the quantity and the quality of information into one measurement.

Compared with previous research, our study focuses on the two main disclosure characteristics of annual reports, which are the level of disclosure and the readability of annual reports. In particular, it uses a different approach to measure the amount of disclosure via the number of words in annual reports. Our approach thus overcomes the disadvantages of previous measurements, as it is objective, is less costly, has high availability for all users and firms, and is easy to apply to a large sample. Moreover, we also test the relationship between the transparency of annual reports and information asymmetry using readability, which has not thus far been attempted in previous research. Using different measurements for the readability of annual reports, different proxies of information asymmetry, and different windows in our tests, we also test the robustness of our results.

Our research shows that the level of disclosure of annual reports, measured by the number of words in 20-F forms, is negatively associated with information asymmetry, while the readability of annual reports does not improve information asymmetry among investors. As the first research on the associations among the disclosure characteristics of annual reports and information asymmetry, we provide optimistic views about the recent tendency of annual reports to improve in terms of information asymmetry. In detail, we show that an increase in the length of annual reports helps reduce information asymmetry. This means that longer annual reports provide more information to investors and reduce the gaps between informed and uninformed investors. We also use a different approach to confirm the negative association between the level of disclosure and information asymmetry (Verrecchia, 1983; Diamond & Verrecchia, 1991; Heflin et al., 2005; Leuz & Verrecchia, 2000; Petersen & Plenborg, 2006; Brown & Hillegeist, 2007; Botosan, 1997).

Moreover, this is the first empirical research to show that investors pay more attention to the amount of information than to how that information is displayed. In other words, investors need information regardless of the level of complexity. The association between the readability of annual reports and information asymmetry is negative or insignificant. We thus confirm that complex annual reports still contain more information, as mentioned by Bushee, Gow, and Taylor (2018).

Previous research shows that the performance of firms correlates with their disclosure policies. The annual reports of loss-making firms are longer and more complicated than those of profit-making firms (Li, 2008; Aymen, Mhamed, & Badreddine, 2016). We also test whether the associations among disclosure characteristics and information asymmetry differ between loss-making and profit-making firms. Although loss-making firms have higher information asymmetry

than profit-making firms, the level of disclosure in their annual reports improves information asymmetry more than for profit-making firms. The longer annual reports of loss-making firms provide more information to investors with the aim of reducing information asymmetry among investors; by contrast, the less readable annual reports provided by those firms cannot influence information asymmetry.

Finally, we test the relationships among the disclosure characteristics of annual reports on information asymmetry under different accounting standards. Although there is no difference in the impacts of the level of disclosure and readability of annual reports on the information asymmetry between annual reports following IFRS and those adopting US-GAAP, we find that firms following IFRS have lower information asymmetry than firms adopting US-GAAP. Moreover, we find that the fact that annual reports under IFRS are statistically longer than those under US-GAAP generally makes the information asymmetry of the former lower than that of the latter because of the negative association between the length of annual reports and information asymmetry.

The remainder of this article is organized as follows. The following section presents the literature review on the disclosure characteristics of annual reports and their associations with information asymmetry. Section 3 develops the hypotheses. Section 4 illustrates the research design and explains the variable measurement in our analysis. The sample selection is described in section 5. Section 6 shows our findings and discussion. Finally, section 7 concludes.

2. Literature Review

2.1 Disclosure characteristics of annual reports and stock market reaction

According to efficient market theory, stock prices reflect publicly available information, which is mostly extracted from the official disclosure of firms. Consequently, the US Securities and Exchange Commission (SEC) "requires public companies to disclose meaningful financial and other information to the public. This provides a common pool of knowledge for all investors to use to judge for themselves whether to buy, sell, or hold particular security" (SEC, 2013). This regulation encourages firms to disclose timely information to investors throughout fiscal years. Among firms' disclosure, annual reports contain more price-value information than other disclosures provided that they are considered to be the primary information source on firms for market participants. Additionally, those reports guarantee equal access to information for all investors, which is expected to reduce information asymmetry among them (You & Zhang, 2009).

As a result, the characteristics of annual reports have attracted the attention of both practitioners and researchers, notably the level of disclosure (how much is disclosed), the tone (what you say), and the transparency (how information is communicated) of such reports (Li, 2010). There is extensive research on the impacts of those characteristics on the stock market. Most extant research shows a significant increase in the length and complexity of annual reports

over time (Li, 2008; You & Zhang, 2009; Miller, 2010; Lehavy et al., 2011; Lundholm et al., 2014; Cazier & Pfeiffer, 2016; see also Chapter 2). Unfortunately, such an increase causes a negative impact on the stock market. Courtis (1986) shows a negative relationship between readability and risks, while Subramanian, Insley, and Blackwell (1993) report a negative association between firm performance and readability. Similarly, Li (2008) and Aymen et al. (2016) show that firms with bad performance tend to hide adverse information by issuing more complex annual reports compared with firms performing well.

Additionally, researchers have shown the adverse effects of the length and readability of annual reports on stock markets through the underreaction of investors (Callen et al., 2013; Lee, 2012; You & Zhang, 2009) as well as the lack of small investors trading (Lawrence, 2013; Miller, 2010). The readability of annual reports also affects analysts' behaviors. Less readable annual reports cause greater dispersion, lower accuracy, greater overall uncertainty in analysts' earnings forecasts, and less readable analyst reports (Lehavy et al., 2011; De Franco et al., 2015).

In terms of the tone used in annual reports, a negative tone is recorded in both 10-K forms and MD&A disclosures. The linguistic features of annual reports are attributed to firm performance and earnings persistency (Li, 2008, 2010). In addition, the textual characteristics of annual reports are associated with stock returns and volume trading during windows (De Franco et al., 2015; Henry & Leone, 2016; Loughran & McDonald, 2011).

In general, all three of these disclosure characteristics of annual reports are attributed to the stock market. Investors pay attention not only to the quantity of information but also to the quality of information including the tone and transparency of communication in annual reports (Loughran & McDonald, 2016; Drake, Roulstone, & Thornock, 2016).

2.2 Disclosure characteristics of annual reports and information asymmetry

Despite the popularity and certainty of the research findings on the association between the disclosure characteristics of annual reports and market reaction, research on the associations between the disclosure characteristics of annual reports and information asymmetry is controversial and limited because of the use of the rating system in the measurement[11]. The conclusions about these relationships are therefore mixed. Theoretically, an increase in disclosure reduces information asymmetry by preventing informed investors from using private information to make investment decisions and improving the liquidity of stocks (Verrecchia, 1983; Diamond & Verrecchia, 1991). Supporting this conclusion, some empirical research reports a negative association between disclosure and information asymmetry (Heflin et al., 2005; Leuz & Verrecchia, 2000; Petersen & Plenborg, 2006; Brown & Hillegeist, 2007; Botosan, 1997). Regardless of

[11] The advantages and disadvantages of rating systems are discussed by Healy et al. (1999). We also discuss more details about these measurements in the research design.

whether the increase in disclosure is voluntary or mandatory, more frequent disclosure additionally improves the information asymmetry of stocks (Fu, Kraft, & Zhang, 2012).

Some research confirms that providing accounting information or public information is the main way to reduce information asymmetry. Disclosing supplementary accounting information on the share of assets, liabilities, revenues, and expenses improves information asymmetry (Diamond & Verrecchia, 1991; Lim, Yeo, & Liu, 2003; Petersen & Plenborg, 2006; Vanza, Wells, & Wright, 2018). Additionally, high-quality auditing and accounting standards contribute to the improvement of information asymmetry (Zhou, 2007). Other research shows the association between the format of annual reports and information asymmetry (Yoon, Zo & Ciganek, 2011; Gajewski & Li, 2015). In particular, Internet-based disclosures or annual reports that adopt XBRL lower information asymmetry in the stock market by "enhancing the communication quality, improving reputation, attracting potential investors, reducing information distribution costs, facilitating the information gathering process" (p.115) (Gajewski & Li, 2015).

By contrast, other research finds that an increase in disclosure exacerbates information asymmetry and lowers liquidity, but raises the trading volume. Investors' different processing abilities cause the difference in analyzing and understanding the information disclosed by firms. Such research assumes that if firms do not disclose, there is no opportunity for analytical traders to exploit their ability to process public information. This lessens the possibility of information asymmetries arising (Kim & Verrecchia, 1994; Krinsky & Lee, 1996; Coller & Yohn, 1997). Amiram, Owens, and Rozenbaum (2016) propose that the increase or decrease in information asymmetry due to firms' disclosures depends on the timeliness of the information. In detail, if the information released is new to all investors, it enlarges information asymmetry among them because of their different cognitive abilities. Conversely, disclosure that is new to only some investors decreases information asymmetry.

Overall, research concludes that firms' disclosure affects information asymmetry either positively or negatively. However, extant research concentrates on the level of disclosure and cannot separate the impacts of the quality and quantity of disclosure on information asymmetry.

3. Hypothesis Development

3.1 Level of disclosures in annual reports and information asymmetry

The relationship between the level of disclosure and information asymmetry is a controversial problem in both theoretical and empirical research. Some research shows that such a relationship is positive because of the limited processing ability of informed and uninformed investors (Kim & Verrecchia, 1994). By contrast, other research shows the inverse association between these two variables (Heflin et al., 2005; Leuz & Verrecchia, 2000; Petersen & Plenborg, 2006; Brown & Hillegeist, 2007; Botosan, 1997). Overall, however, all research concludes that the level of disclosure affects information asymmetry.

You and Zhang (2009) show that annual reports contain more price-relevant information than the other disclosures of firms. Annual reports strictly regulated by the SEC are considered to be the main official communication of firms to investors. Indeed, the level of disclosures in annual reports is associated with information asymmetry. In detail, there is a negative relationship between the amount of financial information published in annual reports and information asymmetry (Petersen & Plenborg, 2006; Brown & Hillegeist, 2007). Consequently, we suppose that the length of annual reports, as the measure of quantity disclosures in annual reports, correlates with information asymmetry, as reflected in the following hypothesis:

Hypothesis 1. Longer annual reports provide more information to investors, which leads to a reduction in information asymmetry.

3.2 Transparency of disclosures in annual reports and information asymmetry

As mentioned above, the transparency of the disclosure of annual reports is defined as how the information is communicated, measured by the readability of such reports. In contrast to numerous studies that test the association between the level of disclosure and information asymmetry, no research directly shows the relationship between the transparency of disclosure and information asymmetry; however, this relationship is indirectly explained in related research.

You and Zhang (2009) assume that the complexity of information affects the speed of its diffusion, which is expected to cause a difference in investors' understanding because of the different processing abilities of market participants. Similarly, poor disclosure delays the market response and creates another layer of information asymmetry between informed and uninformed users (Lee, 2012; Callen et al., 2013). Hsieh, Hui, and Zhang (2016) and Loughran and McDonald (2014a) explain that more readable annual reports enhance certainty, which is expected to decrease information asymmetry between sophisticated and unsophisticated investors. Another study explains the relationship between the readability of annual reports and information asymmetry via demand for high-cost information. In detail, complexity in annual reports raises demand for private information for trading, which is expected to increase information asymmetry among investors (Filzen & Schutte, 2017).

Conversely, some research expects a negative association between the readability of annual reports and information asymmetry. For example, Bushee et al. (2018) assume that complex language does not mean it is less informative. Less readable language results from complex business or regulations, but still provides information that improves information asymmetry for investors. Dyer, Lang, and Stice-Lawrence (2016) provide an additional explanation of this negative association through the advance in technology, which helps investors deal with complex reports; therefore, the readability of annual reports is not expected to affect information asymmetry (Bonsall, Leone, Miller, & Rennekamp, 2017).

In summary, the relationship between the readability of annual reports and information asymmetry is still debated and empirical results are lacking. We support the argument of the positive impact of readability on information asymmetry. We hypothesize that:

Hypothesis 2. More readable annual reports enhance the understanding of investors, which also improves information asymmetry.

3.3 Disclosure characteristics of loss-making firms and information asymmetry

The performance of firms has substantial impacts on their disclosure policies. To obfuscate bad news, the annual reports of loss-making firms are longer and less readable than those of profit-making firms (Subramanian et al., 1993; Li, 2008; Aymen et al., 2016). Additionally, Ertimur (2004) shows that loss-making firms have higher information asymmetry than profit-making firms owing to their poorer disclosures including both mandatory and voluntary disclosures. The poor disclosure policy of loss-making firms leads to more noise, higher demand for private information, and less certainty in valuing, which reduces the liquidity of stocks; therefore, the information asymmetry of loss-making firms is higher.

However, some research shows a different view of the disclosure policy of loss-making firms. They find that such firms compensate for the weakness in their performance by increasing voluntary disclosures to investors about the reasons for their loss and estimates of future forecasts (Chen, DeFond, & Park, 2002; Lougee & Marquardt, 2004). Supporting this opinion, Guay, Samuels, and Taylor (2016) find a positive relationship between the complexity of annual reports and voluntary disclosures, which infers that loss-making firms with more complex annual reports provide more voluntary disclosures. Additionally, other research assumes that the increase in the complexity of the annual reports issued by loss-making firms aims not only for obfuscation but also for information provision. Loss-making firms have to provide reasons for their poor performance, and it is more difficult to explain poor performance with simple language. Consequently, the longer and more complex annual reports of loss-making firms are expected to provide more information to investors, which lowers information asymmetry among them (Bloomfield, 2008; Bushee et al., 2018).

In conclusion, the informativeness of the disclosures issued by loss-making firms is still debated. In particular, there is limited evidence on the relationship between the disclosure characteristics and information asymmetry of loss-making firms. Formally, we propose the following hypotheses:

Hypothesis 3 (a): Loss-making firms have higher information asymmetry than profit-making firms.

Hypothesis 3 (b). There are no differences in the associations among the disclosure characteristics of the annual reports filed by loss-making firms and information asymmetry compared with those associations in the case of profit-making firms.

3.4 Accounting standards and information asymmetry

The SEC allows foreign firms listed on the US stock exchange to choose which accounting standards are followed when preparing their annual reports. Annual reports issued by foreign firms following IFRS are longer but more readable than reports adopting US-GAAP (Lundholm et al., 2014; Cheung & Lau, 2016; see also Chapter 2). Since 2007, the SEC has allowed removing reconciliation from 20-F forms since there is sufficient convergence between IFRS and US-GAAP. Kim, Li, and Li (2012) assume that removing the reconciliation from annual reports does not have significant impacts on the level of disclosure or liquidity of firms following IFRS. By contrast, this action of the SEC has not had a positive impact on information asymmetry since it helps lower cost, and removing reconciliation constrains the effect of IFRS, which improves the quality of the information in annual reports adopting IFRS and reduces information asymmetry (Chen, Deng, Gupta, & Sami, 2015). Moreover, annual reports under IFRS are assessed as good quality ones with more disclosures, less repetition, and more comparability (Lang & Stice-Lawrence, 2015). Some research shows that changing from national accounting standards to IFRS improves information asymmetry among investors (Leuz & Verrecchia, 2000; Zhou, 2007). From that, we expect the advantages of the disclosure characteristics in annual reports following IFRS to affect information asymmetry. The hypotheses are as follows:

Hypothesis 4 (a). Annual reports following IFRS have lower information asymmetry than annual reports adopting US-GAAP.

Hypothesis 4 (b). The impacts of the disclosure characteristics of annual reports on information asymmetry depend on the accounting standards that firms follow.

4. Research Design

In contrast to previous research on how the disclosure characteristics (level of disclosure and transparency) of annual reports affect information asymmetry, our study uses a different approach by adopting the length of annual reports to measure the level of disclosure. Additionally, we suppose that the readability or transparency of annual reports affects information asymmetry. We also assume that the differences in the disclosure policies of firms with different performance levels and following different accounting standards influence information asymmetry. Therefore, we run a number of regressions to test the hypotheses mentioned above. Appendix 4.1 summarizes the definitions and expected signs.

The following equations summarize the regression models used in this research. The first model regresses the impacts of the level of disclosures and readability of annual reports on information asymmetry. The second model is used to test the associations among the disclosure characteristics of annual reports, information asymmetry, and loss-making firms. For the third model, we want to check the information asymmetry of firms following IFRS:

$$\text{Infor.Asymmetry} = \beta_0 + \beta_1 Lnwords_{it} + \beta_2 LnFog_{it} + \beta_3 LnPrice_{it} + \beta_4 LnTurnover_{it} + \beta_5 Volatility_{it} + \varepsilon \quad (1)$$

$$\text{Infor.Asymmetry} = \beta_0 + \beta_1 Lnwords_{it} + \beta_2 LnFog_{it} + \beta_3 LnPrice_{it} + \beta_4 LnTurnover_{it} + \beta_5 Volatilityit + \beta_5 Loss_{it} + \beta_5 Lnword*Loss_{it} + \beta_5 LnFog*Loss_{it} + \varepsilon \quad (2)$$

$$\text{Infor.Asymmetry} = \beta_0 + \beta_1 Lnwords_{it} + \beta_2 LnFog_{it} + \beta_3 LnPrice_{it} + \beta_4 LnTurnover_{it} + \beta_5 Volatility_{it} + \beta_5 IFRS_{it} + \beta_5 Lnword*IFRS_{it} + \beta_5 LnFog*IFRS_{it} + \varepsilon \quad (3)$$

4.1 Measuring information asymmetry

Information asymmetry occurs when one party possesses superior information about the value of firms than other parties. Because of the advantages of more information, informed investors enjoy more trading benefits than others. Consequently, information asymmetry has negative impacts on investors' trading decisions and the efficiency of stock markets (Kim & Verrecchia, 1994; Hu & Prigent, 2018). However, as we cannot directly measure information asymmetry, several proxies are used in extant research such as the bid–ask spread, price impact, trading volume, and probability of information-based trading (Leuz & Verrecchia, 2000; Brown & Hillegeist, 2007; Amiram et al., 2016). The most popular measurement used in research is the bid–ask spread (Kim & Verrecchia, 1994; Chung & Charoenwong, 1998; Leuz & Verrecchia, 2000; Heflin et al., 2005; Fu et al., 2012; Chen et al., 2015). Therefore, we use the bid–ask spread as the main proxy for information asymmetry.

The bid–ask spread (*spread*) is set by dealers who expect to compensate their losses when trading with informed investors by setting a wider spread. It includes three components: the order processing cost, inventory holding cost, and adverse selection cost. Of these, the adverse selection component is generally linked to the information flow, and is normally called the "information asymmetry risk" (Lee & Yahn, 1997; Zhou, 2007). This element is the main component of the bid–ask spread (George, Kaul, & Nimalendran, 1991; Zhou, 2007).

The effect of information releases on information asymmetry is temporary, existing only for around 10 days after announcement (Amiram et al., 2016). The effect of information on trading volume as well as return volatility, as the other proxies of information asymmetry, is also short-lived (You & Zhang, 2009; Miller, 2010). Lehavy et al. (2011) and Li (2017) suggest that testing the impacts of disclosures or informativeness on stock markets should use a window around the announcement dates. As a result, we use the average daily bid–ask spread around the dates on which annual reports are available on the EDGAR website. In the main results, we use the spread/price impact over three days (-1; +1) for testing, while the spread/price impacts over five days (-2; +2) and 21 days (-10; +10) are used in the robustness tests.

We follow previous research to calculate the average daily bid–ask spread during windows as follows: (Ask-Bid)/((Ask+Bid)/2) (Mohd, 2005; Heflin et al., 2005; Fu et al., 2012; Chen et al., 2015; Amiram et al., 2016). Price impact, a measure of the illiquidity of stocks, is identified by the

daily absolute return divided by the daily trading value, as mentioned by Amihud (2002) and Daske, Hail, Leuz, and Verdi (2008).

4.2 Level of disclosures in annual reports

Prior research on the relationship between the level of disclosures and information asymmetry has used different measurements to identify disclosures. Most studies use the analysts' ratings by the Association of Investment Management and Research (Lang & Lundholm, 1993; Healy et al., 1999; Botosan & Plumlee; 2002; Brown & Hillegeist, 2007). Other research has used the Financial Analysts' Rating and amount of disclosure (Heflin et al., 2005; Botosan, 1997). However, these measurements are subjective, unavailable for all firms and all investors, expensive, and difficult to use for a large sample (Healy et al., 1999). In addition, they do not measure the level of disclosure and use a combination of the quantity and quality of disclosures via ratings.

Fortunately, Li (2008), as the pioneer of applying the Perl computing language to measure the disclosure characteristics of annual reports, provides a solution to overcome the disadvantages of previous measurements. He uses Perl language to identify the number of words in annual reports as well as their readability given the negative association between the number of words in annual reports and stock prices or market reactions (Callen et al., 2013; Lee, 2012; You & Zhang, 2009; Miller, 2010; Lehavy et al., 2011; Lawrence, 2013). Therefore, the number of words in annual reports is assumed to measure the quantity of the information contained therein objectively, leading to fewer errors; it is also convenient for a large sample (Lang & Stice-Lawrence, 2015; Bonsall et al., 2017). Understanding the relationship between the length of annual reports and information asymmetry helps accounting standard setters identify whether the current tendency toward issuing longer annual reports provides more information to investors or is repetitive and unnecessary. As a result, we use the number of words to measure the level of disclosures in annual reports. Following the suggestion of Li (2008), Loughran and McDonald (2014a, 2014b), and Bonsall et al. (2017), we use the package Lingua::EN::Fathom of Perl language to identify the number of words. As mentioned above, we expect the number of words in the annual report to have a negative association with the spread/price impact.

4.3 Transparency of annual reports

To measure how information is communicated in annual reports, we use two measurements of the readability of annual reports. The first measurement is the Fog Index, which is the most popular in previous research (Lee, 2012; Lehavy et al., 2011; Li, 2008; Lundholm et al., 2014; Loughran & McDonald, 2014b; Miller, 2010). The Fog Index, first introduced by Gunning in 1952, is based on the number of educational years that readers need to understand the text. It contains two components, namely the number of words per sentence and the percentage of complex words in a document, as follows:

$$\textbf{Fog Index} = 0.4 * (\textbf{words\_per\_sentence} + \textbf{percentage\_complex\_words})$$

The higher the Fog Index is, the less readable annual reports are. The Fog Index is also acquired by the package Lingua::EN::Fathom of Perl. However, this measurement has been criticized by researchers because of the irrelevance of the second component and simple definition of complex words (Loughran & McDonald, 2014a; Bonsall et al., 2017)[12]. Based on the recommendation of the SEC about the language used in firms' disclosures, other research suggests using plain English as regulated in Rule 421 (SEC, 1998) to measure the readability of annual reports. This measurement is referred to as the Bog Index[13] (Miller, 2010; Bonsall et al., 2017):

Bog Index = Sentence Bog + Word Bog – Pep

Similarly, reports are assessed as more difficult to read when they have a higher Bog Index. This index is identified by StyleWriter software without coding the scripts; moreover, it combines various elements of plain English in one measurement compared with the Fog Index. We follow the instructions of Miller (2010) and Bonsall et al. (2017) to identify the Bog Index of the annual reports issued by foreign firms listed on the US stock exchange. The empirical results of the Bog Index are displayed in the robustness test. We expect the Fog Index/Bog Index to be positively associated with information asymmetry.

4.4 Control variables

The main determinants of the bid–ask spread are stock price, turnover, and volatility[14] (Copeland & Galai, 1983; Chung & Charoenwong, 1998; Fu et al., 2012; Amiram et al., 2016). Therefore, we control for these variables in our empirical regressions. To test Hypotheses 3 and 4, two more dummy variables (e.g., *loss* and *IFRS*) are included in model (1). The definitions and expected signs of the control variables are as follows:

Stock price: Stocks with a low price are likely to be new and riskier than others; therefore, they are less informative and less liquid. Consequently, a negative correlation between stock prices and information asymmetry is expected. The stock price is identified by taking the natural logarithm of the average daily stock price in year t (Gajewski & Li, 2015).

Turnover: Trading volume means the willingness of market participants to trade stocks. Previous research shows that trading volume is negatively associated with information asymmetry since stocks are traded more frequently; consequently, they are more attractive to investors, which

[12] The Fog Index defines complex words based on their syllables. Therefore, Loughran and McDonald (2014a) show that 45,000 of the most popular "complex" words according to the Fog Index are "simple and common business terms" (p. 1645). Bonsall et al. (2017) find that the Fog Index only focuses on average sentence length and complex words and overlooks the other elements of plain English.

[13] The Bog Index is identified by using those elements of plain English regulated in Rule 421 of the SEC such as using short sentences, the active voice, and good writing as well as avoiding legal and financial jargon.

[14] Firm size is also a key determinant of the bid–ask spread; however, this variable is shown to be highly correlated with the other explanatory variables such as the complexity of annual reports, stock price, volume, and returns. Therefore, it is not mentioned in this research (Chung & Charoenwong, 1998; Ertimur, 2004; Li, 2008; Loughran & McDonald, 2014a; Miller, 2010; Lehavy et al, 2011). In any case, including size into the regressions does not change the results, which are available on request.

improves the liquidity of stocks (Copeland & Galai, 1983; Leuz & Verrecchia, 2000). The improvement in the liquidity of stocks is assumed to negatively affect information asymmetry. We follow Amiram et al. (2016) to identify the turnover ratio by taking the natural logarithm of the average daily turnover ratio in year t. The turnover ratio is calculated by multiplying the daily volume trading by the daily stock price and then dividing into the market cap of stocks.

Volatility: Stock volatility means uncertainty or risk in the capital market. The more volatile stocks are, the more risk investors take. Stocks with more risk become less attractive to investors and have less liquidity. Consequently, the volatility of the stock return is proposed to be positively associated with the bid–ask spread (Glosten, 1987). We calculate the volatility of the stock return by taking the standard deviation of daily stock returns.

Loss: Loss-making firms are assumed to obfuscate bad information, which leads to more information asymmetry among investors (Li, 2008; Bushee et al., 2018; Ertimur, 2004). Based on the earnings of firms in their annual reports, we set 0 for a firm-year without a loss and 1 for a firm-year with a loss.

IFRS: The disclosures of annual reports following IFRS are reported as high quality; therefore, we expect annual reports following IFRS to be negatively correlated with information asymmetry. Based on firms' announcements at the beginning of the 20-F forms, we set 0 for annual reports following US-GAAP and 1 for annual reports adopting IFRS.

5. Sample Selection

Our research focuses on the annual reports of foreign firms listed on the US stock exchange. According to the requirements of the SEC, foreign firms, normally identified as "foreign private issuers," trade less than 50% of their shares on the US stock exchange and annually fill in a 20-F form. We use Python to auto-download all the 20-F filings of foreign firms from 2005 to 2014 from EDGAR and find 7,588 filings. Using this sample of foreign firms from different countries, cultures, and business environments, we aim to provide a global overview of the effects of the disclosure characteristics of annual reports on information asymmetry.

We only choose filings whose fiscal year ends on December 31 and remove all filings operating in the finance and insurance industries. We also adopt filings in text files that suit analysis using the Perl language. We include the filings of firms listed on the NASDAQ and NYSE that have RIC codes (Reuters Instrument Code) available in Thomson Reuters Datastream. As bid and ask prices are available from June 4, 2006 in Thomson Reuters Datastream, our research period runs from 2006 to 2014. All filings with missing data are removed from our sample. Our final sample has 1,348 observations (Table 4.1).

[INSERT TABLE 4.1 HERE]

To measure the length and readability of annual reports, it is necessary to clean the raw text files downloaded from the EDGAR website. Following Bonsall et al. (2017), Li (2008), and

Loughran and McDonald (2014a, 2014b), we use the packages of Perl language to clean the raw text files that contain the text and other elements. Appendix 4.2 presents the details of cleaning the 20-F forms. We then use the Lingua::EN:: Fathom package of the Perl language and StyleWriter Professional Edition to identify the length, Fog Index, and Bog Index of the post-cleaned annual reports.

6. Findings and Discussion

We use different measurements of disclosure characteristics and two proxies of information asymmetry with a difference in the number of window days to test our hypotheses. Additionally, we report the regression statistics with and without fixed effects including industrial fixed effects and year fixed effects to demonstrate that our inferences are robust.

6.1 Descriptive analysis

Figure 4.1 summarizes the fluctuation in the two proxies of information asymmetry around the announcement date. Panel A of Figure 1 presents the fluctuation in the bid–ask spread over the windows. The date on which the 20-F forms are available on the EDGAR website is called date zero. Overall, the bid–ask spread is almost stable in the 21-day window. The bid–ask spread fluctuates from 0.013 to 0.014 over these 21 days; however, there is an increasing tendency in the bid–ask spread from one day before to two days after the available date on the EDGAR website. Then, the bid–ask spread decreases slightly after the third ex-post day, remaining stable again in the subsequent days. As shown in Panel A of Figure 4.1, the release of annual reports thus affects the bid–ask spread, or information asymmetry in general.

[INSERT FIGURE 4.1 HERE]

As shown in Panel B of Figure 4.1, the tendency of price impact, another proxy of information asymmetry, fluctuates rapidly during the window. The fluctuation in trading volume is explained by the dramatic change in price impact, whereas the tendency of daily stock returns during the windows remains stable. Panel C and Panel D of Figure 4.1 show the change in daily stock returns and daily trading volume during this period. The period from one day before to three days after the announcement date witnesses the most dramatic change in price impact, trading volume, and stock returns, which is consistent with the findings of You and Zhang (2009). Overall, Figure 4.1 shows that annual reports affect both main proxies of information asymmetry during windows. Previous research suggests using three- or five-day windows to test the informativeness of annual reports, which are also used in our analysis.

Table 4.2 presents the descriptive statistics of all the variables used in our research. The mean of the natural logarithm of the bid–ask spread during three-day windows is -5.791, equivalent to 0.01483 of the bid–ask spread. The average bid–ask spread in our sample is lower than that in Chung and Charoenwong (1998), 0.0183, since the information environment has improved significantly over time. The bid–ask spread in the US stock market is also lower than that in other

stock markets such as the German stock market and the Chinese Stock market since the US stock market is highly developed (Leuz & Verrecchia, 2000; Zhou, 2007). The average daily price impact over three-day windows is 0.1126, taking the natural logarithm of this price impact at -7.315. Table 4.2 also illustrates the daily average bid–ask spread and price impact in five- and 21-day windows.

[INSERT TABLE 4.2 HERE]

The natural logarithm of the number of words in 20-F forms is 11.128, equivalent to 73,388. The annual reports of foreign firms listed on the US stock exchange are much longer than those of US firms (Lundholm et al., 2014; see also Chapter 2). The expectation to improve the comparability and reliability of the annual reports issued by foreign firms is assumed to explain the difference in the length of 20-F and 10-K forms (French & Poterba, 1991; Lundholm et al., 2014).

The average Fog Index of 20-F forms is 19.95, which means that readers need more than 19 education years to read and understand the annual reports issued by foreign firms listed on the US stock exchange[15]. Concurring with our result, Lang and Stice-Lawrence (2015) also report a Fog Index of 19.52 for English-language annual reports in 42 countries from 1998 to 2011. Lundholm et al. (2014) record an average Fog index of the MD&A part of 20-F forms of 17.54 in 2000–2012. However, Li (2008) shows that the readability of the MD&A part is lower than the rest of the filings. In comparison with the readability of 10-K forms, the Fog Index of 20-F forms is slightly higher (Li, 2008; Lehavy et al., 2011; Bonsall et al., 2017).

Table 4.2 shows the means of the control variables. In our sample, 26.7% of the observations show negative earnings. Based on the declarations of firms at the beginning of their annual reports, we identify that 47.9% observations in our sample adopt IFRS to file their 20-F forms. The average natural logarithms of turnover, price, and volatility are 0.5325, 2.703, and -3.592, respectively.

Table 4.3 shows the Pearson correlations among the variables used in our analysis. In general, the correlations among the bid–ask spreads and price impacts are high at more than 80%. Conversely, the independent variables are correlated with the dependent variables at a low level. The length of annual reports, turnover, and price are negatively correlated with both proxies of information asymmetry, whereas volatility, the Fog Index, and the Bog Index have positive correlations with the dependent variables. These results are consistent with previous research.

[INSERT TABLE 4.3 HERE]

6.2 Disclosure characteristics of annual reports and information asymmetry

[15] The unfriendly writing style used in 20-F forms causes the high Fog Index with longer sentences and complex words (three or more syllables). In fact, long sentences and a high percentage of complex words violate the SEC's suggestions about communicating to users. These results are available upon request.

Table 4.4 presents the empirical results of Equation (1), which illustrate the impacts of both the disclosure characteristics of annual reports on information asymmetry. In detail, it shows the effects of both the number of words and the Fog Index on the two proxies of information asymmetry with the control variables and fixed effects.

[INSERT TABLE 4.4 HERE]

6.2.1 Level of disclosure and information asymmetry

As shown in Table 4.4, our results support the finding that an increase in the level of disclosures reduces information asymmetry among investors (Verrecchia, 1983; Diamond & Verrecchia, 1999; Heflin et al., 2005; Leuz & Verrecchia, 2000; Petersen & Plenborg, 2006; Brown & Hillegeist, 2007; Botosan, 1997). The negative coefficient of the number of words in Table 4 is significant at the 99% confidence level, even in the cases of controlling for the fixed effects. The same results are also recorded for both proxies of information asymmetry. The number of words is negatively associated with both the bid–ask spread and the price impact with t-values of -4.088 and -6.875, respectively. Adding fixed effects into the regressions does not change the negative association or significance level. The regression with fixed effects testing the association between the length of annual reports and bid–ask spread has a t-value of -3.110. A similar trend is recorded at a t-value of -6.494 when replacing the bid–ask spread by price impact in the regression with fixed effects.

As a consequence, we assume that longer annual reports improve information asymmetry. We also infer that the recent tendency toward significantly increasing the length of annual reports provides more information to the public with the aim of preventing investors from using private information in trading, leading to an improvement in information asymmetry. Although research shows negative impacts of the substantial increase in the number of words in annual reports on users and on the market reaction, our research provides a positive view about such an increase in terms of information asymmetry.

6.2.2 Readability of annual reports and information asymmetry

We expect a positive association between the readability of annual reports and information asymmetry; however, Table 4.4 shows different results to our expectation, namely a negative relationship between the Fog Index and proxies of information asymmetry. The regression results of Equation (1) using the Bog Index to test the association between the transparency of annual reports and information asymmetry are mentioned in the robustness test.

When using the Fog Index to measure the readability of annual reports in Equation (1) without fixed effects, the Fog Index shows a negative association with the bid–ask spread at the 99% confidence level with a t-value of -2.721. This means that more readable annual reports (lower Fog Index) do not improve information asymmetry. This association remains significant at 95% when fixed effects are included in the regression with a t-value of -1.737. Similarly, the same

association between the Fog Index and information asymmetry is found when replacing with the second proxy of information asymmetry. The Fog Index is negatively associated with price impact at the 99% confidence level with a t-value of -2.769. When including fixed effects into the regression with price impact, the association is still significant at 95%. Overall, we thus reject the hypothesis of a positive association between the readability of annual reports and information asymmetry. In other words, we do not find sufficient evidence to conclude that the improvement in the transparency of annual reports can diminish information asymmetry among investors.

Our result is consistent with the suggestion of Bonsall et al. (2017), who assume that the quantity of information, not the readability of reports, affects information asymmetry. Bushee et al. (2018) also draw the same conclusion about the ambiguous relationship between complexity and information asymmetry. As the linguistic complexity in firms' disclosures results from the complexity of the business environment and intention of managers, the complexity of disclosures has both negative and positive effects on information asymmetry. In other words, it is difficult to identify an accurate relationship between the readability of annual reports and information asymmetry.

Irrespective of including readability into the regressions that test the relationships among the disclosure characteristics and information asymmetry[16], the positive association between the readability of annual reports and information asymmetry cannot be confirmed with certainty. In other words, the improvement in the readability of annual reports cannot bridge the information gap among investors. The more readable annual reports are, information asymmetry could be larger or unaffected. Conversely, the longer annual reports are, the better information asymmetry is.

The control variables are significantly associated with information asymmetry. As shown in Table 4.4, turnover and price are negatively attributed to information asymmetry, similar to our expectation and in line with extant research on the determinants of information asymmetry. By contrast, the more volatile stock returns are, the larger information asymmetry is. The adjusted R square value is more than 40%. In particular, our model can explain more than 50% of the change in price impact.

In conclusion, both the disclosure characteristics of annual reports affect information asymmetry. In particular, the level of disclosure improves information asymmetry, whereas the readability of annual reports worsens it. Hence, information asymmetry is negatively affected by the level of disclosure. The complexity of annual reports cannot prevent investors from extracting information from annual reports to enhance their information environment when making decisions and reduce information asymmetry among investors. Clearly, complex annual reports still contain information for investors. Overall, the level of disclosure rather than the readability of annual

[16] We also test separately the impacts of the disclosure characteristics of annual reports on information asymmetry. Similar results are recorded to the regression including both disclosure characteristics.

reports determines information asymmetry. The significant increase in the length and complexity of annual reports found in previous research does not worsen the information gaps among investors.

6.3 Loss-making firms and information asymmetry

Extant research shows that the annual reports of loss-making firms are longer and less readable than those of profit-making firms (Li, 2008). In our sample, the average number of words in the annual reports of loss-making firms is 74,326 compared with 73,045 words in the annual reports of profit-making firms. The Fog Index and Bog Index are also higher than those of the annual reports issued by profit-making firms (20.1 vs. 19.9 and 83.86 vs. 81.46, respectively)[17]. Table 4.5 presents the regression results of Equation (2) and information asymmetry of loss-making firms.

[INSERT TABLE 4.5 HERE]

As shown in Panel A of Table 4.5, the coefficient of *loss* is positive and significant at 99% for both proxies of information asymmetry and for all regressions (i.e., with and without fixed effects). This result means that loss-making firms have higher information asymmetry than profit-making firms. Our result is consistent with the findings of Ertimur (2004). Panel A of Table 4.5 also shows that the length and readability of annual reports are negatively associated with information asymmetry, while, consistent with previous research, the annual reports of loss-making firms are longer and have a higher Fog Index than those of profit-making firms. Consequently, the length and readability of the annual reports of loss-making firms can compensate for the difference in information asymmetry between loss-making and profit-making firms. In other words, complex annual reports including more disclosures and less readability issued by loss-making firms are informative, which is expected to reduce the information asymmetry of loss-making firms (Bushee et al., 2018). However, such compensation is insufficient to reduce information asymmetry between profit- and loss-making firms in all cases. Overall, loss-making firms still have higher information asymmetry than profit-making firms.

Panel B of Table 4.5 presents the empirical results of Equation (2), which tests the difference in the impacts of the disclosure characteristics on information asymmetry between loss-making and profit-making firms. As displayed in Panel B, the length of annual reports is negatively associated with both proxies of information asymmetry at the 99% confidence level. In particular, the coefficient of the interaction term between the number of words in annual reports and *loss* is significantly negative with a t-value of -3.622 for the bid–ask spread and -5.026 for price impact. Such results do not change when including fixed effects. This means that the longer annual reports issued by loss-making firms improve information asymmetry more than those provided by profit-

[17] The t-values of the number of words, Fog Index, and Bog Index between loss-making and profit-making firms are -1.492, -0.77589, and -7.8126, respectively. The difference between the two groups is significant at the 99% confidence level for the Bog Index.

making firms. Holding the number of words constant, the annual reports of loss-making firms are more informative than those of profit-making firms since they reduce information asymmetry more.

Similarly, we find a negative association between the Fog Index and information asymmetry in all the regressions (i.e., with and without fixed effects). Panel B of Table 4.5 shows that the higher the Fog Index is, the lower information asymmetry is; however, the coefficient of the interaction term is positive and significant at the 99% confidence level. Hence, the readability of the annual reports issued by loss-making firms hampers information asymmetry among investors compared with the readability of those provided by profit-making firms, which is illustrated by the positive coefficient of the interaction term between the Fog Index and loss-making firms. However, the F-test for the sum of the two coefficients (Fog Index and interaction term) at zero is significant at 99%[18]. This means that investors do not care about the readability of the annual reports issued by loss-making firms. Our results are consistent with the suggestion of Bushee et al. (2018). It is difficult for loss-making firms to explain bad information with simple language; however, such a complex explanation still contains the information necessary for investors to analyze and make decisions. Therefore, the complexity of the annual reports issued by loss-making firms cannot affect information asymmetry among investors. Compared with the annual reports of profit-making firms, loss-making firms have higher information asymmetry regardless of whether their longer annual reports provide more information to investors, negatively affecting information asymmetry. However, the less readable annual reports provided by those firms cannot influence information asymmetry.

6.4 Accounting standards and information asymmetry

As mentioned above, the SEC allows firms to file their annual reports following IFRS or US-GAAP. Based on firms' declaration at the beginning of their annual reports, we sort our observations into two groups called IFRS and US-GAAP. We find that annual reports following IFRS are longer and more readable than those adopting US-GAAP (82,227 words vs. 65,276 words; 19.94 Fog Index vs. 19.97 Fog Index; 80.75 Bog Index vs. 83.34 Bog Index)[19]. Our results are consistent with the findings of Lundholm et al. (2014). In this section, we test the difference in information asymmetry between firms following IFRS and US-GAAP. Additionally, we check whether the impacts of the disclosure characteristics of annual reports on information asymmetry differ between these two groups.

[INSERT TABLE 4.6 HERE]

[18] The probability of the F-test for the sum of the coefficients of LnFog and Loss*LnFog is 0.1875 for the regression without fixed effects and 0.04 for the regression with fixed effects when using the bid–ask spread for information asymmetry. The F-tests for the sum of these coefficients when using price impact for information asymmetry are 0.8177 and 0.5154, respectively.

[19] The t-values of the differences in the number of words, Fog Index, and Bog Index between the US-GAAP and IFRS groups are -11.346, 0.25754, and 8.3576, respectively. The differences in the number of words and Bog Index are significant at 99%.

As seen in Panel A of Table 4.6, the information asymmetry of firms following IFRS is lower than that of firms adopting US-GAAP. A negative coefficient is reported for both proxies of information asymmetry and when including fixed effects in the regression (99% in three of four cases and 95% in the other case). This result infers that annual reports following IFRS are more informative than those adopting US-GAAP. The reason is that IFRS is principle-based accounting that allows firms to be sufficiently flexible to adopt accounting policies. Such flexibility is considered to better reflect the economic substances of the transactions, which improves the informativeness of accounting information and reduces information asymmetry (Chen et al., 2015). Moreover, the negative associations among the disclosure characteristics of annual reports and information asymmetry shown in Panel A of Table 4.6 support that the longer annual reports under IFRS, on average, reduce information asymmetry compared with the annual reports adopting US-GAAP.

Panel B of Table 4.6 shows the empirical results of Equation (3), which tests the effect of accounting standards on the relationships among the disclosure characteristics of annual reports and information asymmetry. In terms of the level of disclosure, Panel B of Table 4.6 shows the negative association between the length of annual reports and information asymmetry. While the coefficient of the interaction term between IFRS and the number of words is not significant for the regressions with the bid–ask spread, it is significant at the 99% confidence level for the regressions with price impact. This result shows that the effect of the length of annual reports under IFRS on information asymmetry is weaker than that of annual reports under US-GAAP because of the positive coefficient of the interaction terms[20]. However, this result is sensitive because it changes when using different proxies of information asymmetry. Hence, we cannot conclude that the level of disclosure affects information asymmetry in the group of filings following IFRS differently than the group of filings adopting US-GAAP. Ignoring the interaction terms, the increase in the length of annual reports following IFRS lowers information asymmetry; this increase could be explained by supplying an additional explanation about the material difference between US-GAAP and IFRS and more details about the accounting policies in principle-based accounting. This information is new to some extent for investors, especially unsophisticated investors, thereby reducing information asymmetry among them (Amiram et al., 2016). Similarly, we have insufficient evidence to infer that the impact of readability on information asymmetry depends on the accounting standards that firms follow since the regression results of the interaction terms are sensitive to the proxies used. As shown in Panel B of Table 6, the coefficients of readability are not significant. Hence, there is no association between the readability of annual reports and information asymmetry.

[20] The F-test for the sum of the coefficients of Lnword and the interaction term at zero is significant at 99%. This means that the length of annual reports following IFRS is still negatively associated with information asymmetry.

In summary, there is no difference in the impacts of the disclosure characteristics of annual reports on information asymmetry between the IFRS and US-GAAP groups. Noticeably, firms following IFRS have lower information asymmetry than firms adopting US-GAAP.

6.5 Robustness tests

6.5.1 Testing the associations with the Bog Index

The Fog Index has been criticized for its simple definition of complex words, with the SEC suggesting that firms use plain English in their disclosures. Some research recommends using the Bog Index to measure the readability of annual reports (Miller, 2010; Bonsall et al., 2017). Our research also uses the Bog Index to test our results. Overall, our results remain when replacing the Fog Index with the Bog Index; however, some of our results with the Bog Index are sensitive, especially when using the different proxies of information asymmetry, compared with using the Fog Index. We mention the main tests in this part; the results of the other tests are similar to using the Fog Index and are available on request.

Panel A of Table 4.7 shows the controversial results compared with Table 4.4. Using the Bog Index to run the regressions with the two proxies of information asymmetry causes different results. When replacing the Fog Index with the Bog Index in the regression, the negative association between the level of disclosure and information asymmetry is still the same and significant at 99% regardless of including fixed effects. However, the readability of annual reports measured by plain English is not associated with the bid–ask spread but is positively associated with price impact. Hence, it is too risky to draw a conclusion on the relationship between the Bog Index and information asymmetry. In conclusion, the improvement in the readability of annual reports cannot reduce information asymmetry among investors.

[INSERT TABLE 4.7 HERE]

As displayed in Panel B and Panel C of Table 4.7, the conclusions about the associations among firm performance, accounting standard adoption, and information asymmetry do not change when using the Bog Index. Loss-making firms have larger information asymmetry than profit-making firms. Conversely, firms following IFRS have lower information asymmetry than those adopting US-GAAP. Hence, the Bog Index is not associated with information asymmetry in these regressions with different proxies of information asymmetry[21].

6.5.2 Testing the associations with different windows

We also use different windows to test the results. The bid–ask spread and price impact in five- and 21-day windows are used in our analysis. Overall, the results when using the different windows show no materially significant differences. Table 4.8 presents the results of the main tests

[21] The association between the Bog Index and information asymmetry is only significant at 95% in the regressions of Equation (3) when using price impact as the proxy of information asymmetry and without fixed effects.

(the other results are available on request). Panel A displays the associations among the level of disclosure, the Fog Index, and information asymmetry. As shown in Panel A, the length of annual reports is negatively associated with information asymmetry. Conversely, the improvement in the readability of annual reports does not reduce information asymmetry, which is shown in Panel A of Table 8. Including both the disclosure characteristics into one regression with the five-day and 21-day windows, the associations between the level and transparency of annual reports do not change. Overall, our conclusions are not sensitive since we test our results with different dimensions (i.e., different proxies of readability, different measurements of information asymmetry, and different windows). In our robustness test, we only mention the result of the association between the Fog Index and information asymmetry, since the results for the Bog Index were sensitive, as shown in the previous parts of our research. However, the results of the Bog Index with different windows do not significantly differ from the results with three-day windows[22]. Panel B and Panel C of Table 4.8 show the empirical results of Equations (2) and (3) with five- and 21-day windows. Similar trends are recorded when using different windows, showing that loss-making firms have higher information asymmetry than profit-making firms. Hence, annual reports following IFRS provide more information asymmetry, leading to lower information asymmetry than annual reports adopting US-GAAP[23]. In summary, our conclusions are robust when using different measurements of readability, different proxies of information asymmetry, and different windows.

[INSERT TABLE 4.8 HERE]

7. Conclusion

Our research examines the impacts of the disclosure characteristics of annual reports and information asymmetry. The level of disclosure and transparency of annual reports are tested. We use the number of words in annual reports to measure the level of disclosure, which is more objective, less biased, and suitable for large sample sizes. In terms of the transparency of annual reports, we use the readability of annual reports measured by the Fog Index and Bog Index. We measure information asymmetry via the bid–ask spread and price impact during the windows.

Despite mixed evidence on the association between the level of disclosure and information asymmetry in previous research, our study confirms that an increase in the amount of disclosure improves information asymmetry among investors. Our research also infers that the recent increase in the length of annual reports has helped enhance the information environment of investors, leading to a decline in information asymmetry. Consistent with the suggestion of Bonsall et al. (2017), our research shows that the readability of annual reports is not associated with an improvement in information asymmetry. Complex annual reports contain more information important for investors (Bushee et al., 2018). Loss-making firms have higher information

[22] The empirical results with the Bog Index are available upon request.
[23] The interaction terms of Equations (2) and (3) remain when using different windows in the tests.

asymmetry than profit-making firms. Additionally, the impact of the level of disclosure on information asymmetry is stronger for loss-making firms than profit-making firms. Annual reports following IFRS contain more information, leading to lower information asymmetry than annual reports adopting US-GAAP; however, the impacts of the disclosure characteristics on information asymmetry do not depend on the accounting standards used.

Overall, our research shows the positive viewpoints of the current tendency of annual reports in recent years. In terms of information asymmetry, the increase in the length and complexity of annual reports does not worsen the information environment of investors, especially the former.

APPENDICES

Appendix 4.1: Variable definitions and expected signs

Variable Name	Expected Sign	Definition
LN_BAS3	N/A	Natural logarithm of the average daily bid–ask spread around the three-day window (-1;1).
LN_Price3	N/A	Natural logarithm of the average daily price impact around the three-day window (-1; 1).
LN_Words	(-)	Natural logarithm of the number of words in 20-F forms identified by the Lingua::EN:: Fathom package of the Perl language for the text files after cleaning.
LN_Fog	(+)	Natural logarithm of the Fog Index of 20-F forms identified by the Lingua::EN:: Fathom package of the Perl language for the text files after cleaning.
LN_Bog	(+)	Natural logarithm of the Bog Index of 20-F forms identified by StyleWriter Professional Edition for the text files after cleaning.
LN_Price	(-)	Natural logarithm of the average daily stock price.
LN_Turnover	(-)	Natural logarithm of the average daily turnover ratio.
Volatility	(+)	Natural logarithm of the daily standard deviation of stock returns during a year.
IFRS	(-)	An indicator variable equal to 1 if firms' annual reports adopt IFRS and 0 if firms' annual reports comply with US-GAAP.
LOSS	(+)	An indicator variable equal to 1 if firms have a positive net income and 0 otherwise.
LN_BAS5	N/A	Natural logarithm of the average daily bid–ask spread around the five-day window (-2; 2).
LN_Price5	N/A	Natural logarithm of the average daily price impact around the five-day window (-2; 2).
LN_BAS21	N/A	Natural logarithm of the average daily bid–ask spread around the 21-day window (-10; 10).
LN_Price21	N/A	Natural logarithm of the average daily price impact around the 21-day window (-10; 10).

Appendix 4.2. **Cleaning the raw text files of the 20-F forms**

The raw text files of the 20-F forms automatically acquired from EDGAR contain both text and other data; therefore, we need to extract the text to measure the length and readability of these reports from the raw text files. We follow the instructions of Bonsall et al. (2017), Li (2008), and Loughran and McDonald (2014a, 2014b) to remove any unnecessary elements from the raw text files [24] . We use the packages File::Slurp, HTML::format text and HTML::TreeBuilder; HTML::Entities;Text::Unidecode of the Perl language to parse the raw 20-F files, following these steps:

a. Remove HTML and the format design from the raw files.
b. Remove all content between the <XBRL> and </XBRL> tags.
c. Remove all tables with more than 15% numeric characters.
d. Remove all markup tags.
e. Remove other textual expressions such as newlines, underscored characters, multiple blanks, and spaces.

[24] We follow the scripts of Loughran and McDonald (2014a, 2014b) and Bonsall et al. (2017), which are available at https://sraf.nd.edu/data/stage-one-10-x-parse-data/ and https://kelley.iu.edu/bpm/index.html.

TABLE 4.1: Sample selection

We use Python to auto-collect all the 20-F filings from 2005 to 2014 from the EDGAR website. We remove all filings in the finance and insurance sectors and filings with a fiscal year not on December 31 from our sample. Since Thomson Reuters Datastream has data on the bid–ask price from June 4, 2006, we choose all filings available after that day without missing data from Thomson Reuters Datastream.

Total observations collected from EDGAR	7,588
Number of firms	1,475
Number of firms whose fiscal years end on December 31	966
Number of firms with RIC codes available in Thomson Reuters	264
Number of observations without missing data	**1,348**
In which:	
Agriculture, forestry, and fishing	2
Mining	124
Manufacturing	655
Transportation, communication, electric, gas, and sanitary services	328
Wholesale trade	22
Retail trade	12
Services	205

TABLE 4.2: Descriptive statistics

Table 4.2 presents the descriptive statistics of all the variables used in our analyses. The number of words in annual reports and Fog Index are acquired from the Lingua::EN:: Fathom package of the Perl language. The other variables are extracted from Thomson Reuters Datastream. Appendix 4.2 describes the variables.

	N	**Mean**	**Std. Dev.**	**Q1**	**Median**	**Q3**
LN_BAS3	1,348	-5.791	1.50	-6.819	-5.978	-4.966
LN_Price3	1,348	-7.315	3.10	-9.625	-7.746	-5.240
LN_Words	1,348	11.128	0.41	10.930	11.138	11.343
LN_Fog	1,348	2.988	0.10	2.925	2.974	3.040
LN_Bog	1,348	4.405	0.07	4.357	4.407	4.454
LN_Turnover	1,348	0.5325	1.51	-0.545	0.5139	1.630
LN_Price	1,348	2.703	1.14	1.950	2.720	3.390
LN_Volatility	1,348	-3.592	0.48	-3.945	-3.611	-3.236
Loss	1,348	0.267	0.44	0	0	1
IFRS	1,348	0.479	0.50	0	0	1
LN_BAS5	1,348	-5.760	1.49	-6.773	-5.996	-4.892
LN_BAS21	1,348	-5.694	1.47	-6.694	-5.938	-4.780
LN_Price5	1,348	-7.227	3.13	-9.610	-7.730	-5.117
LN_Price21	1,348	-7.165	3.14	-9.490	-7.705	-5.088

TABLE 4.3: Pearson correlations

This table shows the Pearson correlations among the variables used in our research including the bid–ask spread, price impact, number of words in 20-F forms, Fog Index, volatility, turnover, price, loss, IFRS, and the Bog Index.

	(1)	***(2)***	***(3)***	***(4)***	***(5)***	***(6)***	***(7)***	***(8)***	***(9)***	***(10)***	***(11)***	***(12)***	***(13)***	***(14)***
LN_BAS3 (1)	1													
LN_Price3 (2)	0.79	1												
LN_Word (3)	-0.18	-0.24	1											
LN_Fog (4)	0.004	0.02	-0.14	1										
LN_Turnover (5)	-0.15	-0.26	0.10	-0.08	1									
LN_Price (6)	-0.50	-0.53	0.15	0.04	0.04	1								
LN_Volatility (7)	0.48	0.45	-0.07	0.09	0.33	-0.33	1							
Loss (8)	0.35	0.35	0.04	0.02	0.07	-0.29	0.32	1						
IFRS (9)	-0.20	-0.27	0.20	-0.01	-0.24	0.26	-0.26	-0.17	1					
LN_BAS5 (10)	0.98	0.80	-0.18	0.003	-0.15	-0.51	0.48	0.36	-0.20	1				
LN_BAS21 (11)	0.94	0.80	-0.19	-0.006	-0.14	-0.51	0.47	0.37	-0.20	0.96	1			
LN_Price5 (12)	0.80	0.99	-0.24	0.03	-0.26	-0.54	0.45	0.36	-0.28	0.81	0.81	1		
LN_Price21(13)	0.80	0.98	-0.25	0.03	-0.27	-0.54	0.45	0.37	-0.28	0.81	0.81	0.99	1	
LN_Bog (14)	0.11	0.15	0.12	0.17	0.04	-0.20	0.11	0.19	-0.23	0.12	0.11	0.16	0.16	1

TABLE 4.4: Disclosure characteristics of the annual reports and information asymmetry

Table 4.4 presents the empirical results of the OLS estimation of Equation (1), which show the associations among the disclosure characteristics of annual reports and information asymmetry. The disclosure characteristics mentioned in our research are the number of words in annual reports and readability of annual reports (e.g., the Fog Index). Information asymmetry is measured using the bid–ask spread and price impact during windows.[25]

Independent Variables	**Bid-Ask Spread (LN_BAS3)**		**Price Impact (Ln_Price3)**	
	(1)	**(2)**	**(3)**	**(4)**
Intercept	6.807	5.215	22.717	23.938
	(4.886)***	(3.228)***	(8.492)***	(7.860)***
LN_Word	-0.309	-0.239	-0.997	-0.940
	(-4.088)***	(-3.110)***	(-6.875)***	(-6.494)***
LN_Fog	-0.875	-0.564	-1.710	-1.258
	(-2.721)***	(-1.737)**	(-2.769)***	(-2.056)**
LN_Turnover	-0.300	-0.327	-0.820	-0.931
	(-13.589)***	(-14.749)***	(-19.347)***	(-22.330)***
LN_Price	-0.400	-0.400	-0.878	-0.777
	(-13.613)***	(-13.045)***	(-15.579)***	(-13.492)***
LN_Volatility	1.478	1.564	3.067	3.484
	(20.066)***	(18.619)***	(21.694)***	(22.001)***
Fixed Effects (sector and year fixed effects)	No	Yes	No	Yes
Adjusted R2	0.4419	0.4778	0.5214	0.5685
Observations	1,348	1,348	1,348	1,348

[25] When including size in the regression of Equation (1), the sign and significance of the level of disclosure do not change; however, the Fog Index becomes insignificant when adding fixed effects and downgrades to a 90% confidence level without fixed effects. Our conclusions remain when including size in the regressions.

TABLE 4.5: Disclosure characteristics of loss firms and information asymmetry

Table 4.5 presents the impacts of the disclosure characteristics of loss-making firms on information asymmetry. Based on the net income of firms acquired from Thomson Reuters Datastream, we identify the dummy variable of *loss firm year*. The disclosure characteristics of annual reports are the level of disclosure and readability of annual reports.

Panel A: Loss-making firms and information asymmetry

Independent Variables	**Bid-Ask Spread (LN_BAS3)**		**Price Impact (Ln_Price3)**	
	(1)	**(2)**	**(3)**	**(4)**
Intercept	6.77433	4.92911	22.64455	23.277551
	(4.949)***	(3.102)***	(8.671)***	(7.840)***
LN_Word	-0.35580	-0.27576	-1.10198	-1.025385
	(-4.777)***	(-3.642)***	(-7.755)***	(-7.249)***
LN_Fog	-0.91973	-0.60911	-1.80954	-1.362943
	(-2.910)***	(-1.908)*	(-3.001)***	(-2.285)**
LN_Turnover	-0.29749	-0.32191	-0.81460	-0.921139
	(-13.711)***	(-14.792)***	(-19.679)***	(-22.655)***
LN_Price	-0.35853	-0.36186	-0.78641	-0.691918
	(-12.177)***	(-11.825)***	(-14.000)***	(-12.102)***
LN_Volatility	1.35501	1.41933	2.79380	3.149501
	(18.204)***	(16.643)***	(19.674)***	(19.766)***
Loss	0.51290	0.49025	1.14514	1.132782
	(7.009)***	(6.824)***	(8.203)***	(8.439)***
Fixed Effects (sector and year fixed effects)	No	Yes	No	Yes
Adjusted R2	0.4612	0.4952	0.5439	0.5902
Observations	1,348	1,348	1,348	1,348

Panel B: Disclosure characteristics of loss-making firms and information asymmetry

Independent Variables	Bid-Ask Spread (LN_BAS3) (1)	(2)	Price Impact (Ln_Price3) (3)	(4)
Intercept	6.918	5.667	20.516	21.814
	(4.501)***	(3.291)***	(7.011)***	(6.803)***
LN_Word	-0.247	-0.193	-0.799	-0.744
	(-3.048)***	(-2.370)**	(-5.173)***	(-4.906)***
LN_Fog	-1.430	-1.191	-2.353	-1.966
	(-3.986)***	(-3.303)***	(-3.444)***	(-2.928)**
LN_Turnover	-0.284	-0.310	-0.786	-0.895
	(-13.115)***	(-14.310)***	(-19.053)***	(-22.157)***
LN_Price	-0.353	-0.358	-0.775	-0.680
	(-12.078)***	(-11.803)***	(-13.927)***	(-12.023)***
LN_Volatility	1.314	1.377	2.703	3.059
	(17.686)***	(16.208)***	(19.113)***	(19.332)***
Loss	1.341	-0.179	13.329	13.022
	(0.426)	(-0.058)	(2.224)**	(2.280)**
Loss*LN_Word	-0.680	-0.600	-1.796	-1.793
	(-3.622)***	(-3.283)***	(-5.026)***	(-5.265)***
Loss*LN_Fog	2.262	2.464	2.630	2.716
	(3.134)***	(3.497)***	(1.914)*	(2.070)**
Fixed Effects (sector and year fixed effects)	No	Yes	No	Yes
Adjusted R2	0.4704	0.5039	0.5536	0.5999
Observations	1,348	1,348	1,348	1,348

TABLE 4.6: Disclosure characteristics of the annual reports following IFRS and information asymmetry

Table 4.6 presents the associations among the disclosure characteristics of annual reports adopting IFRS and information asymmetry. Based on firms' announcements, we sort the 20-F forms into two groups: IFRS and US-GAAP. The level of disclosure and readability of annual reports are the disclosure characteristics in our analysis. Information asymmetry is measured using the bid–ask spread and price impact.

Panel A: IFRS and information asymmetry

Independent Variables	**Bid-Ask Spread (LN_BAS3)**		**Price Impact (Ln_Price3)**	
	(1)	**(2)**	**(3)**	**(4)**
Intercept	6.35681	4.83171	20.16878	22.258021
	(4.532)***	(2.986)***	(7.622)***	(7.387)***
LN_Word	-0.27084	-0.20155	-0.78242	-0.776514
	(-3.521)***	(-2.588)***	(-5.392)***	(-5.355)***
LN_Fog	-0.87448	-0.56030	-1.70631	-1.243065
	(-2.724)***	(-1.730)*	(-2.817)***	(-2.061)**
LN_Turnover	-0.31251	-0.33802	-0.89087	-0.982682
	(-13.819)***	(-15.025)***	(-20.882)***	(-23.455)***
LN_Price	-0.38557	-0.39119	-0.79816	-0.742725
	(-12.908)***	(-12.745)***	(-14.164)***	(-12.994)***
LN_Volatility	1.45694	1.52775	2.95063	3.324646
	(19.695)***	(17.997)***	(21.142)***	(21.030)***
IFRS	-0.16506	-0.18860	-0.93483	-0.826849
	(-2.459)**	(-2.691)***	(-7.383)***	(-6.335)***
Fixed Effects (sector and year fixed effects)	No	Yes	No	Yes
Adjusted R2	0.4439	0.4803	0.5397	0.5808
Observations	1,348	1,348	1,348	1,348

Panel B: Disclosure characteristics of annual reports following IFRS and information asymmetry

Independent Variables	Bid-Ask Spread (LN_BAS3) (1)	(2)	Price Impact (Ln_Price3) (3)	(4)
Intercept	6.30556	3.87552	23.21354	26.45366
	(3.037)***	(1.739)*	(5.923)***	(6.377)***
LN_Word	-0.50752	-0.37839	-1.34944	-1.48759
	(-3.314)***	(-2.411)**	(-4.667)***	(-5.092)***
LN_Fog	-0.02059	-0.35039	-0.70039	-0.18295
	(-0.045)	(-0.768)	(-0.817)	(-0.215)
LN_Turnover	-0.30450	-0.33114	-0.87904	-0.97127
	(-13.399)***	(-14.670)***	(-20.487)***	(-23.112)***
LN_Price	-0.37355	-0.38183	-0.78074	-0.72294
	(-12.417)***	(-12.383)***	(-13.746)***	(-12.593)***
LN_Volatility	1.43462	1.50652	2.90451	3.26672
	(19.268)***	(17.658)***	(20.662)***	(20.567)***
IFRS	1.69101	3.25554	-3.72899	-5.16690
	(0.602)	(1.172)	(-0.703)	(-0.999)
IFRS*LN_Word	0.26155	0.16213	0.70074	0.85751
	(1.481)	(0.921)	(5.026)**	(2.615)***
IFRS*LN_Fog	-1.59178	-1.75329	-1.66764	-1.72822
	(-2.452)**	(-2.721)***	(-1.361)	(-1.441)
Fixed Effects (sector and year fixed effects)	No	Yes	No	Yes
Adjusted R2	0.4466	0.4828	0.5536	0.5831
Observations	1,348	1,348	1,348	1,348

TABLE 4.7: Robustness tests with the Bog Index

Table 4.7 shows the empirical results of Equations (1), (2), and (3) using another proxy of the readability of annual reports called the Bog Index. The Bog Index is identified by StyleWriter Professional Software for the post-cleaned 20-F forms.

Panel A: Bog Index of annual reports and information asymmetry

Independent Variables	**Bid-Ask Spread (LN_BAS3)**		**Price Impact (Ln_Price3)**	
	(1)	**(2)**	**(3)**	**(4)**
Intercept	2.66731	2.22847	5.67204	11.56026
	(1.302)	(1.027)	(1.446)	(2.831)***
LN_Word	-0.28759	-0.22668	-1.01256	-0.95676
	(-3.789)***	(-2.936)**	(-6.970)***	(-6.584)***
LN_Bog	0.27135	0.27421	2.67471	2.02885
	(0.608)	(0.612)	(3.131)***	(2.405)**
LN_Turnover	-0.29337	-0.32205	-0.80804	-0.91913
	(-13.337)***	(-14.601)***	(-19.195)***	(-22.140)***
LN_Price	-0.40495	-0.40198	-0.86191	-0.77419
	(-13.592)***	(-13.093)***	(-15.117)***	(-13.398)***
LN_Volatility	1.44848	1.54982	2.99066	3.42712
	(19.762)***	(18.388)***	(21.322)***	(21.604)***
Fixed Effects (sector and year fixed effects)	No	Yes	No	Yes
Adjusted R2	0.4389	0.4768	0.5222	0.569
Observations	1,348	1,348	1,348	1,348

Panel B: Loss-making firms and information asymmetry

Independent Variables	Bid-Ask Spread (LN_BAS3) (1)	(2)	Price Impact (Ln_Price3) (3)	(4)
Intercept	3.96056	3.164702	8.45927	13.67377
	(1.958)*	(1.480)	(2.195)**	(3.423)***
LN_Word	-0.32361	-0.251261	-1.09020	-1.01226
	(-4.326)***	(-3.305)***	(-7.652)***	(-7.128)***
LN_Bog	-0.08251	-0.066051	1.91207	1.26072
	(-0.187)	(-0.149)*	(2.273)**	(1.521)
LN_Turnover	-0.29036	-0.318037	-0.80155	-0.91006
	(-13.424)***	(-14.653)***	(-19.458)***	(-22.448)***
LN_Price	-0.36879	-0.367415	-0.78398	-0.69615
	(-12.395)***	(-11.995)***	(-13.835)***	(-12.167)***
LN_Volatility	1.32852	1.410790	19.355	3.11327
	(17.924)***	(16.514)***	(19.674)***	(19.511)***
Loss	0.51021	0.488618	1.09961	1.10304
	(6.905)***	(6.750)***	(7.814)***	(8.157)***
Fixed Effects (sector and year fixed effects)	No	Yes	No	Yes
Adjusted R2	0.4578	0.4938	0.5426	0.5893
Observations	1,348	1,348	1,348	1,348

Panel C: IFRS and information asymmetry

Independent Variables	Bid-Ask Spread (LN_BAS3)		Price Impact (Ln_Price3)	
	(1)	(2)	(3)	(4)
Intercept	3.22959	2.69615	8.71711	13.54996
	(1.568)	(1.241)	(2.245)***	(3.351)***
LN_Word	-0.24354	-0.18332	-0.77402	-0.77232
	(-3.122)***	(-2.327)**	(-5.263)***	(-5.266)***
LN_Bog	0.03009	0.06668	1.36811	1.14594
	(0.066)	(0.147)	(1.588)	(1.355)
LN_Turnover	-0.30568	-0.33412	-0.87471	-0.97045
	(-13.549)***	(-14.862)***	(-20.567)***	(-23.184)***
LN_Price	-0.39385	-0.39533	-0.80178	-0.74587
	(-13.082)***	(-12.861)***	(-14.127)***	(-13.033)***
LN_Volatility	1.43023	1.51745	2.89182	3.28942
	(19.441)***	(17.853)***	(20.851)***	(20.786)***
IFRS	-0.16420	-0.18726	-0.88927	-0.79665
	(-2.379)**	(-2.628)***	(-6.835)***	(-6.005)***
Fixed Effects (sector and year fixed effects)	No	Yes	No	Yes
Adjusted R2	0.4409	0.4791	0.5379	0.5801
Observations	1,348	1,348	1,348	1,348

TABLE 4.8: Robustness tests with different windows

Table 4.8 shows the regression results of Equations (1), (2), and (3) using the bid–ask spread and price impact in five- and 21-day windows. Other results are available on request.

Panel A: Disclosure characteristics of annual reports and information asymmetry

Independent Variables	Bid-Ask Spread		Price Impact	
	LN_BAS 5	LN_BAS21	LN_Price 5	LN_Price 21
Intercept	6.85209	7.54287	22.67130	22.59988
	(4.983)***	(5.511)***	(8.537)***	(8.559)***
LN_Word	-0.31061	-0.35254	-1.01215	-1.01920
	(-4.168)***	(-4.752)***	(-7.033)***	(-7.122)***
LN_Fog	-0.87432	-0.99735	-1.51285	-1.39876
	(-2.754)***	(-3.156)***	(-2.468)**	(-2.295)**
LN_Turnover	-0.29494	-0.28319	-0.83837	-0.84988
	(-13.534)***	(-13.054)***	(-19.919)***	(-20.309)***
LN_Price	-0.41567	-0.41664	-0.90049	-0.91190
	(-14.342)***	(-14.442)***	(-16.088)***	(-16.385)***
LN_Volatility	1.46492	1.40763	3.12740	3.15300
	(20.156)***	(19.457)***	(22.281)***	(22.592)***
Fixed Effects (sector and year fixed effects)	No	No	No	No
Adjusted R2	0.4522	0.4455	0.5359	0.5442
Observations	1,348	1,348	1,348	1,348

Panel B: Loss-making firms and information asymmetry

Independent Variables	Bid-Ask Spread		Price Impact	
	LN_BAS5	LN_BAS321	Ln_Price5	Ln_Price21
Intercept	6.81762	7.50537	22.59693	22.52116
	(5.061)***	(5.621)***	(8.732)***	(8.785)***
LN_Word	-0.36071	-0.40705	-1.12026	-1.13362
	(-4.921)***	(-5.602)***	(-7.956)***	(-8.126)***
LN_Fog	-0.92176	-1.04897	-1.61522	-1.50710
	(-2.963)***	(-3.402)***	(-2.703)***	(-2.546)**
LN_Turnover	-0.29222	-0.28024	-0.83251	-0.84367
	(-13.686)***	(-13.240)***	(-20.296)***	(-20.761)***
LN_Price	-0.37184	-0.36895	-0.80591	-0.81181
	(-12.833)***	(-12.845)***	(-14.479)***	(-14.721)***
LN_Volatility	1.33458	1.26582	2.84617	2.85534
	(18.220)***	(17.433)***	(20.227)***	(20.483)***
Loss	0.54551	0.59349	1.17701	1.24575
	(7.576)***	(8.315)***	(8.509)***	(9.090)***
Fixed Effects (sector and year fixed effects)	No	No	No	Yes
Adjusted R2	0.4743	0.4723	0.5594	0.5703
Observations	1,348	1,348	1,348	1,348

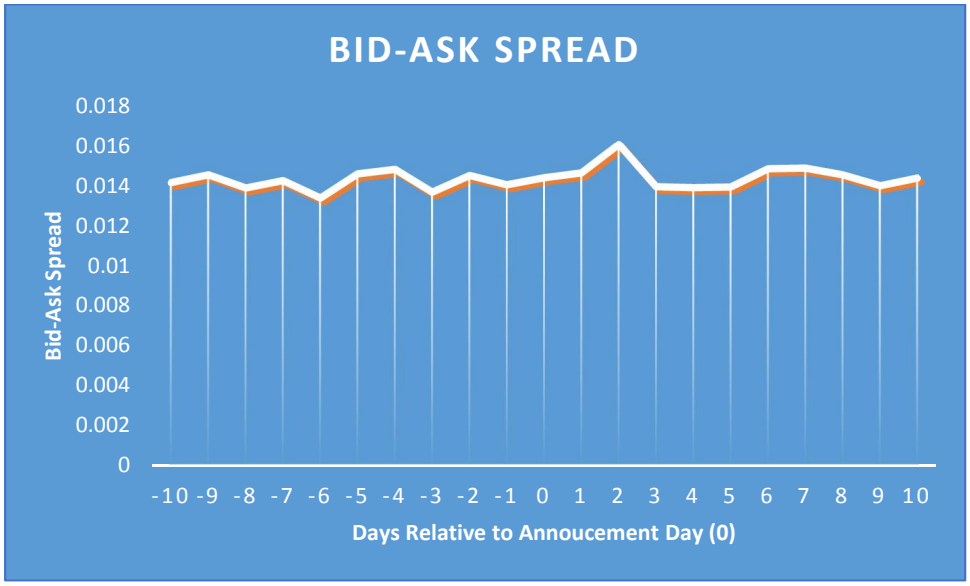
Panel C: IFRS and information asymmetry

Independent Variables	Bid-Ask Spread		Price Impact	
	LN_BAS5	LN_BAS321	Ln_Price5	Ln_Price21
Intercept	6.41003	7.11405	19.98171	19.88205
	(4.630)***	(5.162)***	(7.628)***	(7.639)***
LN_Word	-0.27341	-0.31646	-0.78587	-0.79054
	(-3.601)***	(-4.187)***	(-5.470)***	(-5.539)***
LN_Fog	-0.87369	-0.99674	-1.50901	-1.39488
	(-2.757)***	(-3.160)***	(-2.517)**	(-2.341)**
LN_Turnover	-0.30718	-0.29507	-0.91285	-0.92514
	(-13.763)***	(-13.279)***	(-21.612)***	(-22.046)***
LN_Price	-0.40175	-0.40313	-0.81576	-0.82629
	(-13.627)***	(-13.735)***	(-14.622)***	(-14.907)***
LN_Volatility	1.44466	1.38798	3.00413	3.02843
	(19.786)***	(19.095)***	(21.742)***	(22.061)***
IFRS	-0.16218	-0.15732	-0.98673	-0.99709
	(-2.448)**	(-2.386)**	(-7.871)***	(-8.006)***
Fixed Effects (sector and year fixed effects)	No	No	No	No
Adjusted R2	0.4543	0.4474	0.5561	0.5646
Observations	1,348	1,348	1,348	1,348

Figure 4.1 Information asymmetry around annual report releases

Panel A: Bid-ask spread around the announcement date

Daily bid–ask prices are extracted from Thomson Reuters Datastream. Daily bid–ask spreads are identified by (Ask-Bid)/((Ask+Bid)/2). Announcement dates are the dates on which the 20-F forms are available on the EDGAR website. Day zero is the available date of the 20-F forms. The days before the announcement date are negative dates, whereas ex-post dates are positive dates.

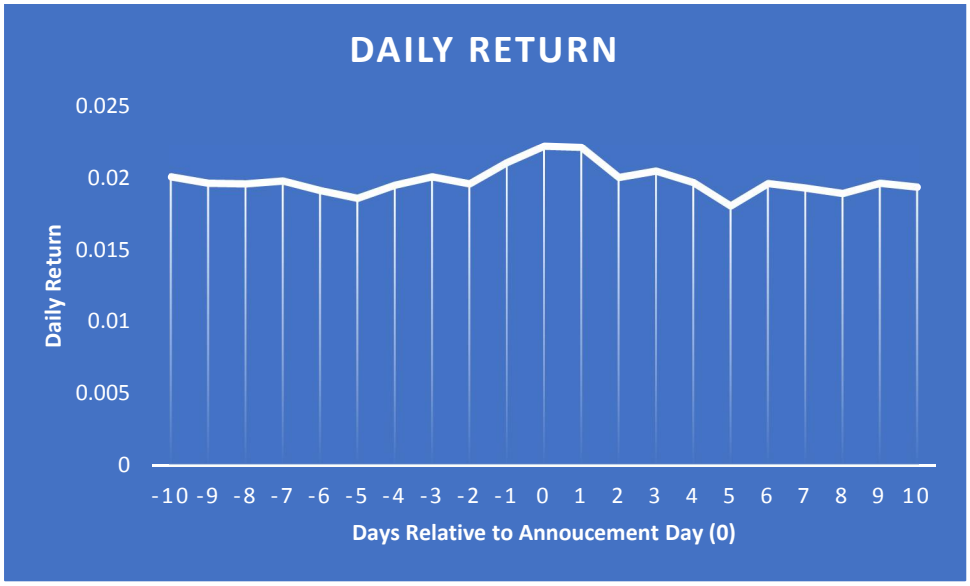


Panel B: Price impact around the announcement date

Daily stock returns, stock price, and trading volume are extracted from Thomson Reuters Datastream. Daily price impacts are identified by (daily stock return × 100)/(daily stock price × daily trading volume). Announcement dates are the dates on which the 20-F forms are available on the EDGAR website.

Panel C: Daily returns around the announcement date

Daily returns are extracted from Thomson Reuters Datastream. Announcement dates are the dates on which the 20-F forms are available on the EDGAR website.



Panel D: Daily trading volume around the announcement date

Daily trading volumes are extracted from Thomson Reuters Datastream. Announcement dates are the dates on which the 20-F forms are available on the EDGAR website.

Chapter 5: Conclusion

Annual reports are the primary information source for investors, regulators, and other users. Since 1998, the SEC's regulations have dictated that firms must write their reports in plain English and in a clear, understandable manner; nonetheless, the readability of annual reports has not yet improved substantially. Most previous research finds that annual reports have become longer and less readable. We still find a dramatic increase in the length of annual reports, but it looks more serious for annual reports issued by foreign firms listed on the US Stock Exchange. Indeed, our results suggest that foreign firms listed on the US stock market have changed their writing styles by better complying with the guidance of the SEC on plain English. They now write shorter sentences and use less jargon, which has reduced the Fog Index or Bog Index scores of their annual reports regardless of their dramatic increase in length. The adopted accounting standards and native language are also determined the length of annual reports and the usage of plain English in firms' disclosures. In terms of the information asymmetry, investors pay more attention to the amount of information than to how that information is displayed. Other words, the increase in the length of annual reports provides more information to investors which reduces the information asymmetry among the investors, meanwhile, the readability of annual reports does not improve the information asymmetry.

These findings have some contributions to academic research and practitioners. By extending to the sample set of 20-F, my finding adds one more piece to the whole picture of the significant increase in the length of annual reports occurring not only in 10-K but also in 20-F. Owing different characteristics as compared to the US firms, we realize new determinants of the number of words in 20-F. The new realized determinants help users and standard setters to deeply understand the behaviors of these firms in issuing their annual reports.

We, as the first researchers, shows the impacts of plain English on the length of annual reports. This research provides the deeply understanding of how issuers adhere to the SEC's regulations on plain English. Overall, foreign firms better comply with the SEC's plain English guidelines in their disclosures over the period from 2004 to 2013. This chapter provides empirical evidence of the types of plain English that lengthen annual reports for regulators. To reduce costs, standard setters should aim to review and modify the guidelines on plain English to improve the readability of annual reports but not significantly increase their length.

With using a different approach to measure the amount of disclosure via the number of words in annual reports, my approach thus overcomes the disadvantages of previous measurements, as it is objective, is less costly, has high availability for all users and firms, and is easy to apply to a large sample. With different approach, we argue again for the fact that the negative association between the amount of disclosure and the information asymmetry, which is still a controversial problem in both theory and empirical research. Moreover, we also test the relationship between the transparency of annual reports and information asymmetry using readability, which has not

thus far been attempted in previous research. As the first research on the associations among the disclosure characteristics of annual reports and information asymmetry, we provide optimistic views about the recent tendency of annual reports to improve in terms of information asymmetry. In addition, the adopted accounting standards is firstly recorded as the factors which enable to improve the information asymmetry.

Despite the novel contributions mentioned above, some open issues need in-depth studies in the future. Firstly, our research uses the indirect comparison with 10-K forms via prior research. It would, however, be more valuable to directly compare the annual reports of US firms. Due to the constraint of available database, the sample size is small as compared to related research. Last but not least, the time constraint leads to the incomplete story of the length of annual reports. The significant increase in the length of annual reports causes other impacts on firms and the users such as the cost of capital or cost of equity. We leave these issues to future work.

REFERENCES

Amihud, Y. (2002). Illiquidity and stock return: Cross-section and time-series effects. Journal of Financial Markets, 5(1), 31–56.

Amiram, D., Owens, E., & Rozenbaum, O. (2016). Do information releases increase or decrease information asymmetry? New evidence from analyst forecast announcements. Journal of Accounting and Economics, 62(1), 121–138.

Arnold, V., Bedard, J., Phillips, J., Sutton, S., (2010). Where do Investors Prefer to Find Nonfinancial Information? Journal of Accountancy.

Aymen, A., Mhamed, L., & Badreddine, M. (2016). Guiding through the Fog: Does annual report readability reveal earnings management? Research in International Business and Finance, 38, 509–516.

Bloomfield, R. (2008). Discussion of "annual report readability, current earnings, and earnings persistence". Journal of Accounting and Economics, 45(2–3), 248–252.

Bloomfield, R. J. (2002). The "Incomplete revelation hypothesis" and financial reporting. Accounting Horizons, 16(3), 233–243.

Bonsall IV, S. B., Leone, A. J., Miller, B. P., & Rennekamp, K. (2017). A plain English measure of financial reporting readability. Journal of Accounting and Economics, 63(2–3), 329–357.

Botosan, C. A. (1997). Disclosure level and the cost of equity capital. Accounting Review, 72(3), 323–349.

Botosan, C. A., & Plumlee, M. A. (2002). A re-examination of disclosure level and the expected cost of equity capital. Journal of Accounting Research, 40(1), 21–40.

Brown, S., & Hillegeist, S. A. (2007). How disclosure quality affects the level of information asymmetry. Review of Accounting Studies, 12(2–3), 443–477.

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

Callen, J. L., Khan, M., & Lu, H. (2013). Accounting quality, stock price delay, and future stock returns. Contemporary Accounting Research, 30(1), 269–295.

Cazier, R. A., & Pfeiffer, R. J. (2016). Why are 10-K filings so long? Accounting Horizons, 30(1), 1–21.

Chen, L. H., Deng, S., Gupta, P. P., & Sami, H. (2015). The SEC's elimination of 20-F reconciliation and information asymmetry. Journal of Contemporary Accounting & Economics, 11(1), 75–87.

Chen, S., DeFond, M. L., & Park, C. W. (2002). Voluntary disclosure of balance sheet information in quarterly earnings announcements. Journal of Accounting and Economics, 33(2), 229–251.

Cheung, E., & Lau, J. (2016). Readability of notes to the financial statements and the adoption of IFRS. Australian Accounting Review, 26(2), 162–176.

Chueng. Yin-Wong, Ng.K.Lilian. (1992). Stock Price Dynamics and Firm Size: An Empirical Investigation. The Journal of Finance, 47 (5), 1985-1997.

Chung, K. H., & Charoenwong, C. (1998). Insider trading and the bid-ask spread. Financial Review, 33(3), 1–20.

Coller, M., & Yohn, T. L. (1997). Management forecasts and information asymmetry: An examination of bid-ask spreads. Journal of Accounting Research, 35(2), 181–191.

Copeland, T., & Galai, D. (1983). Information effects on the bid-ask spread. Journal of Finance, 38(5), 1457–1469.

Courtis, J. K. (1986). An investigation into annual report readability and corporate risk-return relationships. Accounting and Business Research, 16(64), 285-294.

Cutts, M., 2013. Oxford Guide to Plain English. Oxford University Press, New York.

Daske, H., Hail, L., Leuz, C., & Verdi, R. (2008). Mandatory IFRS reporting around the world: Early evidence on the economic consequences. Journal of Accounting Research, 46(5), 1085–1142.

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

Diamond, D. W., & Verrecchia, R. E. (1991). Disclosure, liquidity, and the cost of capital. The Journal of Finance, 46(4), 1325–1359.

Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2016). The usefulness of historical accounting reports. Journal of Accounting and Economics, 61(2–3), 448–464.

Duffee.G.R. (1995). Stock Returns and Volatility: A firm-level analysis. Journal of Financial Economics, 37, 399-420.

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. Journal of Accounting and Economics, 64(2-3), 221-245.

Ernst & Young, (2014). Disclosure effectiveness: What companies can do now. http://www.ey.com/Publication/vwLUAssets/EY-disclosure-effectiveness-what-companies-can-do-now/$FILE/EY-disclosure-effectiveness-what-companies-can-do-now.pdf (accessed 30 May 2018).

Ertimur, Y. (2004). Accounting numbers and information asymmetry: Evidence from loss firms. Source: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=572284.

Filzen, J. J., & Schutte, M. G. (2017). Comovement, financial reporting complexity, and information markets: Evidence from the effect of changes in 10-Q lengths on internet search volumes and peer correlations. The North American Journal of Economics and Finance, 39, 19–37.

French, R. K., & Poterba, M. J. (1991). Investor Diversification and International Equity Markets. American Economic Review, 81(2), 222–226.

Fu, R., Kraft, A., & Zhang, H. (2012). Financial reporting frequency, information asymmetry, and the cost of equity. Journal of Accounting and Economics, 54(2–3), 132–149.

Gajewski, J. F., & Li, L. (2015). Can Internet-based disclosure reduce information asymmetry? Advances in Accounting, 31(1), 115–124.

George, T. J., Kaul, G., & Nimalendran, M. (1991). Estimation of the bid-ask spread and its components: A new approach. The Review of Financial Studies, 4(4), 623–656.

Glosten, L. R. (1987). Components of the bid-ask spread and the statistical properties of transaction prices. The Journal of Finance, 42(5), 1293–1307.

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

Gunning, R. (1952). The technique of clear writing. McGraw-Hill International Book Co., New York, NY.

Healy, P. M., Hutton, A. P., & Palepu, K. G. (1999). Stock performance and intermediation changes surrounding sustained increases in disclosure. Contemporary Accounting Research, 16(3), 485–520.

Heflin, F. L., Shaw, K. W., & Wild, J. J. (2005). Disclosure policy and market liquidity: Impact of depth quotes and order sizes. Contemporary Accounting Research, 22(4), 829–865.

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

Hsieh, C. C., Hui, K. W., & Zhang, Y. (2016). Analyst report readability and stock returns. Journal of Business Finance and Accounting, 43(1–2), 98–130.

Hu, Y., & Prigent, J. L. (2018). Information asymmetry, cluster trading, and market efficiency: Evidence from the Chinese stock market. Economic Modelling.

Kim, O., & Verrecchia, R. E. (1994). Market liquidity and volume around earnings announcements. Journal of Accounting and Economics, 17(1–2), 41–67.

Kim, Y., Li, H., & Li, S. (2012). Does eliminating the Form 20-F reconciliation from IFRS to U.S. GAAP have capital market consequences? Journal of Accounting and Economics, 53(1–2), 249–270.

KPMG and Financial Executives Research Foundation, 2012. Disclosure overload and complexity: Hidden in plain sight. https://www.scribd.com/document/335028195/Disclosure-Overload-Complexity (accessed 17 October 2017).

KPMG and Financial Executives Research Foundation. (2012). Disclosure overload and complexity: hidden in plain sight. Report.

Krinsky, I., & Lee, J. (1996). Earnings announcements and the components of the bid-ask spread. The Journal of Finance, 51(4), 1523–1535.

Lang, M., & Lundholm, R. (1993). Cross-sectional determinants of analyst ratings of corporate disclosures. Journal of Accounting Research, 31(2), 246–271.

Lang, M., & Stice-Lawrence, L. (2015). Textual analysis and international financial reporting: Large sample evidence. Journal of Accounting and Economics, 60(2–3), 110–135.

Lawrence, A. (2013). Individual investors and financial disclosure. Journal of Accounting and Economics, 56(1), 130–147

Lee, C. M. C. C. M., & Yahn, T. L. (1997). Accounting information and bid-ask spreads. Accounting Horizons, 11(4), 50–60.

Lee, Y. J. (2012). The effect of quarterly report readability on information efficiency of stock prices. Contemporary Accounting Research, 29(4), 1137–1170.

Lehavy, R., Li, F., & Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. Accounting Review, 86(3), 1087–1115.

Leuz, C., & Verrecchia, R. E. (2000). The economic consequences of increased disclosure. Journal of Accounting Research, 38, 91–124.

Levitt. A., 1997. Corporate finance in the information age. Remarks delivered to the Securities Regulation Institute, San Diego, California. https://www.sec.gov/news/speech/speecharchive/1997/spch135.txt (accessed 5 October 2017).

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

Li, F. (2010). The information content of forward-looking statements in corporate filings: A naïve Bayesian machine learning approach. Journal of Accounting Research, 48(5), 1049–1102.

Li, H. (2017). Repetitive Disclosures in the MD&A. Working Paper. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2924193. Accessed December 2017.

Lim, C. Y., Yeo, G. H. H., & Liu, C. S. (2003). Information asymmetry and accounting disclosures for joint ventures. International Journal of Accounting, 38(1), 23–39.

Lougee, B. A., & Marquardt, C. A. (2004). Earnings informativeness and strategic disclosure: An empirical examination of "pro forma" earnings. The Accounting Review, 79(3), 769–795.

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. (2014a). Measuring readability in financial disclosures. Journal of Finance, 69(4), 1643–1671.

Loughran, T., & McDonald, B. (2014b). Regulation and financial disclosure: The impact of plain English. Journal of Regulatory Economics, 45(1), 94–113.

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

Loughran, T., McDonald, B., 2011. When is a liability not a liability? J. Finan. 66(1), 35–65.

Loughran, T., McDonald, B., 2015. The use of word lists in textual analysis. J. Behav. Finan. 16(1), 1–11.

Lundholm, R. J., Rogo, R., & Zhang, J. L. (2014). Restoring the tower of babel: How foreign firms communicate with U.S. Investors. Accounting Review, 89(4), 1453–1485.

McNichols, M., & Trueman, B. (1994). Public disclosure, private information collection, and short-term trading. Journal of Accounting and Economics, 17(1–2), 69–94.

Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. The Accounting Review, 85(5), 1227–1253.

Mohd, E. (2005). Accounting for software development costs and information asymmetry. The Accounting Review, 80(4), 1211–1231.

Perez-Quiros, G., & Timmermann, A. (2000). Firm Size and Cyclical Variation in Stock Returns. The Journal of Finance, 55(3), 34

Petersen, C., & Plenborg, T. (2006). Voluntary disclosure and information asymmetry in Denmark. Journal of International Accounting, Auditing, and Taxation, 15(2), 127–149.

PwC, 2016. IFRS and GAAP: Differences and similarities. www.pwc.com/usifrs%0Awww.pwc.com/usifrs (accessed 15 October 2017).

SEC, 1969. The Wheat Report. https://www.sec.gov/about/annual_report/1969.pdf (accessed 1 May 2018).

SEC, 1998a. Staff Legal Bulletin No. 7. Plain English disclosure. https://www.sec.gov/interps/legal/slbcf7.htm (accessed 1 October 2017).

SEC, 1998b. A plain English handbook: How to create clear SEC disclosure. https://www.sec.gov/pdf/handbook.pdf (accessed 15 October 2017).

SEC, 2003. Release No. 33-8350: Commission Guidance Regarding Management's Discussion and Analysis of Financial Condition and Results of Operations. https://www.sec.gov/rules/interp/33-8350.htm (accessed 1 May 2018).

SEC, 2013. What we do. https://www.sec.gov/Article/whatwedo.html (accessed 1 October 2017).

SEC, 2014. Shaping Company Disclosure: Remarks before the George A. Leet Business Law Conference. https://www.sec.gov/news/speech/2014-spch100314kfh (accessed 10 March 2018).

SEC. Commission Guidance Regarding Management's Discussion and Analysis of Financial Condition and Results of Operation (2003). https://www.sec.gov/rules/interp/33-8350.htm Access 2 March 2018

SEC. Mandatory Disclosure Documents Telephone Survey (2008). https://www.sec.gov/pdf/disclosuredocs.pdf Access 2 March 2018

Subramanian, R., Insley, R. G., & Blackwell, R. D. (1993). Performance and readability: A comparison of annual reports of profitable and unprofitable corporations. Journal of Business Communication, 30(1), 49–61.

Topic Gateway Series. Convergence of Accounting Standards (2008). http://www.cimaglobal.com/Documents/ImportedDocuments/cid_tg_accounting_standards_feb08.pdf.pdf Accessed 10 October 2017.

Vanza, S., Wells, P., & Wright, A. (2018). Do asset impairments and the associated disclosures resolve uncertainty about future returns and reduce information asymmetry? Journal of Contemporary Accounting and Economics, 14(1), 22–40.

Verrecchia, R. E. (1983). Discretionary disclosure. Journal of Accounting and Economics, 5, 179–194.

Yoon, H., Zo, H., & Ciganek, A. P. (2011). Does XBRL adoption reduce information asymmetry? Journal of Business Research, 64(2), 157–163.

You, H., & Zhang, X. J. (2009). Financial reporting complexity and investor underreaction to 10-K information. Review of Accounting Studies, 14(4), 559–586.

Zhou, H. (2007). Auditing standards, increased accounting disclosure, and information asymmetry: Evidence from an emerging market. Journal of Accounting and Public Policy, 26(5), 584–620.