Human Readability of Disclosures in a Machine-Readable World

Andrew C. Call*
Arizona State University

Ben Wang
The Hong Kong Polytechnic University

Liwei Weng
The Hong Kong Polytechnic University

Qiang Wu
The Hong Kong Polytechnic University

January 2024

ABSTRACT

While regulators emphasize the need for machine-readable corporate disclosures, we examine how improvements in machine readability of textual and numerical information affect the human readability of these disclosures. Relative to the 2009 XBRL mandate that required a separate XBRL exhibit of financial statement numbers, the 2019 Inline XBRL (iXBRL) regulation improves the machine readability of both textual and numerical content throughout corporate filings. Utilizing the iXBRL mandate as a quasi-exogenous shock to machine readability, we observe a negative effect of machine readability on human readability. In addition, we document that following the iXBRL regulation, disclosures become less informative to retail investors, who generally have less ability to process corporate disclosures with machines and who are more reliant on human readability, and that they reduce ownership in stocks impacted by the iXBRL regulation. Further evidence suggests the reduction in human readability is driven by both opportunistic and non-opportunistic reasons. Our results are robust to a regression discontinuity design, an alternative difference-in-differences design, and alternative measures of human readability. Overall, our findings indicate that improved machine readability has implications for the human processing of disclosures.

Keywords: Machine readability; human readability; retail investor; capital market consequences
JEL Classifications: G14; G18; M41

---

*Corresponding author: Andrew C. Call, W. P. Carey School of Business, Arizona State University. Address: 300 East Lemon Street, Tempe, AZ, USA, 85287. Office: (480) 965-6632. Email: andycall@asu.edu. We thank Kathryn Brightbill, Sean Cao (discussant), Yanru Chang, Amy Chen, Jun Chen, Agnes Cheng, Gus De Franco, Ed deHaan, Devon Erickson, Tracie Frost, Zhaoyang Gu, Bing Han, Chung-Yu Hung, John (Xuefeng) Jiang, Youngtae Kim, Ningzhong Li, Yujing Ma, Stanimir Markov, Linda Myers, Jongwon Park, Lynn Rees, Tjomme Rusticus, Thomas Shohfi, Lisa Sun, Haimeng Teng, Laurence van Lent, Jasmine Wang, Jingjing Wang (discussant), Hong Xiang, Jiawen Yan, Cheng (Colin) Zeng, Harold Zhang, Shubo Zhang, Jingran Zhao, Wuyang Zhao, Vivi Zhu (discussant), Luo Zuo, conference participants at the 2024 FARS Midyear Meeting, the 2024 Hawaii Accounting Research Conference, the 2023 AAA Annual Meeting, and seminar participants at Renmin University of China, The Hong Kong Polytechnic University, and Utah State University for their helpful comments. We gratefully acknowledge financial support from our respective institutions. All errors are our own.

*It's a human plus machine world. It's not a machine-*only model.
Nor do I see it becoming a machine-only model for a long, long time.
—Philip Watson, Chief Innovation Officer at Citi Private Bank

1. Introduction

Machine readability is the ease with which machines can transform both textual and numerical disclosures into usable information. One of the most powerful ways sophisticated investors can leverage machine learning and large language models in capital markets is to take advantage of the textual and numerical information that is fed into machine-learning tools. In his keynote address at the 2018 Financial Information Management Conference, SEC Deputy Chief Economist and Deputy Director Scott Bauguess stated that for "advanced machine learning algorithms to generate unique insights, there must be structure to the information being read." He concludes that data fuels the "machine learning revolution" and that "(s)ophisticated algorithms depend on this data being of high quality and being machine readable." With the growing importance of machine readability in corporate disclosures, the Financial Data Transparency Act of 2022 mandates that, starting in June 2023, the SEC must provide semi-annual reports to Congress on both the public's and the SEC's use of machine-readable data in corporate disclosures (U.S. House of Representatives 2022). Regulators in other countries have also prioritized machine readability of disclosures. For example, on April 29, 2022, the European Securities and Markets Authority issued a proposal requiring disclosures to be machine readable.[1] Additionally, many think tanks now include disclosure modernization in their policy recommendations (Ritz 2020).

In spite of the increased emphasis on machine readability, human readability—the ease with which humans can comprehend written text (Blankespoor, deHaan, and Marinovic 2020)—

[1] See the European Securities and Markets Authority press release at https://www.esma.europa.eu/press-news/esma-news/esma-makes-recommendations-improve-investor-protection.

remains an important component of disclosure processing. The lack of human readability imposes information processing costs on users and delays market reactions to disclosures (Blankespoor et al. 2020). Further, disclosures that are more readily processed by humans are associated with capital market benefits and increases in firm value (Hwang and Kim 2017). Anecdotal evidence from investors and regulators suggests that, even in the AI era, machines cannot replace humans, consistent with the continued importance of human readability of machine-readable disclosures (Bauguess 2017; Egan 2019). For example, an ESG fund manager may use machines to extract ESG-related content from annual reports and then read the extracted ESG information before making investment decisions.

However, the characteristics of corporate disclosures that enhance machine readability do not necessarily facilitate human readability, and vice versa. Therefore, as regulators look to require increasing degrees of machine readability (U.S. House of Representatives 2022; SEC 2023), we explore two related research questions. First, we examine the extent to which machine readability affects human readability. Second, we examine the consequences for retail investors of this potential effect of machine readability on human readability.

We explore these two issues in the context of a recent regulation that creates a significant improvement in the machine readability of textual and numerical content in public firms' entire financial reports. On June 28, 2018, the SEC adopted a new regulation, Inline XBRL Filing of Tagged Data, requiring the use of the Inline eXtensible Business Reporting Language (Inline XBRL or iXBRL) format for the submission of annual and quarterly financial reports (i.e., 10-K and 10-Q filings).[2,3] Specifically, iXBRL requires firms to submit filings using eXtensible HyperText Markup Language (XHTML) and to embed XBRL tags directly into the filings, which

[2] See the introduction of Inline XBRL on the SEC website: https://www.sec.gov/structureddata/osd-inline-xbrl.html.
[3] Some other filings (e.g., Form S-3) are also affected by this regulation, but we focus on 10-K and 10-Q filings.

make it easier for machines to extract and process both textual and numerical content from the entire disclosure, not just the numerical data within the financial statement section of the disclosure. While the XBRL mandate in 2009 made numbers and footnotes in financial statements more machine-readable (Blankespoor 2019), it did not result in an increase in machine readability throughout the entire filing (e.g., Allee, Deangelis, and Moon 2018).[4] In addition to facilitating machine readability throughout the entire disclosure, the iXBRL mandate provides machine users with more context for machine-readable content and also mitigates errors in reading numbers. Thus, this new iXBRL mandate leads to a significant increase in machine readability of both textual and numerical financial disclosures in an era of increased reliance on artificial intelligence, large language models, and machine learning (Bauguess 2017; Bauguess 2018). Therefore, the iXBRL mandate is an ideal setting for evaluating the effect of increased machine readability of textual and numerical information on human readability of the same information.

This regulation phases in implementation over three separate dates, each separated by a one-year interval. Our sample encompasses one year before and one year after the first phase-in date, and the firms included in our treatment group are those in this first phase-in group. We find that relative to control firms, the human readability of periodic reports filed by treatment firms decreases after the implementation of the regulation that increases their disclosures' machine readability. This finding suggests that machine readability reduces human readability of corporate disclosures. Compared to control firms, human readability of disclosures by treatment firms decreases by an average of 3.14% after the passage of the iXBRL mandate, which is equivalent to a 20 percentile drop in human readability for the median firm. We further confirm that this effect

[4] The XBRL exhibit is a stand-alone, machine-readable file that contains only the numbers and footnotes to the financial statements, leaving the main filings unaffected by the XBRL mandate. As an example, an ESG fund manager using a machine to extract ESG information from a 10-K filing cannot rely on XBRL to help extract this information and, therefore, experienced no improvement in the machine readability due to XBRL.

is not driven by any differences in pre-treatment trends between treatment and control firms in our parallel trend analysis. These results are also robust to different measures of human readability, a regression discontinuity design (RDD), an alternative DiD approach, and alternative samples of periodic reports (e.g.,10-Qs or 10-Ks).[5] These findings indicate that improved machine readability leads to a reduction in the human readability of corporate disclosures.

We also explore the consequences of this effect for retail investors. Because retail investors are less likely to benefit from enhancements to machine readability compared to sophisticated institutional investors, reductions in human readability have the potential to adversely affect retail investors who rely more on manual reading of disclosures (Blankespoor, Miller, and White 2014). We find that compared to disclosures for control firms, the disclosures issued by treatment firms become less informative to retail investors after improvements to machine readability. We also find that machine readability discourages retail investing. For example, following the iXBRL adoption, treatment firms experience a 0.7% decrease in retail investor ownership. These results suggest retail investors are negatively impacted by the effect of machine readability on human readability after the iXBRL regulation.

Next, we explore whether this effect is opportunistic, non-opportunistic, or both. Extant literature on disclosure suggests firms try to manage the readability of disclosures due to proprietary costs of disclosures (e.g., Verrecchia 1983; Hayes and Lundholm 1996). We predict that a regulatory requirement that improves the machine readability of disclosures has more influence on human readability among firms with higher proprietary costs of disclosures. For example, after the iXBRL mandate, competitors may find it easier to use machines to extract

[5] It is worth noting that firms may voluntarily adopt iXBRL prior to the iXBRL mandate or before their mandatory compliance date. This voluntary adoption decision is endogenous and can bias our inferences. Therefore, we identify and remove voluntary adopters from our sample to ensure that our treatment firms are subject to an exogenous and mandatory increase in machine readability.

patent- or product-related information from peer firms' 10-K/10-Q filings before manually reading machine-extracted information, which in turn encourages firms with higher proprietary costs of disclosures to further reduce human readability. Following prior research, we identify the existence of proprietary costs using the existence of a Form CT Order (Verrecchia and Weber 2006; Thompson, Urcan, and Yoon 2023), which provides confidential treatment for certain disclosures, and find that the effect is stronger for firms with higher proprietary costs than for firms with lower proprietary costs, suggesting that in response to the mandatory increase in machine readability, managers reduce human readability of disclosures due to proprietary cost considerations.

In addition, prior studies document that firms have greater incentives to make disclosures less readable when performance is poor or when they engage in earnings management (Bloomfield 2002; Li 2008; Lo, Ramos, and Rogo 2017). Consistent with this notion, we find that the reduction in human readability following regulatory increases in machine readability is more pronounced among firms with poor performance and for firms with higher earnings management. Further, we examine whether this effect is more pronounced when managers have greater opportunity to reduce human readability. We find greater reductions in human readability among firms subject to weaker monitoring, suggesting that at least a portion of the effect that we document is opportunistic.

We also perform two sets of cross-sectional tests to examine whether there is a non-opportunistic element to this effect. Prior studies have shown that managers face resource constraints when preparing financial reports (e.g., Doyle, Ge, and McVay 2007; Li, Ye, Zeng, and Zhang 2023). Given that improving the machine readability of disclosures requires additional effort and resources (e.g., more accounting or computer staff), we expect that firms facing more resource constraints experience greater reductions in human readability as they improve their machine readability. Using detailed job posting information prior to iXBRL compliance, we find that treatment firms that fail to hire additional accounting or computer staff experience a stronger effect, consistent with a non-opportunistic mechanism. In addition, we find that the effect is stronger among complex firms that need more resources to comply with the increased machine readability requirement.

This paper contributes to the literature in several ways. First, we contribute to the disclosure literature by documenting that improvements in machine readability affect the human readability of corporate disclosures. Prior literature focuses on the determinants and consequences of human readability without considering machine readability or while holding it constant. This paper extends the literature by investigating a new and underexplored characteristic of disclosures—machine readability— and exploring its effects on human readability. Our findings suggest that neither machine readability nor human readability should be considered in a vacuum, and that enhancements in machine readability have consequences for the ability of humans to process the same disclosures. These findings have policy implications in the U.S. and other countries considering similar regulations on the machine readability of disclosures.

Second, our study contributes to the growing literature on retail investors (e.g., Laudenbach, Loos, Pirschel, and Wohlfart 2021; Barber, Huang, Odean, and Schwarz 2022; Farrell, Green, Jame, and Markov 2022; Welch 2022). Retail investors play an important role in the equity market (Brav, Cain, and Zytnick 2022) and their welfare is of unique interest to regulators (Clayton 2017; Driscoll 2019). Our study provides evidence that improvements in machine readability lead to reductions in human readability, which negatively impact retail investors. These consequences should be of interest to regulators and capital market participants alike.

Finally, our paper extends the literature on disclosure technologies. Modern disclosure technologies impact how firms disclose information and, thus, how investors obtain and process information. As Blankespoor et al. (2020) note, these "technologies are changing how both investors and researchers access qualitative disclosures, which permits research into previously inaccessible hypotheses as well as new hypotheses relevant to modern markets." Our paper provides timely evidence on a new disclosure technology, iXBRL, and supplements research on an earlier disclosure technology, XBRL (Blankespoor et al. 2014; Dong, Li, Lin, and Ni 2016; Bhattacharya, Cho, and Kim 2018; Blankespoor 2019; Kim, Kim, and Lim 2019; Li, Zhu, and Zuo 2021; Guo and Yu 2022).

2. Institutional Background and Hypothesis Development

2.1. Institutional Background

The machine readability of SEC filings has been impacted by two major regulations, XBRL and, more recently, iXBRL. Before the iXBRL regulation, public firms in the U.S. were required to file a separate, XBRL-formatted exhibit that included numbers and footnotes within financial statements, in addition to the main filings in HTML format. As a result, machines utilized two methods to "read" financial reports. For key numbers in financial statements, machines relied upon the XBRL exhibit due to its provision of well-formatted data. The machine automatically converts this exhibit into a tabular dataset, with XBRL tags serving as variable names and their numbers as variable values. Therefore, ensuring the accuracy of numbers, as well as the comparability and interpretability of tag names, has been crucial for the machine readability of numerical content.

In addition, in the XBRL environment, machines extracted text from the main filing in HTML format, which required HTML markups (i.e., tags) to identify the sections, paragraphs, and tables that contain the desired text. If the HTML markups contain errors, machines may misinterpret the textual content, leading to incorrect extraction of words in advance of human consumption and analysis. Hence, while XBRL improved the machine readability of financial statement content (Blankespoor 2019), iXBRL is intended to improve the machine readability of disclosures to minimize these errors and improve the ability of machines to extract information in a context that is useful to users.

The iXBRL format improves the machine readability of 10-K/Q filings in three major ways. First, iXBRL allows machines to more accurately and effectively read textual content in main filings. After iXBRL adoption, firms must transform their HTML-based filings into XHTML format (Basoglu and White 2015). This XHTML format is stricter and more standardized than HTML, which ensures a more predictable document structure and simplifies machine parsing, reducing errors and ambiguity during the parsing process.[6] For example, XHTML enforces rules such as proper nesting and closing of all tags, which reduces errors in identifying the sections, paragraphs, and tables that contain the desired text (see Online Appendix A). Furthermore, XHTML is less likely than HTML to silently ignore errors, increasing the likelihood that preparers correct formatting errors before disclosures are even submitted.[7] The low tolerance for errors in XHTML results in more consistent filings, further benefiting machine readability of textual content.

Second, the iXBRL filings provide users with better context for machine-readable content. One major criticism of XBRL is that its exhibits provide only isolated data items whose tags are chosen at firms' discretion, which undermines comparability across firms and within firms in different periods. For example, in XBRL exhibits, public firms create many different custom tags for basically the same data item, and sometimes individual firms employ different tags for the

[6] Because of the strictness of the XHTML format, the SEC requires firms to clean up tags when adopting iXBRL. The detailed requirements are documented in the Section 5.2.5 of Volume II in the EDGAR Filer Manual, which is available at https://www.sec.gov/files/edgar/filermanual/efmvol2-c5.pdf.
[7] For a detailed comparison between HTML and XHTML, see https://www.w3schools.com/html/html_xhtml.asp.

same item across years.[8] Investors had to reconcile the items with the descriptions in HTML filings before using them (Basoglu and White 2015). In contrast, the tags applicable for machine consumption with iXBRL are surrounded by the supporting XHTML tags and context (see Online Appendix A). Therefore, machine readers can more readily utilize the information from custom tags, allowing users to quickly and accurately process the context for the machine-readable content.

Third, iXBRL mitigates errors when reading numbers in disclosures. In particular, the iXBRL filing mitigates discrepancies between the filing itself and the separate XBRL exhibit by creating a single, consistent file. Prior to iXBRL, data cells that were included in both the main filing and XBRL exhibit were displayed separately, which increased the risk of typos, mismatched numbers, and calculation errors (Harris and Morsfield 2012).[9] Online Appendix A shows an error where -171,099 was reported in the XBRL exhibit while 171,099 was reported in the main filing. The iXBRL adoption mitigates potential discrepancies between the main filing and XBRL exhibit. Solving this inconsistency helps improve the machine readability of 10-K/Q filings. Many institutional investors and practitioners (e.g., ACI, AICPA, CFA Institute, IRIS, Lewis, members of Congress, Merrill, Morningstar, XBRL International, and XBRL US) have commented on these discrepancies between HTML and XBRL data and called for the increased machine readability that iXBRL provides (SEC 2018).

In sum, relative to XBRL, iXBRL provides an incremental improvement in the machine readability of both textual and numerical information throughout quarterly and annual reports.

---

[8] Amit Varshney, the director of Equity Research at Credit Suisse, complained that the XBRL "tagging is inconsistent, for example, you might have five different companies use five different tags for the exact same data or the same company using different tags for the same item over multiple periods" (Harris and Morsfield 2012). For more details, please see https://www.sec.gov/structureddata/reportspubs/osd_assessment_custom-axis-tags.

[9] In a letter to the SEC, directors at Morningstar stated that prior to iXBRL, "XBRL and HTML filings conflict with each other" and that iXBRL "will embed the XBRL tag within the HTML document, and this will greatly improve our analysts' capacities to identify any discrepancies between the XBRL tag and the HTML document quickly and efficiently, helping us to quickly provide higher-quality data to investors." https://www.sec.gov/comments/s7-03-17/s70317-1754317-151974.pdf.

Hoitash, Hoitash, and Morris (2021) and Li et al. (2021) share views that "future research may examine the impending Inline XBRL (iXBRL) mandate" and that more "research is needed to explore the implications of the recent advances of iXBRL," respectively.

2.2. Hypothesis Development

Disclosure readability consists of a human and a machine component. Human readability has been extensively studied, with prior studies indicating that it has both benefits and costs. On the one hand, firms benefit from human-readable disclosures, which can improve corporate investment efficiency (Biddle, Hilary, and Verdi 2009), improve analyst earnings forecast accuracy (Lehavy, Li, and Merkley 2011), benefit retail investors (Lawrence 2013), and reduce stock price crash risk (Kim, Wang, and Zhang 2019). On the other hand, making disclosures more human-readable has proprietary and other costs (e.g., Verrecchia 1983; Hayes and Lundholm 1996). Allee, Do, and Sterin (2021) find that firms facing high product market competition are motivated to lower the human readability of their financial statements to reduce proprietary costs. Similarly, Frankel, Lee, and Lemayian (2018) find that firms with high proprietary costs have less readable 10-Ks.

Given that machine readability is a relatively new topic, studies on machine readability are more limited. Allee et al. (2018) find that machine readability increases the speed of the market response to 10-K and 10-Q filings, confirming that some market participants use machines to read annual and quarterly reports. Recent studies show that investors' use of machines to read corporate filings improves market efficiency (Barbopoulos, Dai, Putniņš, and Saunders 2023) and is positively associated with machine readability (Cao, Jiang, Yang, and Zhang 2023). However, given the increased role of machine readability amid the ongoing need for human readability (particularly among retail investors), we study the relationship between machine readability and human readability of disclosure.

We anticipate a negative effect of machine readability on human readability. In particular, we predict that exogenous increases in machine readability lead to corresponding reductions in human readability through two mechanisms: *opportunistic* and *non-opportunistic* mechanisms. Regarding the opportunistic mechanism, prior studies have documented both the benefits and costs of disclosures. For example, disclosures help users more easily process and interpret financial information, but for a variety of reasons, managers often have incentives to opportunistically obfuscate information in disclosures (e.g., Li 2008; Loughran and McDonald 2014; Ertugrul, Lei, Qiu, and Wan 2017; Lo et al. 2017). We predict that mandated increases to machine readability encourage certain firms to opportunistically reduce the human readability of their disclosures.

First, the prior literature on proprietary costs of disclosure shows that because of the potential for competitive disadvantages resulting from disclosures to competitors, firms tend to reduce disclosure to mitigate proprietary costs (e.g., Verrecchia 1983; Verrecchia and Weber 2006; Dedman and Lennox 2009; Ellis, Fee, and Thomas 2012; Bernard 2016; Bernard, Burgstahler, and Kaya 2018; Liang 2023). Therefore, we predict that firms reduce human readability of their disclosures in the face of mandated improvements in machine readability.

Second, according to the Incomplete Revelation Hypothesis in Bloomfield (2002), firms reduce the readability of disclosures to hide bad news, such as poor performance (Li 2008), and to obfuscate earnings management (Lo et al. 2017). For instance, a firm may try to maintain its current stock price in the face of bad news by making it harder to decipher information in its annual report (Li and Zhang 2015). This strategic disclosure behavior occurs not only in public firms but also in mutual funds. For example, deHaan, Song, Xie, and Zhu (2021) find that low-performing mutual funds have hard-to-read disclosures. Firms also use financial reports and earnings management as complementary tools to hide bad news, particularly when they are subject to higher monitoring (Kim, Li, and Liu 2019). Therefore, we predict that firms are more likely to respond to mandatory improvements in machine readability by obfuscating information via reductions in human readability when their performance is poor and when they are managing earnings.

Third, we argue that managers are more likely to obfuscate human readability when they have more opportunity to do so. Specifically, we argue that firms subject to relatively weak board monitoring have more opportunities to obfuscate disclosures following the iXBRL mandate. In sum, we argue that firms with incentives and opportunities to obfuscate financial information are more likely to respond to the iXBRL mandate by opportunistically reducing the human readability of their disclosures.

We also argue that a non-opportunistic mechanism explains reductions in human readability following improvements in machine readability. The prior literature indicates that the effort, time, and resources available to managers when producing financial reporting are limited (e.g., Doyle et al. 2007; Li et al. 2023). Specifically, Doyle et al. (2007) find that firms have financial reporting issues in the face of a lack of resources and when operations are complex; Li et al. (2023) document an unintentional mechanism through which managers' heavy financial reporting workload results in lower financial reporting quality. Given these constraints, managers' effort to increase machine readability would result in less effort on human readability when the firm is more complex.

In addition, regulatory changes in disclosure requirements (e.g., Sarbanes–Oxley Act; IFRS adoption; XBRL adoption) often impose substantial transition costs, compliance costs, and recurring costs for firms (Linck, Netter, and Yang 2009; Hail, Leuz, and Wysocki 2010; Jamal et al. 2010; De George, Ferguson, and Spear 2013; Hostak, Lys, Yang, and Carr 2013; Li et al. 2021; Li et al. 2023). The SEC acknowledges that the implementation and ongoing compliance with the iXBRL mandate impose costs on filers, including increased time and attention from managers (SEC 2018). Furthermore, the transition to iXBRL forces managers to consider additional training, procedures, risk management, monitoring, and internal controls over financial reporting (PwC 2016; Deloitte 2018). These added burdens increase managers' workload and potentially distract them from efforts that would otherwise result in disclosures that are more readable to humans. Based on both of these mechanisms, we make the following prediction:

> ***Hypothesis 1:*** *The human readability of disclosures is reduced when the machine readability of disclosures increases.*

This negative effect of machine readability on human readability may affect capital market participants differently. Investors have heterogeneous abilities to digest information with the assistance of machines and, therefore, heterogenous demand for human readability (Kalay 2015). Blankespoor et al. (2014) argue that larger investors have superior processing capabilities to use XBRL for informational gains, which puts small investors at an information disadvantage. In terms of iXBRL, retail investors face a similar disadvantage because sophisticated investors are more likely to use machines to obtain and analyze information (e.g., Allee et al. 2018) and, thus, are more likely to benefit from increases in machine readability. As a result, when machine readability is improved, retail investors' information disadvantage becomes larger, weakening their relative preference for firms whose disclosures exhibit superior machine readability (Kalay 2015). In addition, a decrease in human readability requires more cognitive processing costs, and retail investors face more boundaries in processing capacity and resources.

We expect the negative effect of increased machine readability on human readability to affect the informativeness of disclosures to retail investors. Lee (2012) argues that investors respond slowly to 10-Q filings that are low in human readability. Although machine readability is associated with the speed of the market response to SEC filings (Allee et al. 2018), we posit that retail investors are less likely to rely on machines to process corporate disclosures and realize these benefits, all while processing disclosures that exhibit less human readability. Therefore, we predict that the informativeness of disclosures to retail investors decreases following improvements in machine readability.

> ***Hypothesis 2:*** *The informativeness of corporate disclosures to retail investors is reduced when the machine readability of these disclosures increases.*

We note that Luo et al. (2023) examine *voluntary* adopters of iXBRL, prior to the 2019 mandate (i.e., 2016-2018) and find that voluntary iXBRL adoption improves the usefulness of the information disclosures. However, their findings with respect to voluntary adopters do not generalize to mandatory adopters due to the self-selection inherent in the decision to voluntarily adopt. More fundamentally, we focus on the potential effect of machine readability on human readability. In addition, while Luo et al. (2023) conclude that iXBRL helps level the informational playing field between sophisticated and unsophisticated investors, we examine whether iXBRL actually harms retail investors and drives them away from certain stocks.

3. Methodology and Data

3.1. Inline XBRL Regulation

To examine the effects of improvements in the machine readability of corporate disclosures, we use the Inline XBRL Filing of Tagged Data regulation, which was adopted on June 28, 2018, and created an exogenous increase in the machine readability of textual and numerical information in annual and quarterly financial reports (i.e., 10-K and 10-Q filings). The regulation mandates the use of the iXBRL format, which (1) changes the filing format from HTML to XHTML, and (2) embeds machine-readable XBRL tags directly into XHTML documents. This single-document approach in XHTML, embedded with XBRL tags, makes quarterly and annual financial reports easier for machine readers to extract textual and numerical information, with minimal data cleaning and restructuring efforts. In this way, the SEC's iXBRL regulation is a positive shock to the machine readability of financial reports.

The regulation phased in compliance for firms based on their filer category, which is primarily determined by public float or the market value of shares of common equity held by non-affiliates. U.S. filers were phased in over a three-year period. Large accelerated filers (i.e., large firms with an aggregate worldwide public float of at least $700 million, hereafter "large firms"), had to comply no later than the fiscal period ending on June 15, 2019. Accelerated filers (i.e., firms with an aggregate worldwide public float between $75 million and $700 million, hereafter "small firms"), had to comply by June 15, 2020. All other firms that are required to file periodic reports with the SEC, including foreign filers and firms with public float of less than $75 million, had to comply by June 15, 2021. We label these firms as "other firms." See Figure 1 Panel A for a timeline of compliance dates.

We focus on the first compliance date, June 15, 2019, to avoid issues with learning effects if subsequent adopters change their behavior based on the activity of first adopters (Blankespoor 2019). Our sample includes 10-Q and 10-K filings during the fiscal period from June 2018 to May 2020, before the second compliance date (June 15, 2020). Figure 1 Panel B shows our research design. Because we focus on June 15, 2019 as the event date, the treatment group includes large firms, and the control group includes both small firms and other firms.

We exclude firms that voluntarily adopt iXBRL by extracting iXBRL tags in 10-K/Q filings from the EDGAR system and identifying large firms whose filings contain iXBRL tags before June 15, 2019. We identify 684 voluntary adopters that issue 5,102 10-K/Q filings in our sample period, or 29.75% of the total 10-K/Q filings. We exclude these voluntary adopters in order to focus on the firms impacted by the mandatory adoption of iXBRL. Nevertheless, we note that our results are robust to an alternative DiD design in which the treatment firms are large firms that do not voluntarily adopt iXBRL and the control firms are large firms that voluntarily adopt iXBRL.

3.2. Variables and Regressions

To examine the effect of machine readability on human readability, we estimate the following difference-in-differences (DiD) regression model using the iXBRL regulation:

$$Fog\_Index = \alpha + \beta Treat \times Post + \theta Controls + Firm\ FE + Year\_Quarter\ FE + \varepsilon. \quad (1)$$

Following prior literature, we use the Gunning Fog index to measure the human readability of quarterly and annual reports (*Fog_Index*).[10] *Treat* equals one if the firm is a large accelerated filer (i.e., public float ≥ $700 million) and zero otherwise. *Post* equals one if the quarter is after June 15, 2019, and zero otherwise. *Controls* is a vector that includes a set of variables associated with human readability. We follow Li (2008) and Lo et al. (2017) and include earnings (*Earnings*), a loss indicator (*Loss*), firm size (*Size*), the market-to-book ratio (*MTB*), firm age (*Age*), special items (*Special_Items*), return volatility (*Ret_Vol*), earnings volatility (*Earn_Vol*), the number of business segments (*NBSeg*), the number of geographic segments (*NGSeg*), the number of Compustat items (*Nitems*), a seasoned equity offering indicator (*SEO*), an M&A indicator (*MA*), and a Delaware incorporation indicator (*Delaware*) as control variables. All continuous variables are winsorized at the 1st and 99th percentiles. Appendix A contains the variable definitions.

Because the DiD design does not require the treatment and control groups to be similar, in addition to controlling for firm size, we also control for firm fixed effects. The firm fixed effects also control for any time-invariant effects. These controls are important given that the iXBRL

[10] Our results are robust to alternative human readability measures. Please see Table 11 for details.

mandate that went into effect on June 15, 2019 applies only to larger firms. Nevertheless, as we explain below, we also use an alternative research design where both the treatment group and control group include only large firms, and our inferences are similar, suggesting that our results are not driven by differences in firm size.

The unit of our analysis is the firm-fiscal quarter. We examine both quarterly and annual financial reports, namely 10-Q and 10-K filings. We include only firms whose fiscal year ends in December because the readability of quarterly reports may differ from that of annual reports and using firms with a December fiscal year-end mitigates the concern that differences across the readability of annual and quarterly filings lead to spurious results. To mitigate this readability difference across time in a given firm, we include year-quarter fixed effects. In our sample period, 74.25% of firm-quarters have fiscal years ending in December. Our results remain similar if we do not have this December fiscal year-end requirement. Furthermore, our results hold if we include only 10-K or only 10-Q filings.

The main effects of *Treat* and *Post* are subsumed by firm fixed effects and year-quarter fixed effects. Therefore, we report only the coefficient on *Treat*×*Post*. We correct the standard errors by clustering at the firm level. Our inferences are similar if we cluster standard errors by industry.

3.3. Data, Sample, and Summary Statistics

We obtain 10-K/Q filings and firms' public float values from EDGAR.[11] Financial statement information is from Compustat. Segment information is from the segment database in Compustat. Merger and acquisition and seasoned equity offering information is from Thomson/Refinitiv's Securities Data Company (SDC).

[11] Specifically, we search the following regular expression in 10-K/Q filings to get the public float: r'\<dei\:entitypublicfloat.*?dei\:entitypublicfloat'.

We begin identifying our sample by selecting all firm-quarter observations in Compustat with valid PERMNO-GVKEY-CIK identifiers from the second quarter of 2018 to the first quarter of 2020. We delete observations without sufficient data for calculating variables in our main analysis, and we delete firms that change their filer categories during the sample period (e.g., from large to small) because we cannot know exactly when they were mandated to adopt iXBRL. In addition, we delete firms whose fiscal year-end is not in December, as well as firms that voluntarily adopt iXBRL (i.e., large firms that adopt iXBRL before June 15, 2019, and any other firms that adopt before June 15, 2020). Finally, we delete observations that are dropped from regressions due to the inclusion of fixed effects in our regressions. Appendix B shows the sample selection process.

Our final sample contains 16,399 firm-quarter observations, including 8,189 in the pre-treatment period and 8,210 in the post-treatment period. Table 1 presents the summary statistics. About 45% of firms in our sample are treatment firms. On average, the Fog index score for quarterly and annual reports is 21.33, and the standard deviation is 1.65. The mean of the Fog index is a little higher than the 18.23 reported in Li (2008) and 18.02 in Lo et al. (2017), likely because we delete sentences with fewer than five words (e.g., titles, incomplete contents in tables, and meaningless numbers caused by scraping algorithms). The mean of *Size* is 6.409, and the mean of *MTB* is 2.123. About 35% of firm-quarter observations have negative earnings. These values are comparable to those reported in Lo et al. (2017).

3.4. Validation Analysis

We first validate whether the machine readability of financial reports increases following the iXBRL mandate. To measure machine readability, we use the gap in the number of words (*Gap*) computed by two different machine users. Specifically, *Gap* is measured as the absolute difference in the number of words that are machine-read from a filing between two different sources divided by the sum of the number of words from the two sources, as shown below:

$$Machine\ Readability = GAP = \frac{|Number\ of\ Words_{Source\ 1} - Number\ of\ Words_{Source\ 2}|}{Number\ of\ Words_{Source\ 1} + Number\ of\ Words_{Source\ 2}} \tag{2}$$

The intuition behind this measure is that if a filing is more machine readable, scraping algorithms developed by independent programmers will yield more similar outputs.[12] We obtain three sources that provide the number of words in disclosures: Bill McDonald's data, SEC Analytics Suite by WRDS, and our self-developed algorithm. Each of these databases is based on the extraction procedure developed by Bill McDonald (Loughran and McDonald 2011; Loughran and McDonald 2014).[13,14] We compute three machine readability measures: (i) the difference between the WRDS and Bill McDonald sources (*Gap1*), (ii) the difference between Bill McDonald's and our own calculation (*Gap2*), and (iii) the difference between WRDS and our own calculation (*Gap3*). The mean values of *Gap1*, *Gap2*, and *Gap3* are 0.147, 0.017, and 0.163, respectively.

We report our validation results in Table 2. The estimate in Column (1) suggests the absolute difference in the numbers of words between the WRDS database and the McDonald database (*Gap1*) decreases among treatment firms following the iXBRL mandate. This effect is statistically significant at the 1% level (*t*-statistic = -17.07). We rerun the regressions using *Gap2* and *Gap3* in Column (2) and Column (3), respectively, and find that both estimates also yield

---

[12] This intuition aligns with Allee et al.'s (2018) noise in researcher-calculated linguistic measures and gaps between amateur and expert scripters in calculating statistics from filings. Our purpose is not to develop a comprehensive measure of machine readability. Instead, we aim to develop a simple and straightforward measure to validate that iXBRL improves machine readability.

[13] WRDS follows the parsing procedures developed in Loughran and McDonald (2011), and outlines its parsing procedure at https://wrds-www.wharton.upenn.edu/pages/get-data/wrds-sec-analytics-suite/wrds-sec-filings-queries/readability-and-sentiment/.

[14] We develop our parsing program based on the same procedures used by Bill McDonald, which are disclosed on his website at https://sraf.nd.edu/sec-edgar-data/cleaned-10x-files/10x-stage-one-parsing-documentation/. In particular, in addition to procedures in Loughran and McDonald (2011), Bill McDonald removes tables, ASCII-encoded segments, SEC headers/footers, HTML predefined extended characters, and excess linefeeds (Loughran and McDonald 2014).

statistically significant differences at the 1% level. Overall, the evidence in Table 2 suggests our measures of machine readability detect significant increases after the adoption of iXBRL.

4. Empirical Analyses and Results

4.1. Main Analyses

We report our main analysis in Table 3. In Column (1), we include only our variable of interest, *Treat×Post*, with firm fixed effects and year-quarter fixed effects, and in Column (2), we include firm-quarter level control variables following Li (2008) and Lo et al. (2017). The coefficient estimates on *Treat×Post* are significantly positive across both tests, with *t*-statistics of 12.28 in Column (1) and 12.44 in Column (2). The coefficient on *Treat×Post* in Column (2), our baseline specification, is 0.670, which means that the human readability of treatment firms' disclosures decreases by 3.14% after the adoption of iXBRL and is equivalent to a 20 percentile drop in human readability for the median firm.[15] Regarding the control variables, we find patterns similar to the prior literature (Li 2008; Lo et al. 2017). For example, we consistently find that filings documenting losses (*Loss*) are harder to read, confirming that human readability of disclosures decreases when firm performance is poor.

Our DiD design assumes that the treatment and control firms have parallel trends of *Fog_Index* if the adoption of iXBRL does not occur. To test the validity of our empirical strategy, we add several time indicators for quarters before and after the compliance date in our DiD design. Specifically, in Table 4, *Pre4*, *Pre3*, *Pre2*, *Pre1*, *Post1*, *Post2*, *Post3*, and *Post4* equal one if the fiscal quarter ends in June 2018, September 2018, December 2018, March 2019, June 2019, September 2019, December 2019, and March 2020, respectively, and zero otherwise. *Pre4* is the

[15] DiD estimate/mean of *Fog_Index*=0.67/21.33=3.14%.

benchmark, so its interaction term, *Treat×Pre4*, is omitted in the regressions.[16] All other specifications remain the same.

In Column (1), we add indicators only for the pre-treatment period. In Column (2), we add indicators for both the pre-treatment and post-treatment periods. For both Column (1) and Column (2), coefficients on *Treat×Pre3*, *Treat×Pre2*, and *Treat×Pre1* are insignificant, indicating that the parallel trend assumption is satisfied. In Column (1), the coefficient on *Treat×Post* is significantly positive at the 1% level (*t*-statistic=9.51). In Column (2), coefficients on *Treat×Post1*, *Treat×Post2*, *Treat×Post3*, and *Treat×Post4* are all positive and significant at the 1% level, suggesting that human readability of disclosures starts to decrease right after the adoption of iXBRL. The coefficients on *Treat×Post3* and *Treat×Post4* are both significantly larger than the coefficient on *Treat×Post1*, consistent with an increasing influence of iXBRL adoption on human readability.

4.2. Capital Market Consequences to Retail Investors

The iXBRL format aims to "improve the data's usefulness, timeliness, and quality, benefiting investors, other market participants, and other data users …" (SEC 2018). However, our main analysis indicates that a reduction in human readability occurs after the adoption of iXBRL. Hypothesis 2 predicts reduced informativeness of corporate disclosures to retail investors as a result of this effect. In this section, we investigate the effect of the iXBRL mandate on the informativeness of disclosures to retail investors.

4.2.1 Informativeness

Prior studies often use absolute cumulative abnormal returns and abnormal trading volume to proxy for the informativeness of earnings announcements (e.g., Beaver 1968; Landsman and

[16] Results remain the same if we use another pre-treatment period, *Pre1*, as a benchmark.

Maydew 2002; Collins, Li, and Xie 2009; Beaver, McNichols, and Wang 2020). Given that we cannot measure these two variables only for retail investors, we split the sample into high and low retail investor ownership groups, where retail investor ownership is calculated as one minus institutional ownership (Campbell, Drake, Thornock, and Twedt 2023). Specifically, we use absolute cumulative abnormal returns and abnormal trading volume as proxies for informativeness and compare the coefficients on *Treat*×*Post* between the high and low retail ownership groups. We estimate the following model for each group:

$$Informativeness = \alpha + \beta Treat \times Post + \theta Controls + Firm\ FE + Year\_Quarter\ FE + \varepsilon. \quad (3)$$

The dependent variable, *Informativeness*, is one of the two measures for the informativeness of 10-K/Q filings. We follow Blankespoor et al. (2014) to include a set of control variables (*Controls*). The regression specification is the same as that used in our main analysis.

Panel A of Table 5 presents these results. In Columns (1) and (2), we define *Abs_CAR* as the absolute value of cumulative abnormal returns during the event period, where the abnormal return is the stock return minus the value-weighted market return. The event period is the period starting one day before and three days after the quarterly earnings announcement date (-1, +3) (Blankespoor et al. 2014). In Columns (3) and (4), we define *Abnormal_Volume* as the average daily trading volume during the event period minus the average daily trading volume during the non-filing period, divided by the standard deviation of daily trading volume during the non-filing period, which begins 49 days before and ends 5 days before the quarterly earnings announcement date (-49, -5) (Blankespoor et al. 2014). We find abnormally low absolute returns and trading volume only among the treatment firms with high retail investor ownership, with significant differences between the high and low retail investor ownership groups. This evidence suggests treatment firms' filings are less informative to retail investors after the iXBRL mandate, supporting Hypothesis 2 that the informativeness of disclosures is reduced to retail investors.

4.2.2. Other Retail Investor Measures

We also consider three additional measures of informativeness to retail investors: retail investor ownership, the number of retail shareholders, and the net retail flow. First, retail investor ownership, *Ownership_RetailInvestors*, is defined as one minus institutional ownership (Campbell et al. 2023). Second, to measure the number of retail shareholders, we use retail investor data from Robinhood, which is the first brokerage to offer commission-free trading to individual investors. Barber et al. (2022) argue that the simplicity of the Robinhood app attracts retail investors and that Robinhood users are mainly retail investors. Using the dataset from the Robintrack website (https://robintrack.net/), we measure the number of Robinhood shareholders, *Number_RetailInvestors*, as the logarithm of average number of daily Robinhood shareholders in the fiscal quarter. Third, to capture the net retail flow, we use the retail flow data from Nasdaq's Retail Trading Activity Tracker (RTAT), which covers roughly 45% of U.S. retail order flow (Even-Tov, George, Kogan, and So 2023). We define net retail flow, *Netflow_RetailInvestors*, as the average number of 10-day moving average retail net flows in the fiscal quarter, where a 10-day moving average retail net flow is calculated as 100*(buy flow – sell flow)/(buy flow + sell flow) in the most recent 10 trading days.[17] We estimate the following regression:

$$Retail\ Investor = \alpha + \beta Treat \times Post + \theta Controls + Firm\ FE + Year\_Quarter\ FE + \varepsilon. \quad (4)$$

[17] Nasdaq's daily retail flow data are based on the most recent 10 trading days. Therefore, we are not able to use Nasdaq's data to construct *Informativeness* within a short window of (-1, +3).

We follow Hong and Kacperczyk (2009) and include firm size (*Size*), market-to-book ratio (*MTB*), firm age (*Age*), return volatility (*Ret_Vol*), market beta (*Beta*), inverse of the stock price (*PrInv*), monthly stock return (*Return*), and an S&P 500 indicator (*SP500*) as control variables.

We report the results in Panel B of Table 5. In Column (1), we find a significantly negative coefficient on *Treat*×*Post* (-0.007; *t*-statistic=-2.31), indicating that relative to control firms, retail investor ownership of treatment firms decreases by 0.7% after improvements in machine readability. In Column (2), the coefficient on *Treat*×*Post* is -0.054 (*t*-statistic=-2.39), suggesting that after the iXBRL adoption, the number of Robinhood shareholders in treatment firms decreases. In Column (3), the coefficient on *Treat*×*Post* is -0.565 (*t*-statistic=-2.32), meaning that treatment firms have more net retail outflow after the machine readability shock. Therefore, we find that after the increase in machine readability (and the associated decrease in human readability), disclosures become less informative to retail investors and retail investors leave the affected firms.

4.3. Additional Analyses

Next, we explore cross-sectional variation in the effect of machine readability on human readability to test whether managers opportunistically and/or non-opportunistically reduce human readability in response to a mandatory increase in machine readability.

4.3.1. Opportunistic Mechanisms

We argue that certain firms have incentives to obfuscate financial disclosure through lower human readability in response to the increased machine readability from the iXBRL mandate. In particular, the mandate to improve machine readability is likely more costly for firms with high proprietary costs of disclosure. In addition, the Incomplete Revelation Hypothesis suggests the effect will be more pronounced when managers have stronger incentives to obfuscate information in their quarterly and annual reports due to either poor performance or earnings management

(Bloomfield 2002; Li 2008; Lo et al. 2017). Given that improved machine readability makes it easier for users to identify proprietary information or detect these activities, we argue that managers are more likely to lower the human readability level of their disclosures in these settings.

We use confidential treatment of proprietary information to proxy for proprietary costs of disclosure (Verrecchia and Weber 2006; Thompson et al. 2023). When firms request confidential treatment of their proprietary information in their public disclosures, they need to file a Form CT Order, which provides confidential treatment for certain disclosures. We set *Confidential Treatment of Proprietary Information* equal to one if the firm files at least one Form CT Order during 2018Q2 to 2019Q1, which is our pre-treatment period, and equal to zero otherwise. As shown in Columns (1) and (2) of Table 6, the effect for both the high and low proprietary costs groups is statistically significant. However, the magnitude of the effect is significantly larger for the high proprietary costs group than for the low proprietary costs group, suggesting that firms with higher proprietary costs reduce human readability to a greater degree after adopting iXBRL.[18]

In Columns (3) and (4), we identify filings based on whether the firm-quarter filing documents a loss. For loss firms, the coefficient on *Treat×Post* (0.823) is significantly larger than that for non-loss firms (0.641), and this difference in coefficients is significant at the 1% level. The result is consistent with our prediction that managers obfuscate negative information by reducing the human readability of disclosures after a mandatory improvement in machine readability.

We estimate discretionary accruals using the performance-matched modified Jones model (Kothari, Leone, and Wasley 2005) and use its absolute value to measure earnings management, *Earnings Management*. We split the sample into *High* and *Low* groups based on the median value

[18] In untabulated analyses, we use two alternative firm-year level proxies for proprietary costs of disclosure: (1) R&D intensity, calculated as R&D expenditures divided by sales, and (2) intangible assets, measured as intangible assets net of goodwill divided by total assets. Our inference remains the same when we use these proxies.

of *Earnings Management* within the quarter and industry. The coefficient for firms in the *High* earnings management group (0.745) in Column (5) is larger than for firms in the *Low* earnings management group (0.598) in Column (6), and the difference is significant at the 5% level, suggesting that firms with higher earnings management have a stronger incentive to decrease human readability after a mandatory improvement in machine readability.

If the reduction in human readability following regulatory improvements in machine readability arises because of opportunism, the effect should be more pronounced among managers who have greater opportunities to make such adjustments. We expect weak board monitoring to provide such an opportunity. Armstrong, Core, and Guay (2014) argue that independent directors require corporate transparency to perform monitoring and advising roles, limiting management's opportunity to obfuscate disclosures. In Column (7) and Column (8) of Table 6, we partition the sample based on whether the firm's percentage of independent directors exceeds the industry-quarter median. The sample size for this analysis is much smaller because the data of independent directors are from Institutional Shareholder Services (ISS), which covers mainly S&P 1500 companies. The coefficient for firms with less monitoring from independent directors is significantly larger than that for firms with more monitoring (0.658 vs. 0.335), and the difference in coefficients is significant at the 1% level. In sum, the reduction in human readability after the adoption of iXBRL is more pronounced for firms with stronger incentives and opportunities to obfuscate financial disclosures, consistent with the opportunistic mechanism.

4.3.2. Non-Opportunistic Mechanisms

Prior literature documents that managers have limited attention, time, and resources in preparing financial disclosure (Doyle et al. 2007; Li et al. 2023). The adoption of iXBRL entails considerable work for companies, necessitating additional human resources to manage the increased workload. We argue that firms that bring in additional human capital to help implement the iXBRL adoption are better positioned to avoid corresponding reductions in human readability. To identify whether firms add additional accounting or computer-related human resources required for preparing financial disclosures following the iXBRL adoption, we utilize job posting data from RavenPack's Job Analytics database, which contains cleaned and detailed job posting information sourced from LinkUp. As a leading job market data provider, LinkUp sources data from over 50,000 employers, starting from 2007. It provides comprehensive information on job posts, including company identifier, job title, position details, job descriptions, and required skillsets (Chen and Li 2023). Although a job posting does not necessarily mean that hiring actually takes place, Campello, Kankanhalli, and Muthukrishnan (2020) confirm that there is a close link between LinkUp's job postings data and actual job gains recorded at firms. They further note that LinkUp's job posting data provides a reasonable representation of job gains.

Using RavenPack's Job Analytics database, we identify a job posting as an accounting (computer) related job posting if the job position is labelled by RavenPack as an accounting (computer) position based on U.S. Bureau of Labor Statistics's Standard Occupational Classification (SOC) system.[19] Accounting staff in firms that do not hire additional accounting or computer staff will bear the extra workload induced by iXBRL compliance, which in turn distracts them from efforts that would otherwise facilitate human readability.

In Columns (1) and (2) of Table 7, we divide the sample based on whether the firm has accounting job postings from June 28, 2018, the adoption date of the iXBRL regulation, to June 15, 2019, the compliance date for the treatment firms. The coefficient on *Treat×Post* (0.749) for firms that do not issue accounting job postings in the pre-treatment period is significantly larger

[19] Specifically, the accounting's SOC is 43-3000 and the computer's SOC is 15-1200.

than that for firms that issue accounting job postings (0.631). This difference in coefficients is significant at the 10% level. In Columns (3) and (4), we partition the sample based on whether the firm posts computer job openings during the pre-treatment period. We find that firms that do not issue computer job postings reduce human readability further compared with firms that issue such job postings (0.759 vs. 0.631).

In addition to limited resources, firm complexity is another factor that constrains firms from maintaining the human readability of financial disclosures during the transition from XBRL to iXBRL (Lim, Chalmers, and Hanlon 2018). In Columns (5) and (6) of Table 7, *Firm Complexity* is developed by Loughran and McDonald (2023) and is defined as the percentage of complexity words relative to the total number of words. We split the sample into *High* and *Low* groups based on the median value of *Firm Complexity* of 10-K filings of fiscal year 2018, the pre-treatment period, within the industry. The coefficient for firms in the *High* firm complexity group (0.721) in Column (5) is larger than that for firms in the *Low* firm complexity group (0.606) in Column (6), and the difference is significant at the 10% level. This finding suggests more complex firms experience a greater reduction in human readability after the adoption of iXBRL, consistent with the non-opportunistic mechanism. These findings are consistent with non-opportunistic mechanisms playing a role in the decrease in human readability when machine readability improves.

5. Robustness Analyses

5.1. Alternative Research Designs

One concern in our DiD approach is that treatment and control firms are fundamentally different because the iXBRL regulation applies only to firms with at least $700 million in public float. To mitigate this concern, we directly control for firm size and include firm fixed effects in our previous analyses. In addition, we develop two alternative research designs with similarly sized treatment and control firms, which we outline below.

5.1.1. Regression Discontinuity Design (RDD)

In the first alternative design, we employ a regression discontinuity design to compare firms with public float values that barely exceed a threshold (the "treatment group") and firms with public float values barely below the threshold (the "control group"). The public float among these firms is continuous across the threshold, and any significant discontinuity in human readability at the threshold is a result of the treatment (iXBRL adoption). To estimate the discontinuity in human readability, we follow Calonico, Cattaneo, and Titiunik (2014) and Calonico, Cattaneo, Farrell, and Titiunik (2019) to employ a local polynomial-based fuzzy RDD estimation approach. We implement the fuzzy RDD estimation instead of a sharp RDD estimation because the probability of iXBRL treatment increases at the cutoff point (i.e., public float of $700 million) but does not deterministically jump from 0 to 1, consistent with Blankespoor (2019).[20] We use both the second-order and third-order polynomial functions to estimate the treatment effect.

Table 8 reports regression discontinuity estimates. The discontinuity in human readability is statistically significant at the 5% level using both the second-order (Column (1)) and the third-order (Column (2)) polynomials. Thus, these results suggest that our DiD estimates are not driven by differences in public float.

5.1.2. Voluntary Adopters

Some firms voluntarily adopted iXBRL before the mandated adoption date (June 15, 2019 for large firms and June 15, 2020 for small firms), consistent with Luo et al. (2023). In our main

[20] The public float data we obtain from 10-K filings only provides a snapshot as of the fiscal year-end date, which doesn't necessarily coincide with the treatment assignment date. In addition, while the public float is a significant factor, it is not the sole determinant in classifying a firm as a treatment firm (i.e., a large accelerated filer). Prior studies using treatment assignments based on large accelerated filers face these same issues (e.g., Blankespoor 2019).

analysis, we exclude these firms for a cleaner identification. In an alternative DiD design, we compare these voluntary adopters to the firms that adopt at the mandatory adoption date. The primary objective of this design is to compare large firms that newly adopted iXBRL at the June 15, 2019 adoption date to other firms of similar size that did not adopt at the same time.

In this alternative DiD design, *Treat* equals one for large firms that adopt iXBRL only after June 15, 2019 mandate, and zero for large firms that voluntarily adopt iXBRL before June 15, 2019. Table 9 provides the results using this alternative DiD design. In Column (1), we regress *Fog_Index* on the interaction term, and in Column (2), we add control variables. The coefficients on *Treat*×*Post* are 0.295 (t-statistic=3.88) in Column (1) and 0.306 (t-statistic=4.06) in Column (2). These results suggest that compared with disclosures of voluntary adopters, those of mandated adopters (of similar size) experience a decrease in human readability after the iXBRL adoption. These findings mitigate concerns that our main findings are driven by differences in firm size between treatment and control firms.[21]

5.2. Placebo Tests

To further enhance the reliability of our results, we conduct two placebo tests. The first is based on a placebo treatment *date*. In Table 10 Column (1), *Treat* is defined the same as in our main analysis, but *Post_Placebo* equals one if the fiscal quarter is after June 15, 2017, which is a placebo treatment date two years before the actual treatment date, and zero otherwise. The sample comprises 16,641 10-K/Q filings for fiscal quarters from 2016Q2 to 2018Q1, which does not cover any treatment period and does not overlap with the sample period in our main analysis. In Column (1), we do not find any significant coefficients on *Treat*×*Post_Placebo*.

[21] In additional tests reported in Online Appendix B, we re-estimate the main analysis including only 10-Q filings, or alternatively, only 10-K filings, and we find results consistent with those in the main analysis. In addition, we add back filings with non-December fiscal year-ends, and also estimate the effect of machine readability only for filings with non-December fiscal year-ends, and the results remain consistent.

In Column (2), we perform a placebo test using a placebo treatment *group*. In our main analysis, treatment firms are larger (public float > $700 million) than firms in our control (untreated) sample. Therefore, in this test we exclude all large firms and define *Treat_Placebo* as one if the firm has public float between $75 million and $700 million (i.e., firms not subject to the June 15, 2019 mandate), and zero if the firm has public float less than $75 million. In this analysis, we employ the same sample period as in our main analysis, such that the post period does not capture a mandated increase in machine readability for placebo-treatment or placebo-control firms. In this way, the only difference between these firms is their size. As reported in Column (2), we do not find any significant results. These placebo tests provide additional support for the conclusion that the decrease in human readability that we document is a result of an improvement in machine readability imposed by iXBRL adoption.

5.3. Alternative Methods to Handle Tables

When calculating human readability, it is important to delete tables because human readability measures are designed for text instead of tables, yet tables often include short sentences that can artificially inflate human readability levels (Li 2008). Furthermore, when machine readability is poor, it is challenging to accurately identify and exclude all tables. Therefore, as machine readability improves and more tables are correctly excluded, measures of human readability could mechanically decrease. To rule out this alternative explanation for our findings, our main analyses already remove tables using table tags and drop sentences with fewer than five words when computing the Fog index, as these short sentences are likely a part of tables.

In addition, we perform two robustness tests. In both tests, the procedure for handling tables remains consistent across both pre- and post-iXBRL adoption periods. In the first test, we retain all tables in filings from both pre-iXBRL and post-iXBRL periods and, using these parsed filings, we re-compute the Fog index (*Fog_Alternative1*). Thus, *Fog_Alternative1* measures human readability in a way that gives identical treatment (i.e., no exclusion) to tables in both the pre- and post-iXBRL periods. In the second test, we exclude tables in both the pre- and post-iXBRL periods by removing all sentences with fewer than five words, but do so without the help of table tags, which more effectively identify tables in the post-iXBRL period than in the pre-iXBRL period (*Fog_Alternative2*). In both robustness tests (Online Appendix C), the coefficient on *Treat×Post* remains significantly positive, indicating that our primary findings are not driven by measurement error associated with the identification of tables in the post-iXBRL period.

5.4. Alternative Measures of Human Readability

Guay, Samuels, and Taylor (2016) argue that a variety of alternative readability measures can be used as a robustness test in addition to the Fog index. In Table 11, we use six alternative measures of human readability and rerun the regression with the same specification used in the main analysis. Detailed definitions of these alternative measures of human readability are defined in Appendix A. The coefficient estimates on *Treat×Post* are significantly positive at the 1% level when we use these alternative measures. In addition, we find consistent results for control variables when compared to the main analysis. The findings confirm the robustness of the main analysis.

6. Conclusion

We examine the effects of machine readability of corporate disclosures on the human readability of these disclosures. Using a difference-in-differences approach based on an iXBRL regulation that creates a positive shock to machine readability, we find a negative effect of machine readability on human readability. In particular, enhancements to machine readability are associated with reductions in human readability. Further evidence indicates that the mandated improvement in machine readability has adverse consequences for retail investors who disproportionately rely on their own ability (rather than machines) to process corporate disclosures. Additional analyses show that reductions in human readability are more pronounced among firms with incentives and opportunities to obfuscate their disclosures, and among firms with resource constraints and complex operations, suggesting both opportunistic and non-opportunistic reasons for the observed effect. These results provide timely evidence of the consequences of machine readability and have important implications for both investors and regulators.

References

Allee, K. D., M. D. Deangelis, and J. R. Moon. 2018. Disclosure "Scriptability." *Journal of Accounting Research* 56 (2): 363–430.

Allee, K. D., C. Do, and M. Sterin. 2021. Product Market Competition, Disclosure Framing, and Casting in Earnings Conference Calls. *Journal of Accounting and Economics* 72 (1): 101405.

Armstrong, C. S., J. E. Core, and W. R. Guay. 2014. Do Independent Directors Cause Improvements in Firm Transparency? *Journal of Financial Economics* 113 (3): 383–403.

Barber, B. M., X. Huang, T. Odean, and C. Schwarz. 2022. Attention-Induced Trading and Returns: Evidence from Robinhood Users. *The Journal of Finance* 77 (6): 3141–3190.

Barbopoulos, L. G., R. Dai, T. J. Putniņš, and A. Saunders. 2023. Market Efficiency in the Age of Machine Learning. Working paper, New York University.

Basoglu, K. A., and C. E. (Skip) White. 2015. Inline XBRL Versus XBRL for SEC Reporting. *Journal of Emerging Technologies in Accounting* 12 (1): 189–199.

Bauguess, S. W. 2017. The Role of Big Data, Machine Learning, and AI in Assessing Risks: a Regulatory Perspective. https://www.sec.gov/news/speech/bauguess-big-data-ai.

Bauguess. 2018. The Role of Machine Readability in an AI World. SEC Keynote Address: Financial Information Management (FIMA) Conference.

Beaver, W. H. 1968. The Information Content of Annual Earnings Announcements. *Journal of Accounting Research* 6: 67–92.

Beaver, W. H., M. F. McNichols, and Z. Z. Wang. 2020. Increased Market Response to Earnings Announcements in the 21st Century: An Empirical Investigation. *Journal of Accounting and Economics* 69 (1): 101244.

Bernard, D. 2016. Is the Risk of Product Market Predation a Cost of Disclosure? *Journal of Accounting and Economics* 62 (2): 305–325.

Bernard, D., D. Burgstahler, and D. Kaya. 2018. Size Management by European Private Firms to Minimize Proprietary Costs of Disclosure. *Journal of Accounting and Economics* 66 (1): 94–122.

Bhattacharya, N., Y. J. Cho, and J. B. Kim. 2018. Leveling the Playing Field Between Large and Small Institutions: Evidence from the SEC's XBRL Mandate. *The Accounting Review* 93 (5): 51–71.

Biddle, G. C., G. Hilary, and R. S. Verdi. 2009. How Does Financial Reporting Quality Relate to Investment Efficiency? *Journal of Accounting and Economics* 48 (2–3): 112–131.

Blankespoor, E. 2019. The Impact of Information Processing Costs on Firm Disclosure Choice: Evidence from the XBRL Mandate. *Journal of Accounting Research* 57 (4): 919–967.

Blankespoor, E., E. deHaan, and I. Marinovic. 2020. Disclosure Processing Costs, Investors' Information Choice, and Equity Market Outcomes: A Review. *Journal of Accounting and Economics* 70 (2–3): 101344.

Blankespoor, E., B. P. Miller, and H. D. White. 2014. Initial Evidence on the Market Impact of the XBRL Mandate. *Review of Accounting Studies* 19 (4): 1468–1503.

Bloomfield, R. J. 2002. The "Incomplete Revelation Hypothesis" and Financial Reporting. *Accounting Horizons* 16 (3): 233–243.

Bonsall, S. B., A. J. Leone, B. P. Miller, and K. Rennekamp. 2017. A Plain English Measure of Financial Reporting Readability. *Journal of Accounting and Economics* 63 (2–3): 329–357.

Brav, A., M. Cain, and J. Zytnick. 2022. Retail Shareholder Participation in the Proxy Process: Monitoring, Engagement, and Voting. *Journal of Financial Economics* 144 (2): 492–522.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik. 2019. Regression Discontinuity Designs Using Covariates. *Review of Economics and Statistics* 101 (3): 442–451.
Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica* 82 (6): 2295–2326.
Cao, S., W. Jiang, B. Yang, and A. L. Zhang. 2023. How to Talk When a Machine Is Listening: Corporate Disclosure in the Age of AI. *The Review of Financial Studies* 36 (9): 3603–3642.
Campbell, B., M. Drake, J. Thornock, and B. Twedt. 2023. Earnings Virality. *Journal of Accounting and Economics* 75 (1): 101517.
Campello, M., G. Kankanhalli, and P. Muthukrishnan. 2020. Corporate Hiring under COVID-19: Labor Market Concentration, Downskilling, and Income Inequality. Working paper, National Bureau of Economic Research.
Chen, C.-W., and L. Y. Li. 2023. Is Hiring Fast a Good Sign? The Informativeness of Job Vacancy Duration for Future Firm Profitability. *Review of Accounting Studies* 28 (3): 1316–1353.
Clayton, J. 2017. Remarks at the Economic Club of New York. https://www.sec.gov/news/speech/remarks-economic-club-new-york.
Collins, D. W., O. Z. Li, and H. Xie. 2009. What Drives the Increased Informativeness of Earnings Announcements Over Time? *Review of Accounting Studies* 14 (1): 1–30.
De George, E. T., C. B. Ferguson, and N. A. Spear. 2013. How Much Does IFRS Cost? IFRS Adoption and Audit Fees. *The Accounting Review* 88(2): 429–462.
Dedman, E., and C. Lennox. 2009. Perceived Competition, Profitability and the Withholding of Information about Sales and the Cost of Sales. *Journal of Accounting and Economics* 48 (2): 210–230.
deHaan, E., Y. Song, C. Xie, and C. Zhu. 2021. Obfuscation in Mutual Funds. *Journal of Accounting and Economics* 72 (2–3): 101429.
Deloitte. 2018. SEC Requires the Use of Inline XBRL in Certain Filings. https://www2.deloitte.com/content/dam/Deloitte/ch/Documents/audit/deloitte-ch-audit-sec-requires-the-use-of-inline-xbrl-in-certain-filings.pdf.
Dong, Y., O. Z. Li, Y. Lin, and C. Ni. 2016. Does Information-Processing Cost Affect Firm-Specific Information Acquisition? Evidence from XBRL Adoption. *Journal of Financial and Quantitative Analysis* 51 (2): 435–462.
Doyle, J., W. Ge, and S. McVay. 2007. Determinants of Weaknesses in Internal Control over Financial Reporting. *Journal of Accounting and Economics* 44 (1): 193–223.
Driscoll, P. B. 2019. How We Protect Retail Investors. https://www.sec.gov/news/speech/speech-driscoll-042919.
Egan, M. 2019. How Elite Investors Use Artificial Intelligence and Machine Learning to Gain an Edge. https://www.cnn.com/2019/02/17/investing/artificial-intelligence-investors-machine-learning/index.html.
Ellis, J. A., C. E. Fee, and S. E. Thomas. 2012. Proprietary Costs and the Disclosure of Information About Customers. *Journal of Accounting Research* 50 (3): 685–727.
Ertugrul, M., J. Lei, J. Qiu, and C. Wan. 2017. Annual Report Readability, Tone Ambiguity, and the Cost of Borrowing. *Journal of Financial and Quantitative Analysis* 52 (2): 811–836.
Even-Tov, O., K. George, S. Kogan, and E. C. So. 2023. Fee the People: Retail Investor Behavior and Trading Commission Fees. Working Paper, University of California, Berkeley.
Farrell, M., T. C. Green, R. Jame, and S. Markov. 2022. The Democratization of Investment Research and the Informativeness of Retail Investor Trading. *Journal of Financial Economics* 145 (2): 616–641.

Frankel, R., J. Lee, and Z. Lemayian. 2018. Proprietary Costs and Sealing Documents in Patent Litigation. *Review of Accounting Studies* 23 (2): 452–486.

Guay, W., D. Samuels, and D. Taylor. 2016. Guiding Through the Fog: Financial Statement Complexity and Voluntary Disclosure. *Journal of Accounting and Economics* 62 (2–3): 234–269.

Guo, K. H., and X. Yu. 2022. Retail Investors Use XBRL Structured Data? Evidence from the SEC's Server Log. *Journal of Behavioral Finance* 23 (2): 166–174.

Hail, L., C. Leuz, and P. Wysocki. 2010. Global Accounting Convergence and the Potential Adoption of IFRS by the US (Part I): Conceptual Underpinnings and Economic Analysis. *Accounting Horizons* 24(3): 355–394.

Harris, T. S., and S. Morsfield. 2012. An Evaluation of the Current State and Future of XBRL and Interactive Data for Investors and Analysts. White paper, Columbia Business School.

Hayes, R. M., and R. Lundholm. 1996. Segment Reporting to the Capital Market in the Presence of a Competitor. *Journal of Accounting Research* 34 (2): 261–279.

Hoitash, R., U. Hoitash, and L. Morris. 2021. eXtensible Business Reporting Language (XBRL): A Review and Implications for Future Research. *Auditing: A Journal of Practice & Theory* 40 (2): 107–132.

Hong, H., and M. Kacperczyk. 2009. The Price of Sin: The Effects of Social Norms on Markets. *Journal of Financial Economics* 93 (1): 15–36.

Hostak, P., T. Lys, Y.G. Yang, and E. Carr. 2013. An Examination of the Impact of the Sarbanes–Oxley Act on the Attractiveness of US Capital Markets for Foreign Firms. *Review of Accounting Studies* 18: 522–559.

Hwang, B.-H., and H. H. Kim. 2017. It Pays to Write Well. *Journal of Financial Economics* 124 (2): 373–394.

Jamal, K., R. Bloomfield, T. E. Christensen, R. H. Colson, S. Moehrle, J. Ohlson, S. Penman, T. Stober, S. Sunder, and R. L. Watts. 2010. A Research-based Perspective on the SEC's Proposed Rule—Roadmap for the Potential Use of Financial Statements Prepared in Accordance with International Financial Reporting Standards (IFRS) by US Issuers. *Accounting Horizons* 24 (1): 139–147.

Kalay, A. 2015. Investor Sophistication and Disclosure Clienteles. *Review of Accounting Studies* 20 (2): 976–1011.

Kim, C. (Francis), K. Wang, and L. Zhang. 2019. Readability of 10-K Reports and Stock Price Crash Risk. *Contemporary Accounting Research* 36 (2): 1184–1216.

Kim, J.-B., J. W. Kim, and J. Lim. 2019. Does XBRL Adoption Constrain Earnings Management? Early Evidence from Mandated U.S. Filers. *Contemporary Accounting Research* 36 (4): 2610–2634.

Kim, J.-B., B. Li, and Z. Liu. 2019. Information-Processing Costs and Breadth of Ownership. *Contemporary Accounting Research* 36 (4): 2408–2436.

Kothari, S. P., A. J. Leone, and C. E. Wasley. 2005. Performance Matched Discretionary Accrual Measures. *Journal of Accounting and Economics* 39 (1): 163–197.

Landsman, W. R., and E. L. Maydew. 2002. Has the Information Content of Quarterly Earnings Announcements Declined in the Past Three Decades? *Journal of Accounting Research* 40 (3): 797–808.

Laudenbach, C., B. Loos, J. Pirschel, and J. Wohlfart. 2021. The Trading Response of Individual Investors to Local Bankruptcies. *Journal of Financial Economics* 142 (2): 928–953.

Lawrence, A. 2013. Individual Investors and Financial Disclosure. *Journal of Accounting and Economics* 56 (1): 130–147.

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., F. Li, and K. Merkley. 2011. The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts. *The Accounting Review* 86 (3): 1087–1115.

Li, F. 2008. Annual Report Readability, Current Earnings, and Earnings Persistence. *Journal of Accounting and Economics* 45 (2–3): 221–247.

Li, X., H. Zhu, and L. Zuo. 2021. Reporting Technologies and Textual Readability: Evidence from the XBRL Mandate. *Information Systems Research* 32 (3): 1025–1042.

Li, Y., and L. Zhang. 2015. Short Selling Pressure, Stock Price Behavior, and Management Forecast Precision: Evidence from a Natural Experiment. *Journal of Accounting Research* 53 (1): 79–117.

Li, Z., K. Ye, C. Zeng, and B. Zhang. 2023. Ending at the Wrong Time: The Financial Reporting Consequences of a Uniform Fiscal Year-End. *The Accounting Review* 98 (3): 367–396.

Liang, C. 2023. Advertising Rivalry and Discretionary Disclosure. *Journal of Accounting and Economics* forthcoming.

Lim, E. K., K. Chalmers, and D. Hanlon. 2018. The Influence of Business Strategy on Annual Report Readability. *Journal of Accounting and Public Policy* 37 (1): 65–81.

Linck, J. S., J. M. Netter, and T. Yang. 2009. The Effects and Unintended Consequences of the Sarbanes-Oxley Act on the Supply and Demand for Directors. *The Review of Financial Studies* 22 (8): 3287–3328.

Lo, K., F. Ramos, and R. Rogo. 2017. Earnings Management and Annual Report Readability. *Journal of Accounting and Economics* 63 (1): 1–25.

Loughran, T., and B. McDonald. 2011. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance* 66 (1): 35–65.

Loughran, T., and B. McDonald. 2014. Measuring Readability in Financial Disclosures. *The Journal of Finance* 69 (4): 1643–1671.

Loughran, T., and B. McDonald. 2023. Measuring Firm Complexity. *Journal of Financial and Quantitative Analysis* forthcoming.

Luo, X., T. (David) Wang, L. Yang, X. Zhao, and Y. Zhang. 2023. Initial Evidence on the Market Impact of the iXBRL Adoption. *Accounting Horizons* 37(1):143–171.

PwC. 2016. What Public Companies Should Know About Inline XBRL. https://www.pwc.com/us/en/risk-assurance/publications/inline-xbrl.pdf.

Ritz, D. Understanding Machine Readability in Modern Data Policy. *Data Foundation*. https://www.datafoundation.org/understanding-machine-readability-in-modern-data-policy-2020.

SEC. 2018. Inline XBRL Filing of Tagged Data. https://www.sec.gov/rules/final/2018/33-10514.pdf.

SEC. 2023. Semi-Annual Report to Congress Regarding Public and Internal Use of Machine-Readable Data for Corporate Disclosures.

Thompson, A. M., O. Urcan, and H. Yoon. 2022. Do Companies Redact Material Information from Confidential SEC Filings? Evidence from the FAST Act. *The Accounting Review* 98 (3): 229.

U.S. House of Representatives. 2022. James M. Inhofe National Defense Authorization Act for Fiscal Year 2023.

Verrecchia, R. E. 1983. Discretionary Disclosure. *Journal of Accounting and Economics* 5: 179–194.

Verrecchia, R. E., and J. Weber. 2006. Redacted Disclosure. *Journal of Accounting Research* 44 (4): 791–814.

Welch, I. 2022. The Wisdom of the Robinhood Crowd. *The Journal of Finance* 77 (3): 1489–1527.

Appendix A. Variable Definitions

Variables	Descriptions
Dependent Variable	
Fog_Index	The Gunning Fog index of annual and quarterly reports, measured as 0.4 * ((number of words / number of sentences) + 100 * (number of words with more than two syllables / number of words)). (Source: SEC EDGAR)
Independent Variables	
Treat	Equals one if the firm is a large accelerated filer (i.e., public float ≥$700 million) and zero otherwise. (Source: SEC EDGAR)
Post	Equals one if the fiscal quarter is after June 15, 2019, and zero otherwise.
Public Float (in RDD Test)	Public float extracted from SEC 10-K filings. (Source: SEC EDGAR)
Treat_Alternative	Equals one if large firms initiate iXBRL filings after the treatment date, i.e., June 15, 2019 (mandatory adopters), and zero if large firms voluntarily adopt iXBRL before the treatment date (voluntary adopters). (Source: SEC EDGAR)
Treat_Placebo (in Placebo Test)	Equals one if the firm has public float between $75 million and $700 million (i.e., firms not subject to the June 15, 2019 mandate), and zero if the firm has public float less than $75 million. (Source: SEC EDGAR)
Post_Placebo (in Placebo Test)	Equals one if the fiscal quarter is after June 15, 2017 (i.e., two years before the actual treatment date), and zero otherwise.
Control Variables	
Earnings	Operating earnings scaled by total assets at the fiscal quarter-end. (Source: Compustat)
Loss	Indicator variable that equals one if *Earnings*<0, and zero otherwise. (Source: Compustat)
Size	Logarithm of market value of equity at the fiscal quarter-end. (Source: Compustat)
MTB	Market-to-book ratio, measured as (market value of equity + book value of liabilities) / book value of total assets at the fiscal quarter-end. (Source: Compustat)
Age	Number of years since a firm first appears in the CRSP monthly stock return file. (Source: CRSP)
Special_Items	Special items divided by total assets at the fiscal quarter-end. (Source: Compustat)
Ret_Vol	Standard deviation of monthly stock returns in the last 4 fiscal quarters. (Source: Compustat)
Earn_Vol	Standard deviation of operating earnings in the last 12 fiscal quarters. (Source: Compustat)
NBSeg	Logarithm of 1 plus the number of business segments. (Source: Compustat)
NGSeg	Logarithm of 1 plus the number of geographic segments. (Source: Compustat)
Nitems	Logarithm of number of items in Compustat with non-missing values. (Source: Compustat)

SEO	Indicator variable that equals one if a firm has seasoned equity offering in this quarter, and zero otherwise. (Source: SDC Global New Issues)
MA	Indicator variable that equals one if a firm appears as an acquirer in this quarter in the SDC Platinum M&A database, and zero otherwise. (Source: SDC Platinum M&A)
Delaware	Indicator variable that equals one if a firm is incorporated in Delaware, and zero otherwise. (Source: SEC EDGAR)
Consequence Variables	
Abs_CAR	Absolute value of cumulated abnormal returns during the 5-day event window (-1, 3), where day 0 is the 10-K/10-Q filing date and abnormal return is return minus value weighted market return. (Source: CRSP)
Abnormal_Volume	Mean daily trading volume during the 5-day event window (-1, 3) minus the mean daily trading volume during the non-filing period (-49, -5), deflated by the standard deviation of daily trading volume during the non-filing period, where day 0 is the 10-K/10-Q filing date. (Source: CRSP)
Number_RetailInvestors	Logarithm of average number of daily Robinhood shareholders in the quarter. (Source: Robintrack)
Ownership_RetailInvestors	Retail investor ownership, which equals 1 minus *Institutional Ownership*. (Source: Thomson Reuters 13F)
NetFlow_RetailInvestors	Average number of 10-day moving average retail net flows in the fiscal quarter, where a 10-day moving average retail net flow is calculated as 100*(buy flow – sell flow)/(buy flow + sell flow) in the most recent 10 trading days. (Source: Nasdaq's Retail Trading Activity Tracker)
Other Variables Used in Consequence Analyses	
Institutional Ownership	Percentage of shares held by institutional investors in the quarter. (Source: Thomson Reuters 13F)
Analyst	Natural logarithm of 1 plus the number of analysts following the firm in the quarter. (Source: IBES)
Abs_SUE	Absolute value of earnings surprise, where earnings surprise is defined as actual earnings per share minus the latest analyst consensus forecast in the quarter, scaled by the stock price on the fiscal quarter end date. (Source: IBES)
Leverage	Total liabilities divided by total assets at the fiscal quarter end. (Source: Compustat)
FileLag	Natural logarithm of 1 plus the number of days from the earnings announcement date to the 10-K/Q filing date for the fiscal quarter. (Source: Compustat and SEC EDGAR)
LnPrc	Natural logarithm of stock price on the fiscal quarter end date. (Source: CRSP)
Return	Average monthly return during the fiscal period. (Source: CRSP)
Beta	Firm's industry market beta calculated over last 36 months. (Source: CRSP)
PrInv	Inverse of stock price on the fiscal end date. (Source: CRSP)
SP500	Equals one if the stock is in the S&P 500 index, and zero otherwise (Source: Standard & Poor)
Cross-Sectional Variables	

Confidential Treatment of Proprietary Information	*Yes* indicates the firm issues at least one confidential treatment order (Form CT Order) during 2018Q2 to 2019Q1, and *No* indicates 0 confidential treatment orders during 2018Q2 to 2019Q1. (Source: SEC EDGAR)
Firm Performance	*Loss* indicates *Earnings*<0, and *Non-Loss* indicates *Earnings*>=0. (Source: Compustat)
Earnings Management	Absolute value of performance-matched abnormal accruals, calculated using the method in Kothari et al. (2005). *High* indicates the firm's *Earnings Management* is above the median value of that within the quarter and industry, and *Low* indicates the firm's *Earnings Management* is below or equal to the median value of that within the quarter and industry. (Source: Compustat)
Monitoring	Percentage of independent directors. *High* indicates the firm's *Monitoring* is above the median value of that within the fiscal year and industry, and *Low* indicates the firm's *Monitoring* is below or equal to the median value of that within the fiscal year and industry. (Source: ISS)
Accounting Job Postings	*Yes* indicates the firm issues at least one accounting job posting from July 2018 to June 2019, and *No* indicates the firm does not issue such job posting in the period. (Source: RavenPack's Job Analytics)
Computer Job Postings	*Yes* indicates the firm issues at least one computer job posting from July 2018 to June 2019, and *No* indicates the firm does not issue such job posting in the period. (Source: RavenPack's Job Analytics)
Firm Complexity	The number of complexity-related words divided by the number of words in 2018 fiscal year's 10-K filings, where complexity-related words are defined in Loughran and McDonald (2023). *High* indicates *Firm Complexity* is above the median value within the industry, and *Low* indicates *Firm Complexity* is below or equal to the median value within the industry. (Source: SEC EDGAR)
Validation Variables	
Gap1	Absolute difference in numbers of words between WRDS SEC Analytics Suite and Bill McDonald's database divided by sum of numbers of words in the two sources. (Source: Bill McDonald's website and WRDS SEC Analytics Suite)
Gap2	Absolute difference in numbers of words between Bill McDonald's database and self-developed parsed filings divided by sum of numbers of words in the two sources. (Source: Bill McDonald's website and SEC EDGAR)
Gap3	Absolute difference in numbers of words between WRDS SEC Analytics Suite and self-developed parsed filings divided by sum of numbers of words in the two sources. (Source: WRDS SEC Analytics Suite and SEC EDGAR)
Alternative Dependent Variables	
Flesch-Kincaid	The Flesch-Kincaid Readability Index, defined as 0.39 * (number of words / number of sentences) + 11.8 * (number of syllables / number of words) - 15.59. (Source: SEC EDGAR)

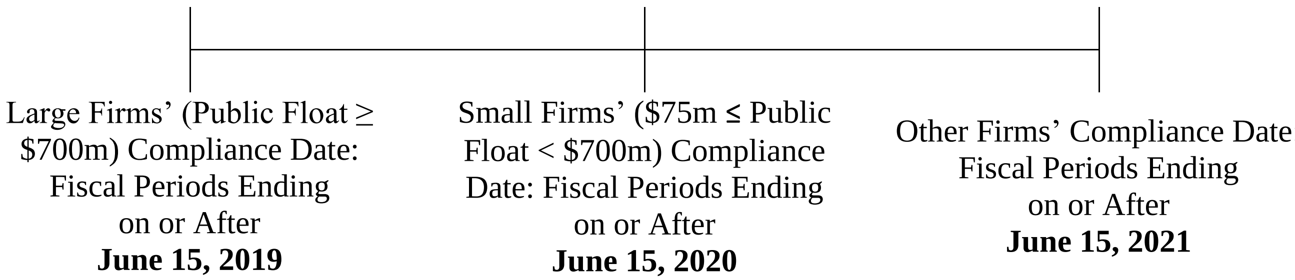
LIX	The LIX Readability Index, defined as (number of words / number of sentences) + (number of words over 6 letters * 100/ number of words). (Source: SEC EDGAR)
RIX	The RIX Readability Index, defined as (number of words with 7 characters or more) / (number of sentences). (Source: SEC EDGAR)
ARI	The Automated Readability Index, defined as 4.71 * (number of characters / number of words) +0.5 * (number of words / number of sentences) -21.43. (Source: SEC EDGAR)
SMOG	The SMOG Index, defined as 1.043 * sqrt(30 * number of words with more than two syllables / number of sentences) + 3.1291. (Source: SEC EDGAR)
Bog Index	The measure of human readability developed by Bonsall, Leone, Miller, and Rennekamp (2017). The plain English factors include sentence length, passive voice, weak verbs, overused words, complex words, and jargon. (Source: Website of Professor Brian P. Miller)

Appendix B. Sample Construction and Selection

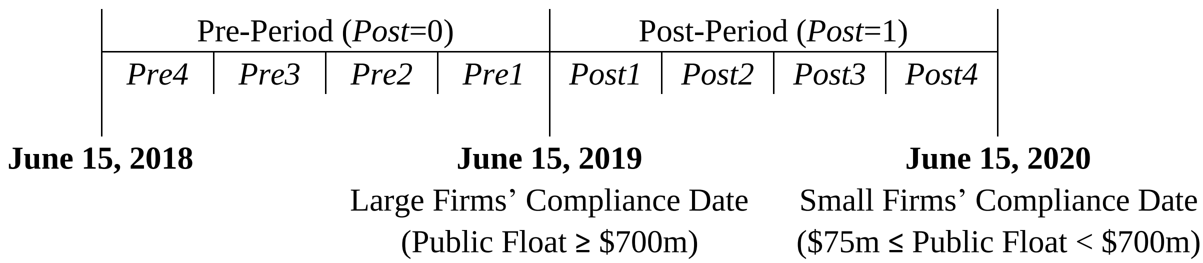
Description	**Observations**
Firm-quarter observations in Compustat with valid PERMNO-GVKEY-CIK identifiers	**41,127**
Delete: Observations without sufficient data for calculating variables in main analysis	(10,189)
Delete: Observations in firms whose filer category changes during the sample period	(2,960)
Delete: Observations in firms whose fiscal year-end is not in December	(5,729)
Delete: Observations in firms that voluntarily adopted iXBRL	(5,102)
Delete: Observations dropped from regressions because of fixed effects	(748)
Final sample	**16,399**

Figure 1
Timeline of Inline XBRL Regulation

Panel A. Timeline of Inline XBRL Compliance



Panel B. Timeline in Difference-in-Differences Design



Notes: Panel A illustrates the timeline of the Inline XBRL compliance. Panel B shows the timeline in our difference-in-differences design. *Post* equals 1 if the quarter is after June 15, 2019, and 0 otherwise. *Pre4*, *Pre3*, *Pre2*, *Pre1*, *Post1*, *Post2*, *Post3*, and *Post4* are equal to 1 if the fiscal quarter ends in June 2018, September 2018, December 2018, March 2019, June 2019, September 2019, December 2019, and March 2020, respectively, and 0 otherwise.

Table 1
Summary Statistics

Variable	N	Mean	SD	P25	P50	P75
Fog_Index	16,399	21.330	1.650	20.190	21.060	22.180
Treat	16,399	0.449	0.497	0	0	1
Post	16,399	0.501	0.500	0	1	1
Earnings	16,399	-0.024	0.091	-0.019	0.006	0.019
Loss	16,399	0.351	0.477	0	0	1
Size	16,399	6.409	2.254	4.720	6.350	8.046
MTB	16,399	2.123	1.990	1.026	1.349	2.337
Age	16,399	18.551	17.658	5.008	14.093	25.436
Special_Items	16,399	-0.004	0.015	-0.001	0	0
Ret_Vol	16,399	0.135	0.094	0.070	0.107	0.171
Earn_Vol	16,399	0.038	0.086	0.003	0.009	0.029
NBSeg	16,399	0.938	0.395	0.693	0.693	1.386
NGSeg	16,399	1.005	0.457	0.693	0.693	1.386
Nitems	16,399	5.624	0.070	5.580	5.620	5.666
SEO	16,399	0.034	0.181	0	0	0
MA	16,399	0.106	0.308	0	0	0
Delaware	16,399	0.561	0.496	0	1	1

Notes: This table provides descriptive statistics for the sample used in the main analyses. The variables are as defined in Appendix A, and all continuous variables are winsorized at 1% and 99%.

Table 2
Validation Test: Impact of Inline XBRL on Machine Readability

	Dependent Variable: Gap in Number of Words		
	WRDS vs. McDonald	McDonald vs. Self	WRDS vs. Self
	Gap1	*Gap2*	*Gap3*
	(1)	(2)	(3)
Treat×Post	**-0.072*****	**-0.011*****	**-0.088*****
	(-17.07)	**(-44.87)**	**(-20.72)**
Earnings	0.044	-0.003**	0.038
	(1.24)	(-2.41)	(1.09)
Loss	0.009*	-0.000*	0.009*
	(1.81)	(-1.67)	(1.74)
Size	-0.004	0.001***	-0.004
	(-1.14)	(4.11)	(-0.92)
MTB	0.004**	-0.000***	0.003**
	(2.46)	(-5.38)	(2.15)
Age	0.025***	-0.001***	0.024***
	(5.74)	(-3.73)	(5.44)
Special_Items	-0.028	-0.000	-0.032
	(-0.38)	(-0.05)	(-0.43)
Ret_Vol	0.006	-0.001	0.005
	(0.29)	(-0.96)	(0.23)
Earn_Vol	-0.029	-0.002	-0.030
	(-0.83)	(-1.51)	(-0.84)
NBSeg	0.025	-0.003*	0.020
	(1.20)	(-1.94)	(0.97)
NGSeg	0.001	0.001	0.003
	(0.12)	(1.22)	(0.27)
Nitems	0.039	0.014***	0.057
	(0.74)	(5.10)	(1.09)
SEO	-0.003	0.000	-0.003
	(-0.48)	(0.10)	(-0.44)
MA	0.009**	0.000	0.009**
	(2.08)	(0.79)	(2.09)
Delaware	-0.023	0.000	-0.023
	(-0.93)	(0.41)	(-0.93)
Firm FE	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes
No. of Obs.	13,708	16,399	13,708
Adj. R-squared	0.197	0.736	0.223

Notes: This table provides difference-in-differences estimates of the effect of Inline XBRL on machine readability. *Treat* equals 1 if the firm is a large firm (i.e., public float ≥ \$700 million), and 0 otherwise. *Post* equals 1 if the quarter is after June 15, 2019, and 0 otherwise. Coefficients for *Treat* and *Post* are subsumed by the firm and year-quarter fixed effects, respectively. *Gap1* is the absolute difference in numbers of words between WRDS SEC Analytics Suite and Bill McDonald's database divided by sum of numbers of words in the two sources. *Gap2* is the absolute difference in numbers of words between Bill McDonald's database and self-developed parsed filings divided by sum of numbers of words in the two sources. *Gap3* is the absolute difference in numbers of words between WRDS SEC Analytics Suite and self-developed parsed filings divided by sum of numbers of words in the two sources. Coefficients are provided with *t*-statistics in parentheses below. The sample consists of firm-quarter 10-K/Q filings for fiscal quarters from 2018Q2 to 2020Q1. Standard errors are corrected by clustering at the firm level. Variables are defined in Appendix A. Continuous variables are winsorized at the 1% and 99% levels. Constants are not tabulated because of fixed effects. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, using a two-tailed *t*-test.

Table 3
Effect of Machine Readability on Human Readability

	Dependent Variable: Human Readability (*Fog_Index*)	
	(1)	(2)
Treat×Post	**0.654*** **	**0.670*** **
	(12.28)	**(12.44)**
Earnings		0.081
		(0.30)
Loss		0.131***
		(2.75)
Size		-0.064*
		(-1.69)
MTB		0.011
		(0.80)
Age		0.268***
		(4.10)
Special_Items		-1.996***
		(-2.85)
Ret_Vol		0.018
		(0.09)
Earn_Vol		0.118
		(0.37)
NBSeg		0.107
		(0.48)
NGSeg		0.036
		(0.17)
Nitems		-0.890*
		(-1.73)
SEO		-0.054
		(-1.04)
MA		0.064
		(1.53)
Delaware		-0.137
		(-0.58)
Firm FE	Yes	Yes
Year-quarter FE	Yes	Yes
No. of Obs.	16,399	16,399
Adj. R-squared	0.468	0.470

Notes: This table provides difference-in-differences estimates of the effect of machine readability on human readability (*Fog_Index*), where Column (1) does not include control variables and Column (2) includes control variables. *Treat* equals 1 if the firm is a large firm (i.e., public float ≥ $700 million), and 0 otherwise. *Post* equals 1 if the quarter is after June 15, 2019, and 0 otherwise. Coefficients for *Treat* and *Post* are subsumed by the firm and year-quarter fixed effects, respectively. Coefficients are provided with *t*-statistics in parentheses below. The sample consists of 16,399 firm-quarter 10-K/Q filings for fiscal quarters from 2018Q2 to 2020Q1. Standard errors are corrected by clustering at the firm level. Variables are defined in Appendix A. Continuous variables are winsorized at the 1% and 99% levels. Constants are not tabulated because of fixed effects. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, using a two-tailed *t*-test.

Table 4
Effect of Machine Readability on Human Readability: Dynamic Difference-in-Differences

	Dependent Variable: Human Readability (*Fog_Index*)	
	(1)	(2)
Treat×Pre3	-0.038	-0.039
	(-0.59)	(-0.60)
Treat×Pre2	0.051	0.050
	(0.76)	(0.75)
Treat×Pre1	0.030	0.030
	(0.41)	(0.42)
Treat×Post	**0.682*****	
	(9.51)	
Treat×Post1		**0.479*****
		(5.45)
Treat×Post2		**0.577*****
		(6.49)
Treat×Post3		**0.623*****
		(8.06)
Treat×Post4		**1.081*****
		(11.61)
Control Variables	Included	Included
Firm FE	Yes	Yes
Year-quarter FE	Yes	Yes
No. of Obs.	16,399	16,399
Adj. R-squared	0.470	0.473

Notes: This table provides dynamic difference-in-differences estimates of the effect of machine readability on human readability. *Treat* equals 1 if the firm is a large firm (i.e., public float ≥ $700 million), and 0 otherwise. *Post* equals 1 if the quarter is after June 15, 2019, and 0 otherwise. *Pre4*, *Pre3*, *Pre2*, *Pre1*, *Post1*, *Post2*, *Post3*, and *Post4* are equal to 1 if the fiscal quarter is in June 2018, September 2018, December 2018, March 2019, June 2019, September 2019, December 2019, and March 2020, respectively, and 0 otherwise. *Pre4* is the benchmark, so it is omitted in the regression. Coefficients for *Treat, Post, Pre4*, *Pre3*, *Pre2*, *Pre1*, *Post1*, *Post2*, *Post3*, and *Post4* are subsumed by the firm and year-quarter fixed effects, respectively. Column (1) presents dynamic estimates for pre-treatment period, and Column (2) presents dynamic estimates for both the pre- and post-treatment periods. Coefficients are provided with *t*-statistics in parentheses below. The sample consists of 16,399 firm-quarter 10-K/Q filings for fiscal quarters from 2018Q2 to 2020Q1. Standard errors are corrected by clustering at the firm level. Variables are defined in Appendix A. Continuous variables are winsorized at the 1% and 99% levels. Constants are not tabulated because of fixed effects. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, using a two-tailed *t*-test.

Table 5
Capital Market Consequences to Retail Investors

Panel A. Consequences on Informativeness of Disclosures to Retail Investors

Dependent Variable:	*Abs_CAR*		*Abnormal_Volume*	
	(1)	(2)	(3)	(4)
Retail Investor Ownership:	High	Low	High	Low
Treat×Post	**-0.012*****	-0.006	**-0.259*****	-0.115
	-(3.41)	(-1.44)	**(-3.26)**	(-1.28)
Earnings	-0.020	-0.029	-0.563	0.115
	(-0.60)	(-0.60)	(-0.83)	(0.15)
Loss	-0.003	0.006	-0.266**	0.091
	(-0.72)	(1.22)	(-2.19)	(0.73)
Size	-0.010*	-0.002	-0.155	0.024
	(-1.65)	(-0.15)	(-1.23)	(0.14)
MTB	0.002	-0.001	-0.017	-0.048
	(1.12)	(-0.49)	(-0.52)	(-1.21)
Institutional Ownership	0.006	0.032	0.783	1.382**
	(0.22)	(1.27)	(1.37)	(2.31)
Analyst	0.011**	0.002	0.019	0.308**
	(2.00)	(0.31)	(0.16)	(2.30)
Abs_SUE	0.011	0.195**	-0.606	2.159
	(0.29)	(2.50)	(-0.80)	(1.30)
Leverage	0.008	-0.004	0.158	-0.036
	(0.58)	(-0.30)	(0.54)	(-0.13)
FileLag	-0.011***	-0.011***	-0.363***	-0.600***
	(-5.82)	(-7.91)	(-8.63)	(-15.24)
LnPrc	-0.004	-0.021**	0.129	0.008
	(-0.77)	(-2.17)	(1.21)	(0.05)
Return	0.008	0.002	0.172	-0.423
	(0.48)	(0.11)	(0.62)	(-1.07)
Firm FE	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes
No. of Obs.	6,052	6,955	6,053	6,963
Adj. R-squared	0.273	0.299	0.123	0.265
Difference in Coefficients	**-0.006****		**-0.144*****	
P-Value	0.030		0.004	

Panel B. Alternative Measures for Retail Investors

Dependent Variable:	*Ownership_ RetailInvestors*	*Number_ RetailInvestors*	*NetFlow_ RetailInvestors*
	(1)	(2)	(3)
Treat×Post	**-0.007****	**-0.054****	**-0.565****
	(-2.31)	**(-2.39)**	**(-2.32)**
Size	-0.057***	-0.029	-0.383
	(-12.16)	(-0.90)	(-1.57)
MTB	0.005***	-0.021***	0.070
	(4.30)	(-2.74)	(0.88)
Ret_Vol	0.077***	1.882***	0.588
	(3.80)	(8.77)	(0.38)
Beta	-0.011***	0.068***	-0.376
	(-4.22)	(2.75)	(-1.63)
PrInv	0.004	0.290***	0.523
	(1.18)	(6.30)	(1.23)
Return	0.103***	0.290***	-1.383*
	(10.28)	(4.14)	(-1.72)
SP500	0.012	-0.174*	-0.562
	(1.09)	(-1.75)	(-0.60)
Firm FE	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes
No. of Obs.	16,243	14,499	15,639
Adj. R-squared	0.979	0.960	0.160

Notes: This table provides the results of capital market consequences of increased machine readability to retail investors. Panel A presents difference-in-differences estimates of the effect of machine readability on the informativeness of disclosures around earnings announcements. We partition the sample based on whether the firm's retail investor ownership is high (i.e., above the industry-quarter medium). *Abs_CAR* is the absolute value of 5-day cumulated abnormal returns around earnings announcements. *Abnormal_Volume* is the mean daily trading volume during the event period minus the mean daily trading volume during the non-filing period, deflated by the standard deviation of daily trading volume during the non-filing period. Control variables are included following Blankespoor et al. (2014). Panel B presents difference-in-differences estimates of the effect of machine readability on retail investor ownership, the number of Robinhood shareholders, and net retail flow. *Ownership_RetailInvestors* is defined as one minus institutional ownership. *Number_RetailInvestors* is defined as the average number of daily Robinhood shareholders in the fiscal quarter, where Robinhood is a brokerage that offers commission-free trading to retail investors. *NetFlow_RetailInvestors* is defined as the average number of 10-day moving average retail net flows in the fiscal quarter, where a 10-day moving average retail net flow is calculated as 100*(buy flow – sell flow)/(buy flow + sell flow) in the most recent 10 trading days. Control variables are included following Hong and Kacperczyk (2009). The empirical *p*-value for the difference in coefficients is estimated through a bootstrapping procedure with 1,000 repetitions. Standard errors are corrected by clustering at the firm level. Variables are defined in Appendix A. Continuous variables are winsorized at the 1% and 99% levels. Constants are not tabulated because of fixed effects. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, using a two-tailed *t*-test.

Table 6
Cross-Sectional Variation in Effect of Machine Readability on Human Readability: Incentives and Opportunities

	Dependent Variable: Human Readability (*Fog_Index*)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Confidential Treatment		Firm Performance		Earnings Management		Monitoring	
	Yes	No	Loss	Non-Loss	High	Low	Low	High
Treat×Post	**0.736*** **	**0.665*** **	**0.823*** **	**0.641*** **	**0.745*** **	**0.598*** **	**0.658*** **	0.335
	(5.42)	**(11.32)**	**(6.06)**	**(10.13)**	**(7.79)**	**(7.20)**	**(4.03)**	(1.30)
Difference in Coefficients	**0.071*** **		**0.182*** **		**0.147** **		**0.323*** **	
P-Value	<0.001		<0.001		0.022		0.007	
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of Obs.	2,660	13,739	5,566	10,551	5,614	5,387	2,908	2,244
Adj. R-squared	0.450	0.445	0.551	0.423	0.465	0.449	0.403	0.357

Notes: This table presents results of the effect of machine readability on human readability conditional on firms' incentives and opportunities to decrease human readability. In Columns (1) and (2), we partition the sample based on whether the firm has confidential treatment of proprietary information, i.e., whether the firm issues at least one Form CT Order from 2018Q2 to 2019Q1 (pre-treatment period). In Columns (3) and (4), we partition the sample based on whether the firm is experiencing a loss. In Columns (5) and (6), we partition the sample based on whether the firm's earnings management exceeds the industry-quarter median, where earnings management is measured by the absolute value of discretionary accruals estimated from the performance-matched modified Jones model (Kothari et al. 2005). In Columns (7) and (8), we partition the sample based on whether the firm has high monitoring, measured by whether the firm's percentage of independent directors exceeds the industry-quarter median. Difference-in-differences estimates are provided with *t*-statistics in parentheses below. Standard errors are corrected by clustering at the firm level. The empirical *p*-value for the difference in coefficients is estimated through a bootstrapping procedure with 1,000 repetitions. Variables are defined in Appendix A. Continuous variables are winsorized at the 1% and 99% levels. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, using a two-tailed *t*-test.

Table 7
Cross-Sectional Variation in Effect of Machine Readability on Human Readability: Workload and Firm Complexity

	Dependent Variable: Human Readability (*Fog_Index*)					
	(1)	(2)	(3)	(4)	(5)	(6)
	Accounting Job Postings		Computer Job Postings		Firm Complexity	
	No	Yes	No	Yes	High	Low
Treat×Post	**0.749*** **	**0.631*** **	**0.759*** **	**0.631*** **	**0.721*** **	**0.606*** **
	(8.54)	**(7.58)**	**(7.38)**	**(8.24)**	**(9.21)**	**(7.20)**
Difference in Coefficients	**0.118***		**0.128***		**0.115***	
P-Value	0.094		0.086		0.099	
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
No. of Obs.	9,748	6,531	8,532	7,762	7,509	7,776
Adj. R-squared	0.528	0.404	0.541	0.415	0.417	0.520

Notes: This table presents results of the effect of machine readability on human readability conditional on accounting or computer job postings and firm complexity. In Columns (1) and (2), we partition the sample based on whether the firm has accounting job postings from the iXBRL adoption date to the compliance date. In Columns (3) and (4), we partition the sample based on whether the firm has computer job postings from the iXBRL adoption date to the compliance date. In Columns (5) and (6), we partition the sample based on firm complexity, where the measure of firm complexity is a machine-learning based measure developed by Loughran and McDonald (2023). Difference-in-differences estimates are provided with *t*-statistics in parentheses below. Standard errors are corrected by clustering at the firm level. The empirical *p*-value for the difference in coefficients is estimated through a bootstrapping procedure with 1,000 repetitions. Variables are defined in Appendix A. Continuous variables are winsorized at the 1% and 99% levels. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, using a two-tailed *t*-test.

Table 8
Effect of Machine Readability on Human Readability:
Regression Discontinuity Design (RDD)

	Dependent Variable: Human Readability (*Fog_Index*)	
	(1)	(2)
Discontinuity	**0.659****	**0.842****
	(2.23)	**(2.16)**
Nonparametric Estimator	Yes	Yes
Polynomials	2nd Order	3rd Order
Control Variables	Yes	Yes
No. of Obs.	7,497	7,497

Notes: This table reports the regression discontinuity nonparametric estimates and *t*-statistics (in parentheses) for the discontinuity in human readability around the public float threshold of large firms. The table reports bias-corrected RD estimates with a bandwidth-robust variance estimator using triangular kernel and clustering with plug-in residuals at the running variable level (public float). The regressions include control variables used in the main analysis. The optimal bandwidth selection is based on a second-generation plug-in bandwidth selection approach and is covariate-adjusted and cluster-robust. Column (1) reports the estimator using the second-order polynomials, and Column (2) reports the estimator using the third-order polynomials. All models have firm-clustered, robust standard errors. Variables are defined in Appendix A. Continuous variables are winsorized at the 1% and 99% levels. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, using a two-tailed *t*-test.

Table 9
Alternative Difference-in-Differences Design

	Dependent Variable: Human Readability (*Fog_Index*)	
	(1)	(2)
Treat_Alternative×Post	**0.295*****	**0.306*****
	(3.88)	**(4.06)**
Earnings		-0.848
		(-0.87)
Loss		0.169*
		(1.78)
Size		-0.106
		(-1.23)
MTB		0.002
		(0.07)
Age		0.282***
		(5.21)
Special_Items		-1.644
		(-1.27)
Ret_Vol		-0.452
		(-0.62)
Earn_Vol		-0.120
		(-0.08)
NBSeg		0.208
		(0.64)
NGSeg		-0.100
		(-0.30)
Nitems		-0.598
		(-0.87)
SEO		-0.030
		(-0.27)
MA		0.086*
		(1.78)
Delaware		-0.727**
		(-2.16)
Firm FE	Yes	Yes
Year-quarter FE	Yes	Yes
No. of Obs.	10,003	10,003
Adj. R-squared	0.397	0.399

Notes: This table provides the results of alternative DiD Design by utilizing the voluntary adoption of iXBRL. *Treat_Alternative* equals 1 if large firms initiate iXBRL filings after the treatment date, i.e., June 15, 2019 (mandatory adopters), and 0 if large firms voluntarily adopt iXBRL before the treatment date (voluntary adopters). *Post* equals 1 if the fiscal quarter is after June 15, 2019, and 0 otherwise. Coefficients for *Treat* and *Post* are subsumed by the firm and year-quarter fixed effects, respectively. Coefficients are provided with *t*-statistics in parentheses below. The sample consists of 10,003 firm-quarter 10-K/Q filings for fiscal quarters from 2018Q2 to 2020Q1. The sample only includes large firms. Standard errors are corrected by clustering at the firm level. Variables are defined in Appendix A and are winsorized at the 1% and 99% levels. Intercepts are not reported because of fixed effects. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, using a two-tailed *t*-test.

Table 10
Placebo Tests

	Dependent Variable: Human Readability (*Fog_Index*)	
	(1)	(2)
Treat×Post_Placebo	0.030	
	(0.90)	
Treat_Placebo×Post		0.047
		(1.11)
Earnings	-0.145	0.175
	(-0.63)	(0.63)
Loss	0.082**	0.121**
	(2.06)	(2.43)
Size	-0.020	-0.077**
	(-0.53)	(-2.03)
MTB	0.002	0.023*
	(0.16)	(1.68)
Age	-0.075*	0.044
	(-1.79)	(0.13)
Special_Items	-1.684**	-1.389*
	(-2.30)	(-1.77)
Ret_Vol	0.137	-0.086
	(0.80)	(-0.42)
Earn_Vol	0.124	-0.053
	(0.60)	(-0.16)
NBSeg	0.108	0.118
	(1.24)	(0.59)
NGSeg	-0.182	0.050
	(-1.54)	(0.29)
Nitems	-0.611	-0.996
	(-1.49)	(-1.60)
SEO	0.002	-0.043
	(0.03)	(-0.85)
MA	0.159***	0.105*
	(4.56)	(1.81)
Delaware	-0.327**	0.185
	(-2.16)	(0.94)
Firm FE	Yes	Yes
Year-quarter FE	Yes	Yes
No. of Obs.	16,641	9,041
Adj. R-squared	0.481	0.559

Notes: This table reports results of two falsification tests based on a placebo treatment date in Column (1) and a placebo treatment group in Column (2). In Column (1), the sample consists of 16,641 10-K/Q filings for fiscal quarters from 2016Q2 to 2018Q1. *Treat* equals 1 if the firm is a large firm (i.e., public float ≥ $700 million), and 0 otherwise, which is the same as *Treat* used in the main analysis. *Post_Placebo* equals 1 if the fiscal quarter is after June 15, 2017 (two years before the actual treatment date), and 0 otherwise. In Column (2), the sample consists of 9,041 10-K/Q filings for fiscal quarters from 2018Q2 to 2020Q1. *Treat_Placebo* equals 1 if the firm is a small firm, who does not receive treatment during the sample period, and 0 otherwise. We exclude large firms, who receive treatment during the sample period, in Panel B. *Post* equals 1 if the quarter is after June 15, 2019, and 0 otherwise, which is the same as *Post* used in the main analysis. Coefficients for *Treat, Treat_Placebo, Post*, and *Post_Placebo* are subsumed by the firm and year-quarter fixed effects, respectively. Coefficients are provided with *t*-statistics in parentheses below. Standard errors are corrected by clustering at the firm level. Variables are defined in Appendix A. Continuous variables are winsorized at the 1% and 99% levels. Constants are not tabulated because of fixed effects. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, using a two-tailed *t*-test.

Table 11
Alternative Measures of Human Readability

	Dependent Variable: Alternative Measures of Human Readability					
	Flesch-Kincaid	*LIX*	*RIX*	*ARI*	*SMOG*	*Bog Index*
Treat×Post	**0.676*** **	**2.016*** **	**0.647*** **	**0.947*** **	**0.489*** **	**0.028*** **
	(12.50)	**(14.63)**	**(13.77)**	**(13.30)**	**(12.95)**	**(11.49)**
Earnings	0.073	0.226	0.115	0.140	0.070	-0.013
	(0.26)	(0.33)	(0.48)	(0.38)	(0.38)	(-0.68)
Loss	0.126***	0.285**	0.098**	0.162**	0.092***	-0.003
	(2.60)	(2.38)	(2.41)	(2.54)	(2.88)	(-0.91)
Size	-0.072*	-0.168*	-0.060*	-0.094*	-0.052**	0.001
	(-1.89)	(-1.76)	(-1.86)	(-1.85)	(-2.07)	(0.60)
MTB	0.014	0.034	0.014	0.022	0.009	-0.001
	(1.00)	(1.02)	(1.17)	(1.24)	(0.93)	(-1.11)
Age	0.265***	0.659***	0.225***	0.351***	0.172***	-0.011
	(4.24)	(4.71)	(4.22)	(4.40)	(3.75)	(-0.96)
Special_Items	-2.212***	-5.754***	-1.798***	-2.935***	-1.380***	-0.019
	(-3.13)	(-3.35)	(-3.06)	(-3.20)	(-2.89)	(-0.32)
Ret_Vol	0.070	0.045	0.059	0.129	0.040	-0.007
	(0.34)	(0.09)	(0.34)	(0.48)	(0.29)	(-0.52)
Earn_Vol	0.071	0.076	0.033	0.058	0.101	-0.011
	(0.22)	(0.10)	(0.12)	(0.14)	(0.48)	(-0.70)
NBSeg	0.084	0.186	0.109	0.131	0.072	0.021
	(0.37)	(0.32)	(0.54)	(0.45)	(0.44)	(0.92)
NGSeg	0.006	-0.024	-0.020	-0.055	0.026	-0.033
	(0.03)	(-0.04)	(-0.11)	(-0.20)	(0.16)	(-0.90)
Nitems	-0.757	-2.904**	-0.590	-1.026	-0.656*	0.083**
	(-1.46)	(-2.29)	(-1.34)	(-1.52)	(-1.84)	(2.17)
SEO	-0.047	-0.075	-0.019	-0.048	-0.029	-0.003
	(-0.90)	(-0.57)	(-0.42)	(-0.69)	(-0.79)	(-0.49)
MA	0.070	0.189*	0.068*	0.110*	0.039	0.000
	(1.63)	(1.81)	(1.85)	(1.93)	(1.36)	(0.05)
Delaware	-0.115	-0.310	-0.142	-0.189	-0.092	-0.001
	(-0.47)	(-0.49)	(-0.67)	(-0.58)	(-0.60)	(-0.06)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	No
Year FE	No	No	No	No	No	Yes
No. of Obs.	16,399	16,399	16,399	16,399	16,399	3,820
Adj. R-squared	0.462	0.493	0.477	0.430	0.516	0.869

Notes: This table provides difference-in-differences estimates of the effect of machine readability on human readability using alternative measures of human readability. *Flesch-Kincaid* is the Flesch-Kincaid Readability Index defined as 0.39 * (number of words / number of sentences) + 11.8* (number of syllables / number of words) - 15.59. *LIX* is the LIX Readability Index, defined as (number of words / number of sentences) + (number of words over 6 letters * 100/ number of words). *RIX* is the RIX Readability Index, defined as (number of words with 7 characters or more) / (number of sentences). *ARI* is the Automated Readability Index, defined as 4.71 * (number of characters / number of words) +0.5 * (number of words / number of sentences) -21.43. *SMOG* is the SMOG Index, defined as 1.043 * sqrt(30 * number of words with more than two syllables / number of sentences) + 3.1291. *Bog Index* is a measure of human readability to capture the plain English attributes of 10-K filings and is downloaded from Professor Brian Miller's website https://host.kelley.iu.edu/bpm/activities/bogindex.html. *Treat* equals 1 if the firm is a large firm (i.e., public float ≥ $700 million), and 0 otherwise. *Post* equals 1 if the quarter is after June 15, 2019, and 0 otherwise. Coefficients for *Treat* and *Post* are subsumed by the firm and year-quarter fixed effects, respectively. Coefficients are provided with *t*-statistics in parentheses below. The sample consists of 16,399 firm-quarter 10-K/Q filings for fiscal quarter 2018Q2-2020Q1. Standard errors are corrected by clustering at the firm level. Variables are defined in Appendix A. Continuous variables are winsorized at the 1% and 99% levels. Intercepts are not reported because of fixed effects. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, using a two-tailed *t*-test.

Online Appendix for "Human Readability of Disclosures in a Machine-Readable World"

Contents

Online Appendix A
Examples of How Inline XBRL Improves Machine Readability

1. Reads Texts in Filings More Accurately

Before iXBRL adoption, General Electric Corp.'s 2018 10-K filing used HTML <table> tags in the main filing to identify the text beneath tables, which machine users typically rely upon when removing tables before extracting information from filings. However, this approach risks machines overlooking crucial footnoted information nested within such tags, as show in the following image.

```
<table><tr>
    <td><div><font>(a)</font></div></td>
    <td><div><font>Power segment revenues represent 24% and 22% of total industrial segment revenues and total segment
    revenues, respectively, for the year ended December 31, 2018.</font></div></td>
</tr></table>
```

The iXBRL mandate resolves this issue by changing the filing format from HTML to XHTML, which prohibits the nesting of HTML table elements. In General Electric's subsequent 2019 10-K filing, which uses XHTML format following the adoption of iXBRL, the footnotes in question are properly enclosed only by <div> tags rather than <table>, ensuring their machine-readable inclusion and preservation of important contextual details.

```
<div><span>(a) Power segment revenues represent </span><span>21%</span><span> and </span><span>19%</span><span> of total
industrial segment revenues and total segment revenues, respectively, for the year ended </span><span>December 31, 2019
</span><span>.</span></div><div><span>(b) Power segment profit represents </span><span>4%</span><span> of total industrial
segment profit for the year ended </span><span>December 31, 2019</span><span>.</span></div>
```

2. Provides Users with Context for Machine-readable Content

Before iXBRL adoption, Coherent Inc.'s 2019Q1 10-Q filing assigned an extremely lengthy custom tag to a line item in its income statement: "cohr:Incomefromoperationsbeforeotherincomeincometaxesandlossfromdiscontinuedoperations." As seen in the following image, such a long and peculiar tag makes discerning the item's true meaning from the tag alone quite difficult. The unstructured, random ordering of line items in traditional XBRL exhibits exacerbates this interpretability issue for machine readers.

```
<cohr:Incomefromoperationsbeforeotherincomeincometaxesandlossfromdiscontinuedoperations contextRef="FD2019Q1YTD" decimals=
"-3" id="Fact-4B9F780B139865F5DB2039EF0A23A80C" unitRef="usd">52811000
</cohr:Incomefromoperationsbeforeotherincomeincometaxesandlossfromdiscontinuedoperations>
```

As shown in the following image of the XHTML file embedded with iXBRL tags, machines can locate the exact position of the custom tag in the income statement and read its heading: “Income (loss) from operations.” This contextual embedding and reconciliation of tags increases machine comprehension of financial statements across different reporting entities over time.

```
<td><div>
    <span>Income (loss) from operations</span></div></td>
<td><div>
    <span><span>(
        <ix:nonFraction id="d57302743e1420-wk-Fact-26AB3ADCD0D9F0F4232DE8361C327341" name=
        "cohr:Incomefromoperationsbeforeotherincomeincometaxesandlossfromdiscontinuedoperations" contextRef="FD2019Q3QTD"
        unitRef="usd" decimals="-3" scale="3" sign="-" format="ixt:numdotdecimal">
            7,157</ix:nonFraction></span></span></div></td>
<td><div><span>)</span></div></td>
```

3. Mitigates Errors in Reading Numbers

Before iXBRL adoption, Apex Resources Inc.’s 2018Q3 10-Q filing documents “Net Loss including noncontrolling interests” differently between the main filing (in a HTML file) and the XBRL exhibit. The former states the figure as 171,099 while the latter reports it as -171,099. The image below visually depicts this divergence between the HTML and XBRL representations of this key line item. Accordingly, machine users relying solely on the XBRL exhibit will derive an incorrect interpretation of the company’s financial status.

Net Loss including noncontrolling interests	171,099	(1,339)

```
<us-gaap:IncomeLossFromContinuingOperationsIncludingPortionAttributableToNoncontrollingInterest contextRef=
"From2018-07-01to2018-09-30" unitRef="USD" decimals="0">-171099
</us-gaap:IncomeLossFromContinuingOperationsIncludingPortionAttributableToNoncontrollingInterest>
```

Such discrepancies are rectified with iXBRL because both the XBRL and HTML content are unified into a single datapoint, greatly mitigating the potential for errors.

Online Appendix B
Alternative Samples

	Dependent Variable: Human Readability (*Fog_Index*)					
	(1)	(2)	(3)	(4)	(5)	(6)
Treat×Post	**0.706*****	**0.566*****	**0.650*****	**0.673*****	**0.602*****	**0.549*****
	(11.41)	**(9.40)**	**(13.56)**	**(12.21)**	**(11.16)**	**(5.13)**
Earnings	0.141	-0.302	-0.096	-0.055	-0.255	-1.000
	(0.42)	(-0.79)	(-0.37)	(-0.17)	(-0.74)	(-1.25)
Loss	0.096	-0.044	0.119***	0.103**	-0.058	0.053
	(1.57)	(-0.55)	(2.88)	(1.98)	(-0.83)	(0.65)
Size	-0.105**	-0.021	-0.066**	-0.102**	-0.017	-0.037
	(-2.33)	(-0.38)	(-1.97)	(-2.52)	(-0.35)	(-0.46)
MTB	0.011	-0.001	0.007	0.010	-0.007	-0.015
	(0.70)	(-0.03)	(0.57)	(0.68)	(-0.37)	(-0.55)
Age	0.275***	-0.186	0.258***	0.265***	-0.143	-0.528
	(4.86)	(-0.77)	(3.55)	(4.02)	(-0.65)	(-1.04)
Special_Items	-2.711***	0.017	-1.707***	-2.665***	0.892	-0.656
	(-2.78)	(0.02)	(-2.72)	(-2.98)	(0.93)	(-0.48)
Ret_Vol	-0.026	0.044	0.110	0.057	0.383	0.769**
	(-0.10)	(0.16)	(0.61)	(0.26)	(1.56)	(2.04)
Earn_Vol	0.260	-0.119	0.153	0.320	-0.111	0.520
	(0.65)	(-0.34)	(0.51)	(0.83)	(-0.35)	(0.59)
NBSeg	0.056	0.393	0.047	0.007	0.390	-0.159
	(0.22)	(1.37)	(0.23)	(0.03)	(1.48)	(-0.42)
NGSeg	0.047	0.043	-0.007	0.029	-0.032	-0.100
	(0.21)	(0.10)	(-0.04)	(0.15)	(-0.10)	(-0.49)
Nitems	0.931	1.193	-0.485	1.147	1.858**	1.266
	(1.21)	(1.22)	(-1.03)	(1.62)	(2.18)	(1.13)

SEO	-0.052	0.101	-0.048	-0.039	0.016	0.002
	(-0.84)	(0.82)	(-0.99)	(-0.67)	(0.14)	(0.01)
MA	0.028	0.102	0.067*	0.040	0.104	0.060
	(0.51)	(1.18)	(1.73)	(0.80)	(1.34)	(0.66)
Delaware	-0.164	0.018	-0.135	-0.153	0.061	-0.236
	(-0.52)	(0.07)	(-0.63)	(-0.56)	(0.27)	(-0.46)
Include 10-Q	Yes	No	Yes	Yes	No	Yes
Include 10-K	No	Yes	Yes	No	Yes	Yes
Include December Fiscal Year-End	Yes	Yes	Yes	Yes	Yes	No
Include Non-December Fiscal Year-End	No	No	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
No. of Obs.	12,185	3,846	20,330	15,121	4,788	3,930
Adj. R-squared	0.459	0.523	0.467	0.457	0.532	0.448

Notes: This table provides difference-in-differences estimates of the effect of machine readability on human readability using alternative samples. *Treat* equals 1 if the firm is a large firm (i.e., public float ≥ $700 million), and 0 otherwise. *Post* equals 1 if the quarter is after June 15, 2019, and 0 otherwise. Coefficients for *Treat* and *Post* are subsumed by the firm and year-quarter fixed effects, respectively. Coefficients are provided with *t*-statistics in parentheses below. In the main analysis, we use 10-Q/K filings with December fiscal year-end. In Column (1), we only include 10-Q filings. In Column (2), we only include 10-K filings. In Columns (3)-(5), we add back filings with non-December fiscal year-end. In Column (6), we only include filings with non-December fiscal year-end. Standard errors are corrected by clustering at the firm level. Variables are defined in Appendix A. Continuous variables are winsorized at the 1% and 99% levels. Intercepts are not reported because of fixed effects. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, using a two-tailed *t*-test.

Online Appendix C
Alternative Methods to Handle Tables

	Fog_Alternative1		*Fog_Alternative2*	
	(1)	(2)	(3)	(4)
Treat×Post	**0.835*****	**0.833*****	**1.454*****	**1.455*****
	(26.88)	**(26.74)**	**(37.51)**	**(37.34)**
Earnings		-0.120		-0.263
		(-0.54)		(-1.02)
Loss		0.063**		0.075**
		(2.23)		(2.20)
Size		-0.003		-0.011
		(-0.14)		(-0.44)
MTB		-0.001		0.011
		(-0.10)		(0.95)
Age		0.053***		0.085***
		(3.49)		(4.28)
Special_Items		-0.748*		-0.792*
		(-1.69)		(-1.67)
Ret_Vol		-0.001		-0.032
		(-0.01)		(-0.16)
Earn_Vol		0.562***		0.412*
		(2.79)		(1.87)
NBSeg		0.299*		0.295
		(1.79)		(1.49)
NGSeg		-0.032		-0.128
		(-0.16)		(-0.51)
Nitems		0.414		-0.356
		(1.26)		(-0.95)
SEO		-0.043		-0.042
		(-1.14)		(-1.05)
MA		0.020		0.022
		(1.08)		(1.06)
Delaware		0.029		-0.033
		(0.21)		(-0.22)
Firm FE	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes
No. of Obs.	16,399	16,399	16,399	16,399
Adj. R-squared	0.851	0.851	0.860	0.860

Notes: This table provides difference-in-differences estimates of the effect of machine readability on human readability using the Fog index based on filings parsed by two alternative methods to handle tables. In Columns (1) and (2), we calculate the Fog index (*Fog_Alternative1*) based on 10-K/Q filings where we delete all HTML tags through regular expression r'\<.*?' and r\<\/.*?\>. In Columns (3) and (4), we calculate the Fog index (*Fog_Alternative2*) based on 10-K/Q filings where we delete all HTML tags through regular expression r'\<.*?' and r\<\/.*?\> and delete sentences containing fewer than five words. *Treat* equals 1 if the firm is a large firm (i.e., public float ≥ $700 million), and 0 otherwise. *Post* equals 1 if the quarter is after June 15, 2019, and 0 otherwise. Coefficients for *Treat* and *Post* are subsumed by the firm and year-quarter fixed effects, respectively. Coefficients are provided with *t*-statistics in parentheses below. The sample consists of 16,399 firm-quarter 10-K/Q filings for fiscal quarter 2018Q2-2020Q1. Standard errors are corrected by clustering at the firm level. Variables are defined in Appendix A. Continuous variables are winsorized at the 1% and 99% levels. Intercepts are not reported because of fixed effects. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, using a two-tailed *t*-test.