DISCLOSURE SALIENCE IN EARNINGS CONFERENCE CALLS AND INVESTOR INFORMATION PROCESSING

by

KARSON FRONK

(Under the Direction of John L. Campbell)

ABSTRACT

Firms and third parties disclose a substantial amount of information during the earnings announcement window. Managers can alleviate information processing costs for investors by presenting this information using a variety of formats. Using machine learning, I identify the use of data visualizations, tables, and text in earnings conference call slides. I examine the association between the format of the earnings call presentation and capital market outcomes. I find the use of data visualizations in earnings conference calls is associated with decreases in information asymmetry, increases in market liquidity, and lower investor processing costs, specifically for retail investors. Additionally, I find no association between retail trading consensus and disclosure format types. Overall, my results suggest that managers' choices of disclosure formats in earnings conference calls can improve information saliency to reduce investors' processing costs.

INDEX WORDS: Data Visualization; Earnings Conference Calls; Investor Processing Costs; Machine Learning; Voluntary Disclosure

DISCLOSURE SALIENCE IN EARNINGS CONFERENCE CALLS AND INVESTOR INFORMATION PROCESSING

by

KARSON FRONK

Brigham Young University, MAcc, BS, 2018

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

DISCLOSURE SALIENCE IN EARNINGS CONFERENCE CALLS AND INVESTOR

INFORMATION PROCESSING

by

KARSON FRONK

Major Professor:	John Campbell
Committee:	Theodore Christensen
	Santhosh Ramalingegowda

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2023

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Disclosure salience in earnings conference calls and investor information processing

1. Introduction

The days surrounding earnings announcements often include a substantial flow of information from both the firm and third parties, such as financial analysts and the business media. Although much of this information may be value relevant, its sheer volume can lead to adverse market outcomes, as investors with limited processing capacity attempt to acquire and analyze the information (Blankespoor et al. 2020). Specifically, they may not process relevant information if it is presented in formats that are harder to parse or less salient (Hirshleifer and Teoh 2003; Bloomfield 2002), potentially leading to negative capital market outcomes, such as decreased liquidity, increased volatility, and inaccurate valuation (Chapman et al. 2019). To help investors, firms have options regarding how saliently they disclose information, such as altering or simplifying the presentation format.

I investigate variation in earnings conference call presentations to examine the relationship between presentation format (i.e., data visualizations, tables, or text) and proxies for outcomes related to investor processing costs (i.e., information asymmetry, market liquidity, and trading volume).[1] I focus on three related research questions. First, is presentation format associated with lower information asymmetry and thus higher market liquidity? Second, is it associated with decreased investor processing costs, particularly for less-sophisticated investors? Finally, is it associated with more consistent trading by less-sophisticated investors?

[1] Lee (1992) and Bamber (1987) document extremely high trading for several hours following the earnings announcement. For this reason and the difficulty in identifying when the slideshow is disclosed (see section 2.1), I analyze two-day event windows.

Answering these questions matters not only to researchers but also to regulators, managers, and auditors. Most accounting and finance research on disclosure focuses on the textual aspects of firms' disclosures (Bonsall et al. 2017; Li 2008). However, information can be disclosed in nontextual forms, and individuals are affected not only by the information they receive but also by how they receive it (Cheng et al. 2021; Chen et al. 2021; Huang et al. 2018). Moreover, regulators, managers, and auditors claim investors can benefit from visual disclosure. The SEC's *Plain English Handbook* advises companies to use data visualizations (e.g., bar charts, line graphs, pie graphs, and maps) and tables to clearly communicate information (SEC 1998). Additionally, one of the SEC's goals is to protect retail investors. To this end, the agency has implemented numerous disclosure standards to make firm disclosures' more understandable for all types of investors. [2] Similarly, auditors argue that charts, graphs, tables, and infographics increase the salience of disclosures (EY 2017; EY 2014) and help investors make better decisions (Diaz 2021). Finally, anecdotal discussions with investor relations personnel suggest that managers include data visualizations to improve investor comprehension. Whether data visualizations are associated with lower investor processing costs is an empirical question.

Research on disclosure presentation format is limited, potentially because of data constraints. I overcome these constraints using a combination of a novel data source, novel data extraction techniques, and a machine learning model. To construct my sample, I begin with the universe of earnings conference calls from WRDS Capital IQ with matching firm identifiers from Compustat, CRSP, I/B/E/S, and TAQ between 2010 and 2020.[3] I then download all available

[2] For example, opening earnings calls to the public, broader 8-K disclosure requirements, and XBRL tagging.
[3] WRDS CapitalIQ transcripts coverage begins in 2008. However, slideshow collection appears under collected prior to 2010. In robustness analysis, I re-estimate my results including 2008 and 2009 and find consistent inferences.

earnings conference call presentation slideshows from Capital IQ's website (i.e., capitaliq.com). Next I use Python to convert each slide into an image and extract the text from the slides. I then create a custom image recognition machine learning model to identify alternative data visualization types (i.e., bar charts, line graphs, pie graphs, and maps) on each slide. I then use information on the slides to determine whether the slide contains a numeric table. Finally, I classify all nondata-visualization and nontabular slides as textual slides.[4] Importantly, these data allow for cross-sectional variation in disclosure presentation format as well as variation in information processing costs, going from less costly (data visualization) to more (text).

My first research question examines whether disclosure presentation format is associated with lower information asymmetry and thus higher market liquidity. To answer this question, I separately regress measures of abnormal information asymmetry and market liquidity on disclosure format type (i.e., the number slides with data visualizations, tables, or text), holding constant the earnings news, call characteristics, concurrent news events (both from the firm and third parties), the timing of the call, and other firm characteristics as well as industry, year-quarter, and day-of-week fixed effects. I find a negative association between data visualization slides and abnormal information asymmetry. However, I do not find a relation between table or text slides and abnormal information asymmetry. Moreover, slides with a data visualization are associated with greater market liquidity, suggesting that more salient disclosures improve liquidity. Again I find no association between abnormal market liquidity and either table or text slides. These

[4] Based on my classification algorithm, slides that I classify as text-only may still contain relevant images, such as product images, flow charts, or product timelines, because the machine-learning algorithm cannot identify them. To provide evidence on the extent to which this occurs, I manually examine a random sample of 300 text slides and find that only 10.67% have an additional relevant image.

findings suggest that data presented in the more salient formats reduces information asymmetry and increases market liquidity.

My second research question examines whether disclosure presentation format is associated with decreased investor processing costs, particularly for less-sophisticated investors. To examine this question, I re-estimate my first model and vary the dependent variable between overall abnormal trading volume and abnormal retail trading volume, following Boehmer et al. (2021).[5] Consistent with contemporaneous research (Xu 2021), I find evidence of a positive association between data visualization disclosure and abnormal *overall* trading volume but no relation between either table or text slides and abnormal trading volume. However, I find this result is not robust to alternative specifications. Additionally, I find robust evidence that data visualization and table slides are positively associated with abnormal *retail* trading volume. I do not find an association between abnormal retail trading volume and text slides. Overall these results suggest that market-wide investor processing costs are not influenced by different disclosure formats but that *retail* investor processing costs are lower for firms that use more data visualizations in their disclosures.

My third research question examines whether disclosure presentation format is associated with more trading consensus by less-sophisticated investors (i.e., the extent to which these investors agree to buy or sell a stock in aggregate). I regress retail trading consensus on data

[5] I do not test the difference between retail investor trading and another group of traders, due to the difficulty in identifying alternative groups of traders. Boehmer et al. (2021) argue that their method contains low type 1 error (false positive) but high type 2 error (false negative). Thus I cannot use nonretail trades to proxy for institutional traders, due to the high false negative. Another approach (i.e., trade size) contains increasing measurement error, as larger investors increasingly break up their trades into smaller trades to disguise their activity (Puckett and Yan 2011; Campbell et al. 2009). Consistent with these concerns, in an untabulated analysis, I find either (1) no statistically significant differences between abnormal retail trading and abnormal nonretail trading or (2) a stronger associations for nonretail traders and abnormal trading volume, relative to retail traders.

presentation type—including the same set of controls as used in the previous models. I find no association between retail trading consensus and any of the data presentation types. The results to my second and third research questions suggest that data visualizations decrease retail investors' processing costs *without* altering their tendency to either buy or sell the stock more in aggregate.

In supplemental analyses, I address concerns that the results are due to the information itself rather than its presentation. First, I separate out the data visualization by type because information in bar charts and line graphs is largely pulled from the financial statements while pie graphs and maps are ratios and locations, respectively. I find that the results are primarily attributable to bar charts and line graphs, suggesting that presenting financial statement information visually may lower information processing costs. Second, I separate data visualization slides by the number of visualizations on each slide (i.e., one, two, three, or four or more). I find the results are concentrated among slides with one or two data visualizations, suggesting that the associated reduction in processing costs pertains to slides with less visual information to process. Finally, I create word clouds of each type slide and find that the majority of information relates to earnings or finance. In other words, my results are driven by financial information that is disclosed largely for all firms, mitigating concerns that differences in information (rather than disclosure format) drive the findings. Taken together, these additional results help alleviate the concern that underlying information drives my findings rather than how it is presented.

As with any archival study, my tests represent associations for which I cannot definitively ascribe causality. To alleviate concerns about an unidentified correlated omitted variable, I perform several robustness tests. I re-estimate each model (i) using the residual value from a prediction model for the level of each presentation format type, (ii) entropy-balancing on the likelihood of using above-median slide presentation format (McMullin and Schonberger 2020), (iii) applying the control function regression method (Armstrong et al. 2022; Klein and Vella 2010), (iv) using the percentage of slides with each format, (v) applying alternative methods for dealing with outliers (i.e., winsorization and studentized residuals) (Leone et al. 2019), (vi) altering the form of industry fixed effects, (vii) excluding all fixed effects (Jennings et al. 2022), (viii) excluding all controls (Whited et al. 2022), and (ix) replacing all missing format values with zero if the firm does not have a presentation slideshow on Capital IQ's website. In each of these tests, my inferences endure. Finally, I perform the Oster (2019) delta test to assess whether my results could be attributed to correlated omitted variables and find that omitted variables are highly unlikely to explain my results.

I make three contributions to the accounting and finance literatures. First, I contribute to the growing literature examining the consequences of disclosure format. Research finds that the format of textual disclosure (i.e., headline salience, prominence, and ordering) is associated with market outcomes (Cheng et al. 2021; Chen et al. 2021; Huang et al. 2018) and that data visualizations are associated with increased information content (Christensen et al. 2022; Xu 2021; Wu 2021). My evidence suggests that more data visualizations are associated with decreased information asymmetry, increased market liquidity, and decreased investor processing costs, specifically for retail investors. Moreover, studies do not compare the relative usefulness of different types of presentation formats, focusing solely on the visual disclosures in isolation (Christensen et al. 2022) or relative to all others (Xu 2021; Wu 2021). Further, by examining outcomes related to information asymmetry and market liquidity, I respond to Blankespoor et al.'s (2020) call to "examine the effects of disclosure formatting on market outcomes other than price responsiveness, which is the focus of most existing research" (p. 25).[6]

Second, I contribute to the literature examining disclosure processing costs. Research demonstrates harms from increased textual processing costs (Chapman et al. 2019; Lawrence 2013; Lee 2012; Miller 2010). My study is among the first to empirically test the association between data visualization use and processing cost outcomes.[7] The accounting and finance setting differs from much of the literature exploring the usefulness of data visualizations because managers have substantial agency costs and prefer to highlight positive news (Baginski et al. 2018; Kothari et al. 2009; Skinner 1994). Moreover, my evidence regarding the usefulness of visual disclosures underscores the SEC's suggestions in the *Plain English Handbook* to increase disclosure effectiveness and help "level the playing field" (SEC 2022; SEC 2021).

Finally, I contribute to the earnings conference call literature by creating and making publicly available a broad dataset of slide format and content. Research on qualitative disclosures focuses on textual characteristics (Loughran and McDonald 2020; Loughran and McDonald 2016; Leuz and Wysocki 2016; Kearney and Liu 2014). My dataset can be used by researchers to explore questions relating to the dynamics between these alternative disclosure presentation choices and the information sets of managers and the effects they have on market participants' decision-making, as called for other researchers (Blankespoor et al. 2020; Loughran and McDonald 2016).

[6] Information asymmetry and market liquidity are important market factors in their own right because information asymmetry creates market frictions, which increase a firm's cost of capital.
[7] Contemporaneous data visualization studies in accounting (Christensen et al. 2022; Xu 2021; Wu 2021) examine the determinants and information content on data visualizations rather than investor processing cost outcomes. Xu (2021) examines one processing cost outcome (i.e., intraperiod price formation) and finds increased price formation as firms use more data visualizations.

2. Background and Literature

2.1 Background

An increasing number of firms disclose a slideshow to accompany the presentation segment of their earnings call (see Figure 1). The earnings call slide decks are made available at a variety of times on the day of the earnings announcement, specifically (1) simultaneously with the earnings announcement on the firms' investor relations page, (2) with the 8-K on the SECs website, or (3) during the earnings call. These slideshows contain an array of design and formatting choices. Slides can contain data visualizations, tables (e.g., balance sheet, income statement, or reconciliation tables), bullet points, product images, safe harbor disclosures, and any combination of the design choices. (See Appendix A for examples.) Moreover, managers often reference these slide decks during the earnings call to guide their discussion. One interesting note is the medium of these earnings calls vary between a video of the slides to audio-only earnings calls during which managers verbally direct investors and analysts participating on the call to look at a particular slide.

2.2 Investor Processing Costs

The earning announcement window is one of the most information rich events during a firm-quarter. Many firms supplement their announcements with an earnings call, management guidance, and 10-K/Q filing (Arif et al. 2019; Davis and Tama-Sweet 2012). Additionally, third parties will issue news articles, analyst reports, social media posts, etc., to analyze, summarize, and highlight key performance information (Coleman et al. 2022; Campbell et al. 2022; Nekrasov et al. 2022). The effect of conference calls (Frankel et al. 1999; Bushee et al. 2003; Brown et al. 2004) and the bundling of firm disclosures has been studied extensively (Frankel et al. 2022; Beaver et al. 2020; Bozanic et al. 2018; Rogers and Van Buskirk 2013). In periods of increased information processing costs, investors rationally decide how much time and effort they will exert to acquire and integrate information based on their idiosyncratic assessments of the costs and benefits (Blankespoor et al. 2020; Simon 1978). These assessments lead to incomplete stock price formation, as information is more costly to acquire and interpret (Bloomfield 2002; Grossman and Stiglitz 1980) or is presented in less salient formats (Hirshleifer and Teoh 2003). Incomplete stock price formation can lead to detrimental market effects, such as increased information asymmetry, decreased market liquidity, and increased volatility (Chapman et al. 2019; Loughran and McDonald 2014; Lawrence 2013; Miller 2010).

2.3 Disclosure Presentation Style

How information is presented affects individuals' perceptions of and response to the information (Hirshleifer et al. 2004; Taylor and Thompson 1982; Taylor and Fiske 1978). Outside of the accounting and finance literatures, the evidence is mixed regarding the efficacy of pictures and data visualizations in helping users form judgments. On the one hand, pictures or data visualizations when closely linked with text (whether spoken or written) can increase individuals' comprehension and recall by highlighting critical features (Gigerenzer et al. 2007; Houts et al. 2006; Stokes 2002; Glenberg and Langston 1992; Bower et al. 1975). Further, the effort required to acquire and integrate information into decision-making is much lower for data visualizations, relative to other disclosure formats, because people can easily recognize discontinuities and edges; variation in colors, shapes, and motion; and patterns to retrieve information without the strict adherence to rules (Kosslyn 1994). On the other hand, visualizations that inaccurately highlight the underlying data or are poorly designed can harm users' perceptions and judgments (Arunachalam et al. 2002). Further, even if managers do design accurate and clear data visualizations, users' differing levels of graph literacy (i.e., the ability to understand information presented graphically) can affect their ability to incorporate the information from the underlying data into their predictions of outcomes (Okan et al. 2016; Okan et al. 2012; Galesic and Garcia-Retamero 2011).

Research finds that investors respond differently to the information that is recognized versus disclosed not only because of the informational properties but also because of differential difficulty in processing the information. Experimental evidence finds that disclosed items require greater effort and cognitive resources to understand, relative to recognized items (Hodge et al. 2004; Dietrich et al. 2001; Hirst and Hopkins 1998). Archival evidence further finds that investors respond to recognized information even if it was previously available (Hand 1990). Much of the early research in presentation style examines how and where the textual information is disclosed. Bowen et al. (2005), Elliott (2006), and Chen et al. (2021) examine the prominence of non-GAAP earnings relative to GAAP earnings. They find evidence suggesting that investors respond more to the prominent information and that emphasized non-GAAP earnings are associated with higher quality non-GAAP earnings. Allee and DeAngelis (2015), Huang et al. (2018), and Cheng et al. (2021) examine different textual formatting choices outside the non-GAAP setting and find that investors respond more to evenly dispersed, salient, and early information.

Another formatting choice firms can make is how they visually present the information (i.e., data visualization versus tables). The SEC's Plain English Handbook suggests both visualizations and tables because they "convey information more quickly and clearly than text" (SEC 1998, 48) and "illuminate information more clearly and quickly than text" (SEC 1998, 49). Moreover, an SEC study found that investors "prefer that disclosures be written in clear, concise, understandable language using bullet points, tables, charts, and/or graphs" (SEC 2020). Loughran and McDonald (2016) suggest that managers' use of nontextual disclosures should enhance the ability of the investor to understand the information, yet very little research exists exploring alternative disclosure methods. Elliott et al. (2017) experimentally examine the use of data visualizations to convey corporate social responsibility information. They find evidence that visualizations increase investor affect, which leads to increased willingness to invest, particularly for individuals who need more help processing numbers, when the information is more "community" focused.[8] Additionally, Backof et al. (2018) find that auditors are more skeptical of aggressive management assumptions when information is presented graphically versus textually.

Other concurrent studies explore the use of data visualizations in disclosures using archival data. Christensen et al. (2022) explore firms' 10-K disclosures. They find evidence suggesting that larger firms, more volatile firms, growth firms, firms with a Big Four auditor, and firms with more peer reporting are more likely to include a data visualization, while older firms and firms with restructuring costs are less likely to do so. In addition, they find that data visualizations—specifically quantitative data visualizations—are associated with increased market volatility. Two concurrent studies explore earnings call slide presentations. Xu (2021) finds that managers include quantitative data visualizations in earnings calls when "information demand is higher, financial statement processing costs are higher, and operating performance is better." Moreover, data

[8] My tests examining retail investors' processing costs differ from those of Elliott et al. (2017) in three important ways. First, Elliott et al. examine the psychological mechanism (i.e., affect) by which graphics may influence investors' decisions, whereas I examine the generalizable associated effect size. Elliott et al. also examine investors' decisions based on corporate social responsibility reporting, whereas I examine the earnings announcement—a more salient disclosure avenue than corporate social responsibility reporting. Second, they examine investors' willingness to invest in a particular firm (i.e., buy shares); however, I examine investors' decision to trade shares (i.e., buy or sell). In an untabulated analysis, I find increases in both abnormal buying and selling by retail investors, and I find that data visualizations are associated with increased share purchases by both retail and institutional investors.

visualizations in conference call slideshows are associated with increased price reactions, trading, and price formation. Wu (2021) finds additional evidence that data visualizations increase the market reaction to earnings. However, Wu also finds that greater return reversals follow. Thus research suggests that data visualizations increase investor responses. However, the evidence is mixed about whether investors make better decisions or are simply distracted by salient information.[9]

3. Hypothesis Development

Investors rationally allocate the amount of time and effort they expend processing information after weighing the costs and benefits (Blankespoor et al. 2020; Simon 1978; Simon 1955). These trade-offs can lead to less efficient reactions by market participants, which are manifest through increases in market disagreement and processing costs. Research identifies three ways in which managers can help investors' process information. First, they can write clearer disclosures (Bonsall et al. 2017; Loughran and McDonald 2014; Li 2008); however, Bushee et al. (2018) find that complexity in writing is due to both obfuscation *and* firm complexity, and managers have limited control over the complexity of the information. The second way is to provide more time between their disclosures (Chapman et al. 2019), which is contrary to current investor information demands at the earnings announcement (Beaver et al. 2020; Rogers and Van Buskirk 2013). Additionally, Hirshleifer and Teoh (2003) find that information presented in a more

[9] Three other studies—by Brown et al. (2021), Nekrasov et al. (2021), and Ben-Rephael et al. (2021)—explore managers' use of visuals, such as stylized text or management pictures (not necessarily data visualizations). Brown et al. (2021) find that nonprofessional investors rely more on firms' non-GAAP earnings when it is disclosed in an image on Twitter, regardless of textual prominence in the actual disclosure. Nekrasov et al. (2021) also explore visuals on Twitter and find that the market reaction is positively associated with visuals in tweets. However, they find that these visuals also lead to greater return reversals. Ben-Rephael et al. (2021) examine visuals in annual reports and find an association between visual usage and lower risk, cost-of-capital, and analyst disagreement in the year following the annual report. Futher, their evidence suggests that less than one page per annual report contains a data visualization.

salient format (e.g., data visualizations) is better absorbed into market prices, and Bloomfield (2002) argues that information that is less costly to interpret is more fully reflected in prices. Overall these arguments suggest that data visualizations and tables should decrease information asymmetry and thus increase market liquidity, because investors can extract the pertinent information better through the use of data visualizations.

However, the increased salience to more favorable information does not equate to increased value relevance. Managers have incentives to highlight favorable information, especially at the earnings announcement (Davis and Tama-Sweet 2012; Skinner 1994), and control the narrative about the firm (Lee 2016; Hollander et al. 2010; Mayew 2008). They also have significant discretion over how they present the data (e.g., axis, colors, size, etc.), what information they present (e.g., ROA, sales, etc.), and where in the slideshow to display it (e.g., multiple visuals on a single slide, early, etc.). Each of these choices may help or inhibit investors' ability to incorporate the information into their decision-making, focusing them on information more favorable to managers. Thus the information managers present may not help investors process firm information and can lead to more disagreement. In fact, Nekrasov et al. (2022) and Christensen et al. (2022) find that visuals are associated with greater reversals and post-filing return volatility, respectively.[10] Finally, many practitioners have expressed concern that slideshows inhibit presenters' ability to adequately express the relevant information because they prioritize the format of the presentation over the content (Cooper 2009; Tufte 2001). These arguments suggest an unclear relation between information asymmetry or market liquidity and alternative disclosure

[10] Christensen et al. (2022) argue that the increase in post-filing return volatility "suggest(s) that firms include infographics when they expect investors to have greater difficulty interpreting the information in the 10-K, but that those infographics have the unintended consequence of increasing information uncertainty" (page 28).

formats. However, these disclosures occur repeatedly as managers continue to disclose in future quarters.[11] I expect alternative disclosure formats to be negatively (positively) associated with information asymmetry (market liquidity), stated formally:

H1a: Disclosure format is negatively associated with information asymmetry.

H1b: Disclosure format is positively associated with market liquidity.

The prior hypotheses examine broad market implications, similar to prior research. One important aspect of alternative disclosure formats is the influence they can have on investors with differing levels of sophistication. Xu (2021) and Christensen et al. (2022) suggest that data visualizations are associated with aggregate market outcomes, specifically, increased market reactions and post-filing return volatility, respectively. However, they find little evidence suggesting that differing investor types benefit from data visualizations. Studies show that retail investors decisions are affected by how information is presented in the financial statements and in writing (Miller 2010; Maines and McDaniel 2000). Retail investors may benefit from alternative formatting choices, particularly from more salient forms, because visuals can highlight the most relevant information and reduce their processing costs (Cardinaels 2008). Moreover, data visualizations and tables require less effort and expertise to interpret because individuals can easily extract the relevant information from variations in color, size, and discontinuities (Kosslyn 1994). Thus I expect data visualizations to benefit both retail and institutional investors, stated formally:

H2a: Disclosure format is negatively associated with investor processing costs.

H2b: Disclosure format is negatively associated with retail investor processing costs.

---

[11] For example, the correlation between current number of slides with a data visualization (table) [text] and the previous earnings announcement is 0.806 (0.767) [0.778] (untabulated).

Each of these hypotheses examines whether these alternative disclosure formatting choices influence investors' decisions to make trades. They do not address whether the disclosure format affects the quality of the trades, particularly for retail traders. Less-knowledgeable individuals perform better on tasks when they can use data visualizations, relative to alternative disclosure formats (Cardinaels 2008). However, investors with differing levels of "graph literacy" or varying levels of ability to understand numerical information may not accurately incorporate the information into their decisions (Elliott et al. 2017; Galesic and Garcia-Retamero 2011). If disclosure format helps retail investors make better decisions, then I expect increased trading consistency among retail investors. However, if data visualizations distract retail investors from more value relevant information, then I expect decreased trading consistency. Thus I make no directional prediction regarding the trading consistency by retail investors, stated formally, in null form:

H3: Disclosure format is not associated with retail investor trading consistency.

4. Sample Selection and Research Design

4.1 Sample Selection

I begin with the universe of Capital IQ earnings call transcripts (231,232 observations) from 2008 through 2020. I match firm-quarter identifiers from Capital IQ to Compustat, CRSP, I/B/E/S, and TAQ, which results in 140,021 observations remaining. I then use Python to search for and download (if available) every earnings call slideshow presentation from Capital IQ. I download 54,645 slideshows from the available 140,021 earnings calls (39.03%).[12] I eliminate

[12] This percentage is subject to measurement error. I collect my sample from CapitalIQ, which downloads and saves them for their subscribers to use later. I cannot use an alternative source because firms' do not retain their slideshows on their websites for a consistent period, nor do most firms file their slideshows with the SEC.

2008 and 2009 from my main analysis because of sample selection bias concerns (less 893 observations).[13] I then drop observations with a market price less than \$1 and total assets less than \$5 million (less 396 observations) and financial firms (less 10,264 observations).[14] I also drop observations where the extracted slideshow PDF does not have extractable text, the "slideshow" is not an actual slideshow (these were typically annual/quarterly reports, results summary "handout" or other random firm disclosures), or I could not download the slideshow from Capital IQ's website (less 4,516 observations). Finally, I dropped observations missing adequate data from WRDS Capital IQ Transcript (less 899 observations), Compustat, CRSP, IBES, intraday TAQ, or RavenPack (less 6,943 observations). These sample selection criterions yield a sample of 30,734 observations.

4.2 Research Design

To answer each of my research questions, I test the following model:

$$\begin{aligned} DepVar = {} & \beta_0 + \beta_1 \ln DataVis + \beta_2 \ln Tables + \beta_3 \ln Text + \beta_{4-5} EarnNews + \\ & \beta_{6-11} CallChars + \beta_{12-16} \text{Concur} + \beta_{17-18} Timing + \beta_{19-25} \text{FirmChars} + FE. \end{aligned} \quad (1)$$

DepVar is either abnormal bid-ask spread, abnormal price impact, abnormal trading volume, abnormal retail trading volume, or retail trading consensus. Each of these measures is calculated following Blankespoor et al. (2020) or Boehmer et al. (2021). *lnDataVis* (*lnTables*) [*lnText*] is the logged number of slides with at least one data visualization (table) [text]. I calculate

[13] These concerns arise because the percentage of firms that use slides, according to CapitalIQ, in 2008 and 2009 is 0.33% and 8.90%, respectively. However, the percentage of calls that mention "slide(s)," "slideshow," and "powerpoint" in the presentation is 19.67% and 22.29%, respectively. In 2010, I find that 22.36% have slides on CapitalIQ, and 22.89% mention slides (see Figure 1). In an untabulated analysis, I find similar inferences when I include 2008 and 2009 in my main analysis.

[14] In an untabulated analysis, I find similar inferences when I include financial firms in my main analysis.

DataVis using an object detection machine learning model, which is outlined in Appendix C. I estimate the model separately for each of the three aggregate disclosure format variables and then together. Hypothesis 1a (1b) states that each disclosure formatting type is associated with a decrease (increase) in information asymmetry (market liquidity), and thus I expect *lnDataVis*, *lnTables*, and *lnText* to be negatively associated with both abnormal bid-ask spread and abnormal price impact [i.e., $\beta_1, \beta_2, \beta_3 < 0$].[15] I also perform Wald tests between β_1, β_2, and β_3 to test the equality across the three coefficients. I expect a lower level of both abnormal bid-ask spread and abnormal price impact for data visualizations, relative to tables, which is lower relative to text slides. Overall I expect $\beta_1 < \beta_3$, $\beta_2 < \beta_3$, and $\beta_1 < \beta_2$, which would suggest that data visualizations are associated with less information asymmetry and more market liquidity, relative to tables, which are associated with less information asymmetry and more market liquidity, relative to text.

I control for an array of observable characteristics because the underlying decision to choose a particular format may be associated with observable characteristics that are also associated with market outcomes, namely *EarnNews*, *CallChars*, *Concurrent*, *Timing*, and *FirmChars.* First, I include the *EarnNews* variables to capture the earnings announcement news. I include the absolute value of unexpected earnings ($|UnEarn|$) and an indicator for whether the firm reported a loss (*Loss*). Second, the *CallChars* variables measure different characteristics of the earnings call to capture the information conveyed by managers and analysts. Specifically, I include the absolute value of managers' and analysts' tone ($|MgrTone|$ and $|AnalystTone|$), the informative

[15] A decrease in abnormal price impact is an increase to market liquidity (Blankespoor et al. 2020).

component of complex information on the call as well as the obfuscation component, following Bushee et al. (2018) (*InfoMgr*, *ObfuPres*, and *ObfuQA*), and the numerical information by managers on the call (*PctNum*). Third, I control for the *Concurrent* news produced by the firm and third parties. These variables include indicators for whether the firm simultaneously disclosed an earnings forecast (*Guid*) or its 10-K/Q (*Bundled*) and the number of concurrent earnings announcements (*lnConcurEA*), analyst reports (*lnAnalystFore*), and media articles (*lnArticles*). Fourth, I control for the *Timing* of the earnings announcement, namely whether the call was after hours (*AftHrs*), and the number of days since the end of the fiscal quarter (*lnLag*). Finally, I include *FirmChars* to control for firm characteristics. These variables are market size *(lnMVE*), operating complexity (*lnSegs*), growth potential *(MTB*), stock price momentum *(MOM*), institutional ownership (*InstOwn*), analyst following (*lnAnalysts*), firm age (*lnAge*), and earnings volatility *(EarnVol*). I define all variables in Appendix B. I also control for the variation associated with the day of week, industry, and calendar-year-quarter by including fixed effects for each of those variables. I cluster standard errors by firm and calendar-year-quarter. I also estimate all of my results after excluding extreme observations using Cook's Distance values greater than 4/N.[16]

Hypotheses 2a and 2b address whether investor processing costs, particularly those of retail investors, are associated with alternative disclosure formats. I re-estimate equation 1 using abnormal trading volume and abnormal retail trading volume. Hypothesis 2a (2b) states that each disclosure format is associated with decreased investor (retail investor) processing costs, and thus

[16] Leone et al. (2019) find that winsorization does not effectively reduce the effect of significant outliers. Thus I tabulate all results after applying Cook's Distance. In a robustness analysis, I exclude extreme observations using studentized residuals and by only winsorizing and find consistent inferences.

I expect *lnDataVis*, *lnTables*, and *lnText* to be positively associated with each type of abnormal trading volume [i.e., $\beta_1, \beta_2, \beta_3 > 0$]. Similar to hypothesis 1, I test the equality among the coefficients and expect $\beta_1 > \beta_3$, $\beta_2 > \beta_3$, and $\beta_1 > \beta_2$, suggesting data visualizations are associated with lower investor processing costs, relative to tables, which are more, relative to text. Hypothesis 3 examines the consistency of retail investor trades. I re-estimate equation 1 using retail trading consensus. I make no prediction regarding the association between the disclosure format types and retail trading consensus because alternative formats could distract retail investors from value relevant information or highlight important information.

5. Results

5.1 Descriptive Statistics

Panel A of Table 2 presents descriptive statistics for the 30,734 firm-year-quarter observations in my sample. I find that the average number of slides, specifically the number of slides with a data visualization (table) [text], is 5.97 (2.81) [11.25] after dropping title, safe-harbor disclosure, agenda, and question-and-answer slides. The average (median) of abnormal bid-ask spread is 0.194 (0.165), and the abnormal price effect is 1.580 (1.434). In addition, I find that the average (median) abnormal trading volume is 0.673 (0.646), the abnormal retail trading is 0.119, (0.034) and the retail trading consensus is 0.125 (0.091). Each of these variables is slightly higher than evidence from prior research (Blankespoor et al. 2020). However, I examine a specific subset of observations during a later period. The rest of my control variables are generally consistent with prior research.

Panel B of Table 2 reports the correlation matrix. These correlations provide preliminary evidence of hypothesis 1. The correlation for data visualization slides and abnormal bid-ask spread is significantly negative (ρ = -0.048, *p*-value < 0.001), but there is no association with either tables or text. These results are partially consistent with H1a, which states that the different slide formats are negatively associated with information asymmetry. I also find that the correlation between data visualization [text] slides and abnormal price effect is significantly negative (ρ = -0.038, *p*-value < 0.001) [ρ = -0.013, *p*-value = 0.023]. However, I find a marginally positive significant correlation between table slides and abnormal price impact (ρ = 0.010, *p*-value = 0.072). These results are mostly consistent with H1b, which states that the different slide formats are associated with higher market liquidity. With regard to my second hypotheses, I find that the correlation between table slides and abnormal trading volume is significantly positive (ρ = 0.018, *p*-value < 0.001). However, the correlations between data visualization [text] slides and abnormal trading volume is significantly negative (ρ = -0.052, *p*-value < 0.001) [ρ = -0.011, *p*-value = 0.044]. These results are not consistent with H2a, which states that the different slide formats are positively associated with overall investor attention. Moreover, I find mixed preliminary evidence that the differing slide formats are negatively associated with retail investor trading as the correlations are significantly negative in the Spearman correlations, contrary to H2b. Finally, I find some preliminary evidence that retail trading consensus is negatively associated with all three disclosure format types. Interestingly, I find that all three disclosure formatting types are positively associated with each other.

5.2 Main Results

5.2.1 Information Asymmetry and Market Liquidity

Table 3 presents multivariate evidence regarding H1a, which tests the relation between each format choice and information asymmetry. To test H1a, I estimate the model separately for each of the three disclosure format options: (1) data visualization (column 1), (2) tables (column 2), and text (column 3). I also test a model with all format options combined (column 4) and then test the difference between coefficients. In column 1, I find that *lnDataVis* is negatively associated with *Abn BidAsk Spread*, with a coefficient of -0.0099 and a p-value less than 0.01. Also, in column 2 [3], I find a negative [positive] but insignificant association between *lnTables* [*lnText]* and *Abn BidAsk Spread*. These results suggest that only data visualizations are associated with a lower information asymmetry during the earnings announcement. Moreover, these effects remain when I estimate the model including all three types of format choices (column 4). Overall I find evidence consistent with H1a, which predicts that data visualization and table slides will be negatively associated with information asymmetry. However, I do not find similar evidence for text slides. In terms of economic significance, my results suggest that a 1 % increase in slides with a data visualization is associated with a 1.11% decrease in the *Abn BidAsk Spread*.

I also perform Wald tests to test the difference between the differing types of formatting choices. I find that the difference between data visualization and tabular [textual] slides is -0.0085 (p-value < 0.01) [-0.0173 (p-value < 0.01)] with an F-stat of 6.22 [12.42]. Moreover, I find that the difference tabular and textual slides is -0.0088 (p-value < 0.05) with an F-stat of 3.671. However, neither sort of is significantly associated with *Abn BidAsk Spread*. These results suggest that disclosure formatting types have differential effects on abnormal bid-ask spread, with data visualization slides associated with a lower abnormal information asymmetry, relative to both tabular and textual slides.

Table 4 presents multivariate evidence regarding H1b, which tests the effect each format type has on market liquidity and whether some formats increase or reduce liquidity more, relative to the others. Similar to H1a, for H1b I estimate the model separately for each of the three disclosure types and then fully combined for a test of the difference between coefficients. In column 1, I find that *lnDataVis* is negatively associated with *Abn Price Effect*, with a coefficient of -0.0127 (p-value < 0.01). In addition, in column 2 [3], I find a positive but insignificant association between *lnTables* [*lnText*] and *Abn Price Effect*. This result, along with the results in table 3, suggests that data visualization slides are associated with an increase in market liquidity during the earnings announcement.[17] Moreover, the effects in columns 1–3 remain when I estimate the model including all three types of format choices (column 4). In terms of economic significance, my results suggest a 1% increase in slides with a data visualization is associated with a 1.56% decrease in *Abn Price Effect*.

I also perform Wald tests to test the difference between the different formats. I find that the difference between data visualization and tabular [textual] slides is -0.0194 (p-value < 0.05) [-0.0308 (p-value < 0.01)] with an F-stat of 4.63 [6.49]. In addition, I find that the difference tabular and textual slides is insignificant. These results show that disclosure different formats have different abnormal price effects. Overall the results in Tables 3 and 4 are consistent with H1b, which predicts an increase in market liquidity with respect to data visualization use. However, I

[17] Blankespoor et al. (2020) argue that these two measures along with abnormal depth provide evidence of market liquidity. In an untabulated analysis, I find a directionally consistent but statistically insignificant association between abnormal depth and data visualizations.

do not find evidence of an increase in market liquidity for firms that use more tabular or textual slides.

5.2.2 Investor Information Processing

Table 5 presents multivariate evidence regarding H2a, which tests the association between each format and investor information processing. In column 1, I find that *lnDataVis* is positively associated with *Abn Trading Volume*, with a coefficient of 0.0101 and a p-value less than 0.05, consistent with the findings of Xu (2021). Additionally in column 3, I find *lnText* is positively associated with *Abn Trading Volume*, with a coefficient of 0.0109 and a p-value less than 0.10. However, I do not find a statistically significant association between *lnTables* and *Abn Trading Volume*. Further, in column 4, I find only *lnDataVis* is associated with *Abn Trading Volume*. In terms of economic significance, these results suggest a 1% increase in slides with a data visualizations is associated with a 0.94% increase in *Abn Trading Volume*. I also test the difference across coefficients and find no statistical difference between any of the disclosure formats.

I then test the association between disclosure formats and retail investors' processing costs. Table 6 presents the multivariate evidence regarding H2b, which tests whether retail investor processing costs are negatively associated with disclosure formatting choices. In column 1 [2], I find that *lnDataVis [lnTables]* is positively associated with *Abn Retail Trading Volume*, with a coefficient of 0.0051 [0.0033] and a p-value less than 0.05 [0.10]. In addition, in column 3, I find a negative but insignificant association between *lnText* and *Abn Retail Trading Volume*. These results remain when I include all three format types in the same analysis (column 4). In terms of economic significance, I find a 1% increase in slides with a data visualization [tables] is associated with a 0.59% [0.38%] increase in *Abn Retail Trading Volume*. In addition, I also perform Wald tests to test the difference between the differing types of formats. I find that the difference between data visualization [tabular] and textual slides is 0.0109 (p-value < 0.05) [0.0087 (p-value < 0.05)] with an F-stat of 5.639 [3.765]. I also find no difference between data visualization and tabular slides. These results suggests that these two types of slides are associated with lower retail investor processing costs.

5.2.3 Retail Investor Trading Consistency

Table 7 presents multivariate evidence for H3, which estimates the effect of each slide format on retail investor trading consistency. Overall I find no significant association between any of the disclosure presentation formats and retail investor trading consistency. These results, along with those presented in Table 6, suggest that retail investors trade more as firms include more data visualizations and tables in their earnings conference call slideshows, but the increased usage does not increase their trading consistency. Thus the use of data visualizations and tables may only attract retail investors rather than increase the quality of their decisions.

5.3 Additional Analyses

I perform three additional analyses to address whether the results are driven by the underlying information or the presentation of the information. First, I separate the data visualization variable into visualization type (i.e., bar charts, line graphs, pie graphs, and maps). In an untabulated analysis, I find that the average usage of bar charts is 4.07, pie graphs is 0.98, line graphs is 0.81, and maps is 0.74 slides per slideshow, which suggests that bar charts are the graph most commonly used by managers. I then test whether the alternative types of data visualizations have differential effects on my dependent variables of interest. Table 8 Panel A presents the multivariate analyses. Most of the aforementioned results are primarily attributable to bar charts and line graphs, which most often include financial statement information.

Second, I re-examine my main analysis using the number of visualizations per slide. In an untabulated analysis, I find the average number of slides with one data visualization is 3.62, two is 1.41, three is 0.58, and four or more is 0.43. I present the multivariate analysis of slides with multiple data visualizations in Table 8 Panel B. I find that information processing benefits are concentrated among slides that use either one or two data visualizations, which suggests that the associated reduction in investor processing costs from data visualization usage is concentrated in slides with fewer pieces of visual information to process.

Finally, my results have ignored the type of information that is conveyed in the slides. Figure 2 presents word clouds of the information presented in slides of various visual formats. I find that the majority of information is financially related in all types of slides and fairly consistent across slide types. To further test this stipulation, I perform two untabulated regression analyses. First, I search for financial words on each slide, using the dictionaries from Brochet et al. (2018). Second, I perform a topic analysis of the presentation segment of the earnings call using Latent Dirichlet allocation and include the topic discussion values as controls in my main model. In each of these analyses I find consistent results as those in my primary analysis. However, the effects are much weaker for macroeconomic-related slides. Overall, these three additional analyses suggest that my results are not driven by the underlying information but rather the visual disclosure format.

5.4 Sensitivity Analyses

I estimate several sensitivity analyses to additionally validate the main results. One concern is that the choice by firms to issue a certain level of visualization types influences the relation that I find. To alleviate this concern, I re-estimate my model (i) using the residual from a predicted model, (ii) entropy balancing on the likelihood of having more data visualization slides than the median (McMullin and Schonberger 2020), (iii) applying the Klein and Villa (2010) control function regression method (Armstrong et al. 2022), and (iv) replacing industry fixed effects with firm fixed effects. My inferences are unchanged with these alternative specifications, with the exception of (1) market liquidity, when I estimate the model with firm fixed effects and (2) abnormal trading volume when entropy balancing.

Second, in each of my models, I estimate the different types of disclosure formatting as the logged number of slides that meet each criterion. One concern is that firms with many slides in general might influence my results. I do not include these estimations as my main analysis because (1) doing so would force a trade-off between different data visualizations and all other formatting choices and (2) a null result can be caused by either the numerator (i.e., disclosure format) or the denominator (i.e., total number of slides). The percentage of slides with data visualizations tests whether data visualizations relate differentially to the dependent variable of interest, relative to tables and text, not whether they are inherently associated with the outcome. However, to alleviate concerns that my results are attributable to excessive slide usage, I re-estimate my models using the percentage of each type, relative to the total number of slides. I find consistent results for all specifications with the exception of abnormal trading volume.

Third, my main analyses include year-quarter, industry, and day-of-week fixed effects to control for the time-invariant, industry-invariant, and weekday-invariant effects. Jennings, Kim, Lee, and Taylor (2022) suggest that, when the independent variable of interest is measured with measurement error in a model with high-dimensional fixed effects, the fixed effects can bias in favor of falsely rejecting a true null hypothesis. Consistent with prior evidence, my machine learning algorithm is only 86% accurate in identifying data visualizations, leading to small, random measurement error in each of my independent variables of interest. In an untabulated analysis, I estimate the results without any fixed effects and alternative industry fixed effects classifications (Fama-French 12 and two-digit SIC codes). With the exception of the abnormal trading volume results, my inferences are unchanged, which suggests that my fixed effects structure does not induce false positives with respect to abnormal bid-ask spread, price impact, or retail trading volume.

Fourth, in each of my models, I include an array of variables to control for various other factors that may also be associated with both the decision to include data visualizations and outcomes. Whited et al. (2022) argue that too many control variables can introduce bias into well-specified models. To alleviate concerns that measurement error in both my independent variable of interest and my control variables affect my inferences, in an untabulated analysis, I re-estimate each model without including any control variables both with and without fixed effects and find consistent results as presented previously. The exception is abnormal trading volume, which suggests that the inferences are not attributable to “included variable bias.”

Finally, as in all archival studies, a correlated omitted variable may exist that I cannot identify and that may influence the results in a way predicted by my hypotheses (in particular, differentially for retail investors, relative to institutional investors). To address this possibility, I estimate the Oster (2019) delta test and find values greater than the absolute value of 1.9 for all my main results, which suggests that omitted variables would have to be 1.9 times stronger than the variables I have already included to harm the results, again with the exception of abnormal trading volume. Oster (2019) argues that values greater than one provide sufficient evidence against correlated omitted variables significantly altering the results.

6. Conclusion

I explore the association between alternative disclosure format types in earnings conference call slides and information asymmetry, market liquidity, trading volume, and retail investor trade consistency. First, I show that slides with a data visualization are associated with lower abnormal bid-ask spread, which suggests that these slides are associated with less investor disagreement. I also find that the negative association between data visualizations and abnormal bid-ask spread is lower, relative to either tabular or textual slides. Second, I find that data visualization slides are associated lower abnormal price impact, which suggests that they are associated with higher market liquidity. Moreover, I find the negative association between data visualizations and abnormal price impact differs from that for both tabular and textual slides. I also find evidence suggesting that data visualizations are positively associated with abnormal trading volume. However, I find that this relation is not robust. Additionally, I find that abnormal retail investor trading volume is higher for firms that use more data visualization or tabular slides and that these effects are statistically different from textual slides. Finally, I do not find any association with retail investor trading consistency as firms use more data visualizations. Thus retail investors increase their trading for firms that use more data visualizations and tables as part of a larger disclosure strategy, but they do not appear to differentially change their trading for better or worse.

Like all studies that rely on archival data, mine has limitations. One is my inability to provide causal evidence. While I attempt to provide corroborating evidence by examining the predicted level of disclosure format, entropy balancing, and applying the Klein and Villa (2010) control function method, I cannot provide definitive causal evidence that these disclosure formats influence my findings. Another limitation is I cannot separate distorted or misleading data visualizations from those of higher quality because my machine learning process cannot determine the quality or usefulness of the visualizations. Thus I cannot disentangle the quality of the data visualization from its quantity. However, I expect lower quality data visualizations to bias against finding these results, due to the increased difficulty in acquiring and incorporating the information into investor decision-making. Finally, I focus exclusively on the disclosure format of slides disclosed at the earnings announcement to use during the earnings call. I cannot rule out the possibility that firms also include data visualizations in other disclosures, such as social media, or whether these visuals are re-used from other disclosures. While I do not expect data visualization repetition to bias toward rejecting the null, I cannot rule this out.

Overall my evidence is an important step in helping academic researchers, managers, and regulators understand the influence of disclosure format on market participants, thus extending the work of Christensen et al. (2022), Xu (2021), and Wu (2021). Future research could create more sophisticated machine learning techniques to help determine whether these visuals are of higher or lower quality to be able better understand managers' incentives surrounding disclosure format. In addition, future research could examine slide disclosures in conjunction with other disclosure avenues, namely the 10-K/Q, 8-K, or social media to understand how the firms' overall use of graphical disclosures impacts investor decision-making and the informativeness of disclosures in general.

References

Allee, K. D., and M. D. DeAngelis. 2015. The Structure of Voluntary Disclosure Narratives: Evidence from Tone Dispersion: the structure of voluntary disclosure narratives. *Journal of Accounting Research* 53 (2): 241–274.

Arif, S., N. T. Marshall, J. H. Schroeder, and T. L. Yohn. 2019. A growing disparity in earnings disclosure mechanisms: The rise of concurrently released earnings announcements and 10-Ks. *Journal of Accounting and Economics* 68 (1): 101221.

Armstrong, C., A. Nicoletti, and F. S. Zhou. 2022. Executive stock options and systemic risk. *Journal of Financial Economics* 146 (1): 256–276.

Arunachalam, V., B. K. W. Pei, and P. J. Steinbart. 2002. Impression Management with Graphs: Effects on Choices. *Journal of Information Systems* 16 (2): 183–202.

Backof, A. G., T. D. Carpenter, and J. Thayer. 2018. Auditing Complex Estimates: How Do Construal Level and Evidence Formatting Impact Auditors' Consideration of Inconsistent Evidence? 35 (4): 1798–1815.

Baginski, S. P., J. L. Campbell, L. A. Hinson, and D. S. Koo. 2018. Do Career Concerns Affect the Delay of Bad News Disclosure? *The Accounting Review* 93 (2): 61–95.

Bamber, L. S. 1987. Unexpected Earnings, Firm Size, and Trading Volume around Quarterly Earnings Announcements. *The Accounting Review* 62 (3): 510–532.

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): 1–21.

Ben-Rephael, A., J. Ronen, T. Ronen, and M. Zhou. 2021. Do Images Provide Relevant Information to Investors? An Exploratory Study. *SSRN Electronic Journal*.

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): 1–46.

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

Boehmer, E., C. M. Jones, Xiaoyan Zhang, and Xinran Zhang. 2021. Tracking Retail Investor Activity. *The Journal of Finance* 76 (5): 2249–2305.

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.

Bowen, R. M., A. K. Davis, and D. A. Matsumoto. 2005. Emphasis on Pro Forma versus GAAP Earnings in Quarterly Press Releases: Determinants, SEC Intervention, and Market Reactions. *The Accounting Review* 80 (4): 1011–1038.

Bower, G. H., M. B. Karlin, and A. Dueck. 1975. Comprehension and memory for pictures. *Memory & Cognition* 3 (2): 216–220.

Bozanic, Z., D. T. Roulstone, and A. Van Buskirk. 2018. Management earnings forecasts and other forward-looking statements. *Journal of Accounting and Economics* 65 (1): 1–20.

Brochet, F., K. Kolev, and A. Lerman. 2018. Information transfer and conference calls. *Review of Accounting Studies* 23 (3): 907–957.

Bushee, B. J., I. D. Gow, and D. J. Taylor. 2018. Linguistic Complexity in Firm Disclosures: Obfuscation or Information? *Journal of Accounting Research* 56 (1): 85–121.

Campbell, B., M. S. Drake, J. R. Thornock, and B. J. Twedt. 2022. Earnings Virality. *Journal of Accounting and Economics* Forthcoming: 1–30.

Campbell, J. Y., T. Ramadorai, and A. Schwartz. 2009. Caught on tape: Institutional trading, stock returns, and earnings announcements. *Journal of Financial Economics* 92: 26.

Cardinaels, E. 2008. The interplay between cost accounting knowledge and presentation formats in cost-based decision-making. *Accounting, Organizations and Society* 33 (6): 582–602.

Chapman, K. L., N. Reiter, H. D. White, and C. D. Williams. 2019. Information overload and disclosure smoothing. *Review of Accounting Studies* 24 (4): 1486–1522.

Chen, J. V., K. H. Gee, and J. J. Neilson. 2021. Disclosure Prominence and the Quality of Non□GAAP Earnings. *Journal of Accounting Research* 59 (1): 163–213.

Cheng, L., D. T. Roulstone, and A. Van Buskirk. 2021. Are Investors Influenced by the Order of Information in Earnings Press Releases? *The Accounting Review* 96 (2): 413–433.

Christensen, T. E., K. E. Fronk, J. A. Lee, and K. K. Nelson. 2022. Data visualization in 10-K filings. *Working Paper*.

Coleman, B., K. J. Merkley, and J. Pacelli. 2022. Human Versus Machine: A Comparison of Robo-Analyst and Traditional Research Analyst Investment Recommendations. *The Accounting Review*.

Cooper, E. 2009. Overloading on Slides: Cognitive Load Theory and Microsoft's Slide Program PowerPoint. *Association for the Advancement of Computing in Education Review (formerly AACE Journal)* 17 (2): 9.

Davis, A. K., and I. Tama-Sweet. 2012. Managers' Use of Language Across Alternative Disclosure Outlets: Earnings Press Releases versus MD&A*. *Contemporary Accounting Research* 29 (3): 804–837.

Diaz, J. 2021. Data Visualization Leads to Better Decision Making - KPMG Philippines. *KPMG*. https://home.kpmg/ph/en/home/insights/2021/07/data-visualization-leads-to-better-decision-making.html.

Dietrich, J. R., S. J. Kachelmeier, D. N. Kleinmuntz, and T. J. Linsmeier. 2001. Market Efficiency, Bounded Rationality, and Supplemental Business Reporting Disclosures. *Journal of Accounting Research* 39 (2): 243–268.

Elliott, W. B. 2006. Are Investors Influenced by Pro Forma Emphasis and Reconciliations in Earnings Announcements? *The Accounting Review* 81 (1): 113–133.

Elliott, W. B., S. M. Grant, and K. M. Rennekamp. 2017. How Disclosure Features of Corporate Social Responsibility Reports Interact with Investor Numeracy to Influence Investor Judgments. *Contemporary Accounting Research* 34 (3): 1596–1621.

EY. 2014. Disclosure effectiveness in action: what strides are you making? https://assets.ey.com/content/dam/ey-sites/ey-com/en_us/topics/assurance/ey-disclosure-effectiveness-in-action.pdf.

———. 2017. Disclosure effectiveness in action: companies make great strides. https://www.financialexecutives.org/Research/FERF-Files/2016/ey-ferf-disclosure-effectiveness-in-action-compani.aspx.

Frankel, R., J. Jennings, and J. Lee. 2022. Previously-Priced Discretionary Disclosure. *Working Paper*.

Galesic, M., and R. Garcia-Retamero. 2011. Graph Literacy: A Cross-Cultural Comparison. *Medical Decision Making* 31 (3): 444–457.

Gigerenzer, G., W. Gaissmaier, E. Kurz-Milcke, L. M. Schwartz, and S. Woloshin. 2007. Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest* 8 (2): 44.

Glenberg, A. M., and W. E. Langston. 1992. Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language* 31 (2): 129–151.

Grossman, S. J., and J. E. Stiglitz. 1980. On the Impossibility of Informationally Efficient Markets. *The American Economic Review*: 393–408.

Hand, J. R. M. 1990. A Test of the Extended Functional Fixation Hypothesis. *The Accounting Review* 65 (4): 740–762.

Hirshleifer, D. A., S. S. Lim, and S. H. Teoh. 2004. *Disclosure to an Audience with Limited Attention*. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network.

Hirshleifer, D., and S. H. Teoh. 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* 36 (1): 337–386.

Hirst, D. E., and P. E. Hopkins. 1998. Comprehensive Income Reporting and Analysts' Valuation Judgments. *Journal of Accounting Research* 36 (Supplmental): 47–75.

Hodge, F. D., J. J. Kennedy, and L. A. Maines. 2004. Does Search-Facilitating Technology Improve the Transparency of Financial Reporting? *The Accounting Review* 79 (3): 687–703.

Hollander, S., M. Pronk, and E. Roelofsen. 2010. Does Silence Speak? An Empirical Analysis of Disclosure Choices During Conference Calls. *Journal of Accounting Research* 48 (3): 531–563.

Houts, P. S., C. C. Doak, L. G. Doak, and M. J. Loscalzo. 2006. The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling* 61 (2): 173–190.

Huang, X., A. Nekrasov, and S. H. Teoh. 2018. Headline Salience, Managerial Opportunism, and Over- and Underreactions to Earnings. *The Accounting Review* 93 (6): 231–255.

Jennings, J. N., J. M. Kim, J. A. Lee, and D. J. Taylor. 2022. *Measurement Error, Fixed Effects, and False Positives in Accounting Research*. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network.

Kearney, C., and S. Liu. 2014. Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis* 33: 171–185.

Klein, R., and F. Vella. 2010. Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics* 154 (2): 154–164.

Kosslyn, S. 1994. On cognitive neuroscience. *Journal of Cognitive Neuroscience* 6 (3): 297–303.

Kothari, S. P., S. Shu, and P. D. Wysocki. 2009. Do Managers Withhold Bad News? *Journal of Accounting Research* 47 (1): 241–276.

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

Lee, C. M. C. 1992. Earnings news and small traders. *Journal of Accounting and Economics* 15: 265–302.

Lee, J. 2016. Can Investors Detect Managers' Lack of Spontaneity? Adherence to Predetermined Scripts during Earnings Conference Calls. *The Accounting Review* 91 (1): 229–250.

Lee, Y.-J. 2012. The Effect of Quarterly Report Readability on Information Efficiency of Stock Prices: Quarterly Report Readability and Information Efficiency. *Contemporary Accounting Research* 29 (4): 1137–1170.

Leone, A. J., M. Minutti-Meza, and C. E. Wasley. 2019. Influential Observations and Inference in Accounting Research. *The Accounting Review* 94 (6): 337–364.

Leuz, C., and P. D. Wysocki. 2016. The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research. *Journal of Accounting Research* 54 (2): 525–622.

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

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

———. 2016. Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research* 54 (4): 1187–1230.

———. 2020. Textual Analysis in Finance. *Annual Review of Financial Economics* 12 (1): 357–375.

Maines, L. A., and L. S. McDaniel. 2000. Effects of Comprehensive-Income Characteristics on Nonprofessional Investors' Judgments: The Role of Financial-Statement Presentation Format. *The Accounting Review* 75 (2): 179–207.

Mayew, W. J. 2008. Evidence of Management Discrimination Among Analysts during Earnings Conference Calls. *Journal of Accounting Research* 46 (3): 627–659.

McMullin, J. L., and B. Schonberger. 2020. Entropy-balanced accruals. *Review of Accounting Studies* 25 (1): 84–119.

Miller, B. P. 2010. The Effects of Reporting Complexity on Small and Large Investor Trading. *The Accounting Review* 85 (6): 2107–2143.

Nekrasov, A., S. H. Teoh, and S. Wu. 2022. Visuals and attention to earnings news on twitter. *Review of Accounting Studies* 27: 1233–1275.

Okan, Y., M. Galesic, and R. Garcia-Retamero. 2016. How People with Low and High Graph Literacy Process Health Graphs: Evidence from Eye-tracking: Graph Literacy and Health Graph Processing. *Journal of Behavioral Decision Making* 29 (2–3): 271–294.

Okan, Y., R. Garcia-Retamero, E. T. Cokely, and A. Maldonado. 2012. Individual Differences in Graph Literacy: Overcoming Denominator Neglect in Risk Comprehension: Individual Differences in Graph Literacy. *Journal of Behavioral Decision Making* 25 (4): 390–401.

Puckett, A., and X. Yan. 2011. The Interim Trading Skills of Institutional Investors. *The Journal of Finance* 66 (2): 33.

Rogers, J. L., and A. Van Buskirk. 2013. Bundled forecasts in empirical accounting research. *Journal of Accounting and Economics* 55 (1): 43–65.

SEC. 1998. A Plain English Handbook. *Securities and Exchange Commission*. https://www.sec.gov/reportspubs/investor-publications/newsextrahandbookhtm.html.

———. 2020. Recommendation on Disclosure Effectiveness. *Securities and Exchange Commission*. https://www.sec.gov/spotlight/investor-advisory-committee-2012/disclosure-effectiveness.pdf.

———. 2021. What We Do. *Securities and Exchange Commission*. https://www.sec.gov/about/what-we-do.

———. 2022. Our Goals. *Securities and Exchange Commission*. https://www.sec.gov/our-goals.

Simon, H. A. 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics* 69 (1): 99–118.

———. 1978. Rationality as Process and as Product of Thought. *The American Economic Review* 68 (2): 1–16.

Skinner, D. J. 1994. Why Firms Voluntarily Disclose Bad News. *Journal of Accounting Research* 32 (1): 38–60.

Stokes, S. 2002. Visual Literacy in Teaching and Learning: A Literature Perspective. *Electronic Journal for the Integration of Technology in Education* 1 (1): 10–19.

Taylor, S. E., and S. T. Fiske. 1978. Salience, Attention, and Attribution: Top of the Head Phenomena. In *Advances in Experimental Social Psychology*, edited by L. Berkowitz, 11:249–288. Academic Press.

Taylor, S. E., and S. C. Thompson. 1982. Stalking the elusive “vividness” effect. *Psychological Review* 89 (2): 155–181.

Tufte, E. 2001. *The Visual Display of Quantitative Information*. Second Edition. Graphics Press LLC.

Whited, R. L., Q. T. Swanquist, J. E. Shipman, and J. R. Moon. 2022. Out of Control: The (Over) Use of Controls in Accounting Research. *The Accounting Review* 97 (3): 395–413.

Wu, S. 2021. What is the Role of Visuals in Earnings Conference Call Slides? Working Paper, UC Irvine Electronic Theses and Dissertations.

Xu, D. 2021. *Data Visualization in Earnings Conference Calls*. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network.

Table 1: Sample Selection
This table reports the sample composition.

	Obs with Slideshow
Total observations from Capital IQ with matching data on slideshow and matching identifiers on Compustat, I/B/E/S, CRSP, or TAQ data through 2020	54,645
less earnings calls prior to 2010	(893)
less where stock price less than $1 and assets less than $5 million	(396)
less financial firms (one-digit SIC-code = 6)	(10,264)
less 'slideshows' without extractable text or false positives	(4,516)
less missing call characteristics data	(899)
less missing sufficient Compustat, I/B/E/S, CRSP, intraday TAQ, or RavenPack data	(6,943)
Observations for primary analysis	30,734

Table 2: Descriptive Statistics and Correlation Matrix
This table presents descriptive statistics for the variables used in the main empirical analyses (Panel A) and the correlation matrix for the dependent and independent variables of interest (Panel B). I define all variables in Appendix B and winsorize all continuous variables at the 1st and 99th percentiles. The sample spans 2010 to 2020 and includes 30,734 observations for all variables. In Panel B, bold values represent significance at the 5% level.
Panel A: Descriptive Statistics

	Mean	Std Dev	25%	Median	75%
Market Outcomes					
Abn BidAsk Spread	0.194	0.288	0.014	0.165	0.366
Abn Price Impact	1.580	0.780	1.054	1.434	1.949
Abn Trading Vol	0.673	0.557	0.293	0.646	1.024
Abn Retail Trading Vol	0.119	0.260	0.009	0.034	0.104
Retail Trading Consensus	0.125	0.115	0.041	0.091	0.173
Disclosure Formats					
DataVis	5.972	5.207	2	5	9
Tables	2.812	3.016	1	2	4
Text	11.252	7.016	6	10	15
Control Vars					
|UnEarn|	0.006	0.014	0.001	0.002	0.005
Loss	0.256	0.437	0	0	1
|MgrTone|	0.315	0.167	0.184	0.317	0.440
|AnalystTone|	0.221	0.170	0.091	0.185	0.321
InfoMgr	-0.330	1.023	-1.033	-0.459	0.204
ObfuPres	-0.096	1.231	-0.945	-0.130	0.700
ObfuQA	-0.071	1.072	-0.818	-0.147	0.585
PctNum	0.025	0.008	0.019	0.024	0.029
Guid	0.173	0.378	0	0	0
Bundled	0.201	0.401	0	0	0
ConcurEA	227.453	135.623	121	215	337
AnalystsFore	7.994	5.370	4	7	11
Articles	41.376	30.725	19	38	57
AftHrs	0.327	0.469	0	0	1
Lag	33.690	9.919	27	33	38
MVE	13,096.466	30,360.749	985.964	2,924.805	9,948.960
Segs	5.695	4.428	2.000	5.000	8.000
MTB	3.290	5.704	1.356	2.210	3.777
MOM	0.026	0.207	-0.088	0.023	0.130
InstOwn	0.716	0.283	0.603	0.815	0.925
AnalystFoll	7.521	6.033	3.000	6.000	11.000
FirmAge	25.682	22.118	8.570	18.725	37.951
EarnVol	0.017	0.015	0.007	0.012	0.020

Panel B: Correlation Matrix

		(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
(1)	*Abn BidAsk Spread*	1	**0.523**	**0.101**	**0.011**	**-0.022**	**-0.053**	0.005	-0.005
(2)	*Abn Price Impact*	**0.513**	1	**0.201**	**0.105**	**0.027**	**-0.044**	**0.016**	**-0.015**
(3)	*Abn Trading Vol*	**0.075**	**0.183**	1	**0.725**	**-0.073**	**-0.054**	**0.019**	-0.011
(4)	*Abn Retail Trading Vol*	**-0.066**	**-0.013**	**0.515**	1	**-0.136**	**-0.013**	**0.016**	0.002
(5)	*Retail Trading Consensus*	**-0.020**	**0.026**	**-0.096**	**-0.128**	1	**-0.020**	**-0.016**	**-0.04**
(6)	*lnDataVis*	**-0.048**	**-0.038**	**-0.052**	-0.003	**-0.015**	1	**0.035**	**0.245**
(7)	*lnTables*	0.007	0.010	**0.018**	-0.005	**-0.019**	**0.019**	1	**0.161**
(8)	*lnText*	-0.005	**-0.013**	**-0.011**	-0.007	**-0.040**	**0.235**	**0.168**	1

Table 3: Information Asymmetry

This table reports the results of regression analyses of abnormal bid-ask spread on the alternative disclosure format types. The sample includes all observations from 2010 to 2020 with the necessary data to compute the model variables. I include forty-eight-digit Fama-French industry, calendar-year-quarter, and day-of-week fixed effects (unreported) and cluster standard errors by firm and calendar-year-quarter. I define all variables in Appendix A. I exclude extreme observations using Cook's Distance values greater than 4/N. *, **, and *** represent one-sided (two-sided) significance at the 10%, 5%, and 1% levels, respectively, when predictions are (not) made.

	Pred	(1)	(2)	(3)	(4)
Dependent Variable =			***Abn BidAsk Spread***		
lnDataVis	*(-)*	-0.0099***			-0.0111***
		(-4.187)			(-4.433)
lnTables	*(-)*		-0.0017		-0.0026
			(-0.614)		(-0.951)
lnText	*(-)*			0.0017	0.0062
				(0.531)	(1.837)
|UnEarn|		-0.3332***	-0.3337***	-0.3327***	-0.3261***
		(-3.813)	(-3.793)	(-3.795)	(-3.781)
Loss		-0.0452***	-0.0457***	-0.0457***	-0.0459***
		(-7.822)	(-7.936)	(-7.901)	(-7.937)
|MgrTone|		-0.0065	-0.0063	-0.0067	-0.0073
		(-0.468)	(-0.455)	(-0.475)	(-0.525)
|AnalystTone|		-0.0104	-0.0110	-0.0106	-0.0109
		(-1.496)	(-1.563)	(-1.493)	(-1.537)
InfoMgr		-0.0087***	-0.0090***	-0.0090***	-0.0088***
		(-4.782)	(-4.918)	(-4.872)	(-4.799)
ObfuPres		0.0025	0.0027*	0.0027*	0.0026
		(1.625)	(1.698)	(1.733)	(1.647)
ObfuQA		-0.0025*	-0.0026*	-0.0026*	-0.0025*
		(-1.767)	(-1.862)	(-1.872)	(-1.760)
PctNum		-0.3287	-0.3685	-0.3890	-0.2949
		(-1.196)	(-1.305)	(-1.396)	(-1.047)
Guid		0.0151**	0.0165**	0.0161**	0.0155**
		(2.334)	(2.567)	(2.474)	(2.404)
Bundled		0.0232**	0.0240**	0.0238**	0.0239**
		(2.238)	(2.327)	(2.303)	(2.312)
lnConcurEA		0.0109**	0.0108**	0.0110**	0.0111**
		(2.266)	(2.254)	(2.301)	(2.318)
lnAnalystsFore		0.0065	0.0065	0.0067	0.0059
		(1.063)	(1.055)	(1.094)	(0.976)
lnArticles		-0.0011	-0.0011	-0.0011	-0.0011
		(-0.724)	(-0.753)	(-0.714)	(-0.770)
AftHrs		-0.0177***	-0.0183***	-0.0182***	-0.0179***
		(-3.525)	(-3.698)	(-3.651)	(-3.588)
lnLag		0.0545***	0.0554***	0.0553***	0.0541***

	(3.742)	(3.778)	(3.788)	(3.669)
lnMVE	-0.0004	-0.0009	-0.0010	-0.0002
	(-0.122)	(-0.307)	(-0.349)	(-0.077)
lnSegs	0.0026	0.0025	0.0026	0.0027
	(0.618)	(0.595)	(0.619)	(0.654)
MTB	-0.0000***	-0.0000**	-0.0000**	-0.0000**
	(-3.184)	(-2.389)	(-2.336)	(-2.391)
MOM	0.0027	0.0040	0.0039	0.0028
	(0.250)	(0.369)	(0.356)	(0.254)
InstOwn	0.0836***	0.0851***	0.0851***	0.0827***
	(8.183)	(8.293)	(8.325)	(8.103)
lnAnalystFoll	-0.0015	-0.0011	-0.0010	-0.0008
	(-0.317)	(-0.237)	(-0.218)	(-0.182)
lnFirmAge	-0.0022	-0.0016	-0.0016	-0.0026
	(-0.732)	(-0.516)	(-0.531)	(-0.865)
EarnVol	-0.3838***	-0.3829***	-0.3688***	-0.3772***
	(-3.747)	(-3.789)	(-3.648)	(-3.717)
lnDataVis - lnText < 0				-0.0173***
				(12.420)
lnTables - lnText < 0				-0.0088**
				(3.671)
lnDataVis - lnTables < 0				-0.0085***
				(6.223)
Observations	29,328	29,313	29,316	29,326
Adj. R-squared	0.329	0.329	0.328	0.330

Table 4: Market Liquidity

This table reports the results of regression analyses of abnormal price impact on the alternative disclosure format types. The sample includes all observations from 2010 to 2020 with the necessary data to compute the model variables. I include forty-eight-digit Fama-French industry, calendar-year-quarter, and day-of-week fixed effects (unreported) and cluster standard errors by firm and calendar-year-quarter. I define all variables in Appendix A. I exclude extreme observations using Cook's Distance values greater than 4/N. *, **, and *** represent one-sided (two-sided) significance at the 10%, 5%, and 1% levels, respectively, when predictions are (not) made.

	Pred	(1)	(2)	(3)	(4)
Dependent Variable =			***Abn Price Impact***		
lnDataVis	*(-)*	-0.0127**			-0.0156**
		(-2.037)			(-2.380)
lnTables	*(-)*		0.0052		0.0038
			(0.747)		(0.539)
lnText	*(-)*			0.0116	0.0152
				(1.483)	(1.853)
|UnEarn|		-0.7165***	-0.7157***	-0.6921***	-0.6907***
		(-4.486)	(-4.464)	(-4.098)	(-4.102)
Loss		-0.0840***	-0.0841***	-0.0849***	-0.0846***
		(-6.715)	(-6.666)	(-6.686)	(-6.693)
|MgrTone|		-0.0749**	-0.0758**	-0.0775**	-0.0755**
		(-2.197)	(-2.237)	(-2.282)	(-2.241)
|AnalystTone|		0.0010	0.0006	-0.0002	0.0037
		(0.040)	(0.024)	(-0.008)	(0.143)
InfoMgr		-0.0298***	-0.0299***	-0.0300***	-0.0297***
		(-5.800)	(-5.706)	(-5.707)	(-5.733)
ObfuPres		0.0108**	0.0110**	0.0111**	0.0109**
		(2.301)	(2.354)	(2.389)	(2.342)
ObfuQA		-0.0008	-0.0004	-0.0003	-0.0004
		(-0.156)	(-0.085)	(-0.053)	(-0.079)
PctNum		-1.2952*	-1.4660**	-1.4200**	-1.3896**
		(-1.922)	(-2.161)	(-2.109)	(-2.068)
Guid		0.0205	0.0226	0.0232	0.0214
		(1.117)	(1.233)	(1.276)	(1.170)
Bundled		0.0886***	0.0902***	0.0903***	0.0885***
		(3.317)	(3.386)	(3.394)	(3.285)
lnConcurEA		0.0352***	0.0351***	0.0353***	0.0352***
		(3.122)	(3.132)	(3.142)	(3.130)
lnAnalystsFore		-0.0149	-0.0136	-0.0140	-0.0150
		(-0.938)	(-0.858)	(-0.886)	(-0.941)
lnArticles		0.0034	0.0035	0.0033	0.0035
		(0.953)	(0.983)	(0.913)	(0.990)
AftHrs		-0.0293**	-0.0285**	-0.0281**	-0.0276**
		(-2.200)	(-2.142)	(-2.107)	(-2.079)
lnLag		0.0523*	0.0507*	0.0505*	0.0502*

	(1.838)	(1.778)	(1.767)	(1.765)
lnMVE	-0.0394***	-0.0405***	-0.0406***	-0.0399***
	(-7.013)	(-7.213)	(-7.258)	(-7.087)
lnSegs	-0.0061	-0.0056	-0.0050	-0.0056
	(-0.690)	(-0.630)	(-0.564)	(-0.628)
MTB	0.0000	0.0000	0.0000	0.0000
	(0.082)	(0.101)	(0.098)	(0.115)
MOM	0.0460	0.0479*	0.0490*	0.0502*
	(1.672)	(1.724)	(1.760)	(1.799)
InstOwn	0.2538***	0.2551***	0.2552***	0.2520***
	(9.685)	(9.741)	(9.770)	(9.681)
lnAnalystFoll	-0.0280**	-0.0286**	-0.0286**	-0.0281**
	(-2.346)	(-2.398)	(-2.401)	(-2.349)
lnFirmAge	0.0246***	0.0251***	0.0249***	0.0240***
	(3.579)	(3.623)	(3.600)	(3.511)
EarnVol	-0.8941**	-0.8860**	-0.8899**	-0.8789**
	(-2.448)	(-2.428)	(-2.418)	(-2.409)
lnDataVis - lnText < 0				-0.0308***
				(6.487)
lnTables - lnText < 0				-0.0114
				(1.006)
lnDataVis - lnTables < 0				-0.0194**
				(4.629)
Observations	30,328	30,325	30,323	30,330
Adj. R-squared	0.148	0.148	0.148	0.148

Table 5: Investor Information Processing

This table reports the results of regression analyses of abnormal trading volume on the alternative disclosure format types. The sample includes all observations from 2010 to 2020 with the necessary data to compute the model variables. I include forty-eight-digit Fama-French industry, calendar-year-quarter, and day-of-week fixed effects (unreported) and cluster standard errors by firm and calendar-year-quarter. I define all variables in Appendix A. I exclude extreme observations using Cook's Distance values greater than 4/N. *, **, and *** represent one-sided (two-sided) significance at the 10%, 5%, and 1% levels, respectively, when predictions are (not) made.

	Pred	(1)	(2)	(3)	(4)
Dependent Variable =			***Abn Trading Volume***		
lnDataVis	*(+)*	0.0101**			0.0094**
		(2.247)			(2.072)
lnTables	*(+)*		0.0034		0.0017
			(0.690)		(0.335)
lnText	*(+)*			0.0109*	0.0073
				(1.796)	(1.199)
|UnEarn|		-0.1628**	-0.1615**	-0.1638**	-0.1864**
		(-2.283)	(-2.303)	(-2.380)	(-2.648)
Loss		-0.0168	-0.0159	-0.0170	-0.0170
		(-1.624)	(-1.537)	(-1.642)	(-1.629)
|MgrTone|		0.0577**	0.0604**	0.0597**	0.0583**
		(2.162)	(2.274)	(2.249)	(2.196)
|AnalystTone|		0.0253	0.0291*	0.0277*	0.0261*
		(1.642)	(1.937)	(1.829)	(1.734)
InfoMgr		-0.0004	-0.0008	-0.0004	-0.0006
		(-0.098)	(-0.178)	(-0.098)	(-0.133)
ObfuPres		-0.0005	-0.0005	-0.0004	-0.0003
		(-0.149)	(-0.156)	(-0.133)	(-0.095)
ObfuQA		-0.0082**	-0.0083**	-0.0082**	-0.0081**
		(-2.151)	(-2.167)	(-2.159)	(-2.127)
PctNum		-2.1419***	-2.1342***	-2.0547***	-2.1383***
		(-4.009)	(-3.992)	(-3.771)	(-3.988)
Guid		0.0241*	0.0236*	0.0238*	0.0243*
		(1.994)	(1.933)	(1.947)	(1.993)
Bundled		-0.0103	-0.0099	-0.0110	-0.0099
		(-0.875)	(-0.837)	(-0.929)	(-0.850)
lnConcurEA		-0.1013***	-0.1011***	-0.1006***	-0.1013***
		(-13.371)	(-13.422)	(-13.351)	(-13.416)
lnAnalystsFore		0.1604***	0.1589***	0.1586***	0.1592***
		(15.562)	(15.476)	(15.443)	(15.678)
lnArticles		-0.0018	-0.0017	-0.0017	-0.0015
		(-0.606)	(-0.575)	(-0.569)	(-0.517)
AftHrs		0.0493***	0.0497***	0.0506***	0.0497***
		(4.428)	(4.383)	(4.555)	(4.467)
lnLag		-0.0240	-0.0245	-0.0240	-0.0233

	(-1.066)	(-1.086)	(-1.062)	(-1.028)
lnMVE	-0.0532***	-0.0525***	-0.0524***	-0.0532***
	(-12.909)	(-12.622)	(-12.935)	(-13.131)
lnSegs	0.0149**	0.0144**	0.0144**	0.0144**
	(2.090)	(2.029)	(2.019)	(2.021)
MTB	0.0000	0.0001	0.0000	0.0001
	(1.446)	(1.680)	(1.468)	(1.500)
MOM	0.0177	0.0157	0.0172	0.0176
	(0.823)	(0.732)	(0.816)	(0.829)
InstOwn	0.1175***	0.1176***	0.1168***	0.1185***
	(6.554)	(6.603)	(6.607)	(6.663)
lnAnalystFoll	-0.0450***	-0.0449***	-0.0442***	-0.0441***
	(-5.578)	(-5.534)	(-5.443)	(-5.500)
lnFirmAge	-0.0052	-0.0062	-0.0064	-0.0054
	(-0.856)	(-1.025)	(-1.048)	(-0.891)
EarnVol	0.6164**	0.6589**	0.6321**	0.6131**
	(2.268)	(2.475)	(2.301)	(2.260)
lnDataVis - lnText < 0				0.0021
				(0.065)
lnTables - lnText < 0				-0.0057
				(0.490)
lnDataVis - lnTables < 0				0.0077
				(1.326)
Observations	29,139	29,133	29,126	29,133
Adj. R-squared	0.273	0.273	0.273	0.274

Table 6: Retail Investor Information Processing

This table reports the results of regression analyses of abnormal retail trading volume on the alternative disclosure format types. The sample includes all observations from 2010 to 2020 with the necessary data to compute the model variables. I include forty-eight-digit Fama-French industry, calendar-year-quarter, and day-of-week fixed effects (unreported) and cluster standard errors by firm and calendar-year-quarter. I define all variables in Appendix A. I exclude extreme observations using Cook's Distance values greater than 4/N. *, **, and *** represent one-sided (two-sided) significance at the 10%, 5%, and 1% levels, respectively, when predictions are (not) made.

	Pred	(1)	(2)	(3)	(4)
Dependent Variable =		***Abn Retail Trading Volume***			
lnDataVis	*(+)*	0.0051**			0.0059***
		(2.212)			(2.461)
lnTables	*(+)*		0.0033*		0.0038*
			(1.378)		(1.537)
lnText	*(+)*			-0.0020	-0.0049
				(-0.640)	(-1.499)
|UnEarn|		0.3199**	0.3231***	0.3226***	0.3151**
		(2.678)	(2.721)	(2.706)	(2.644)
Loss		0.0200***	0.0198***	0.0200***	0.0196***
		(4.401)	(4.318)	(4.363)	(4.311)
|MgrTone|		-0.0033	-0.0021	-0.0020	-0.0028
		(-0.284)	(-0.179)	(-0.172)	(-0.237)
|AnalystTone|		0.0246***	0.0247***	0.0249***	0.0245***
		(4.304)	(4.295)	(4.335)	(4.217)
InfoMgr		0.0068***	0.0069***	0.0069***	0.0069***
		(3.603)	(3.639)	(3.632)	(3.649)
ObfuPres		0.0008	0.0008	0.0008	0.0008
		(0.571)	(0.564)	(0.567)	(0.543)
ObfuQA		-0.0020	-0.0021	-0.0021	-0.0020
		(-1.295)	(-1.339)	(-1.349)	(-1.302)
PctNum		-0.6238**	-0.6158**	-0.5792**	-0.6702***
		(-2.695)	(-2.642)	(-2.521)	(-2.855)
Guid		0.0053	0.0041	0.0041	0.0049
		(0.780)	(0.601)	(0.605)	(0.716)
Bundled		-0.0114**	-0.0119**	-0.0116**	-0.0112*
		(-2.024)	(-2.100)	(-2.030)	(-1.969)
lnConcurEA		-0.0307***	-0.0310***	-0.0310***	-0.0307***
		(-8.625)	(-8.763)	(-8.779)	(-8.626)
lnAnalystsFore		0.0497***	0.0495***	0.0496***	0.0495***
		(9.736)	(9.684)	(9.651)	(9.688)
lnArticles		-0.0002	-0.0001	-0.0002	-0.0001
		(-0.173)	(-0.090)	(-0.122)	(-0.077)
AftHrs		0.0219***	0.0225***	0.0223***	0.0222***
		(4.566)	(4.651)	(4.598)	(4.582)
lnLag		0.0076	0.0079	0.0083	0.0080

	(0.831)	(0.857)	(0.896)	(0.881)
lnMVE	-0.0198***	-0.0196***	-0.0195***	-0.0198***
	(-9.347)	(-9.250)	(-9.290)	(-9.394)
lnSegs	0.0024	0.0022	0.0023	0.0020
	(0.747)	(0.680)	(0.710)	(0.628)
MTB	0.0000**	0.0000**	0.0000**	0.0000**
	(2.304)	(2.328)	(2.266)	(2.324)
MOM	-0.0146	-0.0147	-0.0150	-0.0143
	(-1.134)	(-1.116)	(-1.129)	(-1.098)
InstOwn	0.0001	-0.0004	-0.0003	0.0002
	(0.011)	(-0.045)	(-0.038)	(0.025)
lnAnalystFoll	0.0111***	0.0110***	0.0110***	0.0114***
	(3.262)	(3.225)	(3.232)	(3.337)
lnFirmAge	-0.0024	-0.0028	-0.0026	-0.0022
	(-0.870)	(-0.987)	(-0.941)	(-0.794)
EarnVol	0.5062***	0.5218***	0.5072***	0.5167***
	(4.004)	(4.083)	(3.931)	(4.087)
lnDataVis - lnText < 0				0.0109**
				(5.639)
lnTables - lnText < 0				0.0087**
				(3.765)
lnDataVis - lnTables < 0				0.0022
				(0.402)
Observations	30,190	30,193	30,194	30,193
Adj. R-squared	0.146	0.146	0.146	0.146

Table 7: Retail Investor Trading Consistency

This table reports the results of regression analyses of retail investor trading consistency on the alternative disclosure format types. The sample includes all observations from 2010 to 2020 with the necessary data to compute the model variables. I include forty-eight-digit Fama-French industry, calendar-year-quarter, and day-of-week fixed effects (unreported) and cluster standard errors by firm and calendar-year-quarter. I define all variables in Appendix A. I exclude extreme observations using Cook's Distance values greater than 4/N. *, **, and *** represent two-sided significance at the 10%, 5%, and 1% levels, respectively.

	Pred	(1)	(2)	(3)	(4)
Dependent Variable =			***Retail Trading Consensus***		
lnDataVis	*(?)*	-0.0005			-0.0005
		(-0.724)			(-0.641)
lnTables	*(?)*		-0.0009		-0.0009
			(-1.138)		(-1.040)
lnText	*(?)*			-0.0007	-0.0004
				(-0.649)	(-0.315)
|UnEarn|		-0.0629***	-0.0631***	-0.0641***	-0.0635***
		(-5.571)	(-5.563)	(-5.506)	(-5.561)
Loss		-0.0051***	-0.0051***	-0.0051***	-0.0052***
		(-4.177)	(-4.222)	(-4.343)	(-4.293)
|MgrTone|		-0.0105**	-0.0100**	-0.0097**	-0.0099**
		(-2.300)	(-2.236)	(-2.146)	(-2.198)
|AnalystTone|		-0.0035	-0.0037	-0.0038	-0.0035
		(-1.012)	(-1.076)	(-1.100)	(-1.006)
InfoMgr		-0.0006	-0.0007	-0.0006	-0.0006
		(-0.867)	(-0.971)	(-0.893)	(-0.870)
ObfuPres		0.0005	0.0004	0.0005	0.0004
		(0.858)	(0.776)	(0.858)	(0.733)
ObfuQA		-0.0009	-0.0009	-0.0009	-0.0009
		(-1.441)	(-1.470)	(-1.514)	(-1.486)
PctNum		-0.0293	-0.0204	-0.0284	-0.0124
		(-0.372)	(-0.250)	(-0.360)	(-0.152)
Guid		-0.0023	-0.0022	-0.0022	-0.0023
		(-1.375)	(-1.290)	(-1.310)	(-1.332)
Bundled		-0.0040**	-0.0040**	-0.0040**	-0.0040**
		(-2.486)	(-2.484)	(-2.469)	(-2.438)
lnConcurEA		0.0045***	0.0045***	0.0045***	0.0045***
		(6.042)	(6.071)	(6.090)	(5.939)
lnAnalystsFore		-0.0185***	-0.0185***	-0.0185***	-0.0184***
		(-13.628)	(-13.802)	(-13.801)	(-13.688)
lnArticles		0.0001	0.0001	0.0001	0.0000
		(0.221)	(0.208)	(0.251)	(0.108)
AftHrs		-0.0024*	-0.0025*	-0.0025*	-0.0025*
		(-1.908)	(-1.963)	(-1.987)	(-1.999)
lnLag		-0.0048*	-0.0043*	-0.0046*	-0.0044*

	(-1.925)	(-1.713)	(-1.846)	(-1.773)
lnMVE	-0.0088***	-0.0088***	-0.0088***	-0.0087***
	(-14.140)	(-14.185)	(-14.104)	(-14.011)
lnSegs	0.0010	0.0011	0.0012	0.0012
	(0.873)	(0.900)	(0.991)	(0.968)
MTB	-0.0000	-0.0000	-0.0000	-0.0000
	(-1.118)	(-1.174)	(-1.137)	(-1.156)
MOM	-0.0045	-0.0044	-0.0044	-0.0046
	(-1.404)	(-1.399)	(-1.402)	(-1.460)
InstOwn	0.0056**	0.0055**	0.0056**	0.0054**
	(2.216)	(2.181)	(2.257)	(2.155)
lnAnalystFoll	-0.0078***	-0.0077***	-0.0077***	-0.0077***
	(-7.560)	(-7.563)	(-7.418)	(-7.406)
lnFirmAge	0.0005	0.0006	0.0006	0.0006
	(0.557)	(0.670)	(0.699)	(0.639)
EarnVol	-0.2490***	-0.2549***	-0.2509***	-0.2490***
	(-6.749)	(-6.787)	(-6.615)	(-6.698)
Observations	29,373	29,372	29,371	29,371
Adj. R-squared	0.100	0.100	0.100	0.100

Table 8: Additional Analysis

This table reports the results of regression analyses of each of the previous analyses with the number of slides with at least one of each different type of data visualizations (Panel A) and the number of slides with one, two, three, or four or more data visualizations on each slide (Panel B). The sample includes all observations from 2010 to 2020 with the necessary data to compute the model variables. I include forty-eight-digit Fama-French industry, calendar-year-quarter, and day-of-week fixed effects (unreported), control variables from previous models (unreported), and cluster standard errors by firm and calendar-year-quarter. I exclude extreme observations using Cook's Distance values greater than 4/N. *, **, and *** represent one-sided significance at the 10%, 5%, and 1% levels, respectively, in Panel A and Panels B, where applicable.

Panel A: Types of Data Visualizations Slides

	(1)	(2)	(3)	(4)	(5)
Dependent Variable =	***Abn BidAsk Spread***	***Abn Price Impact***	***Abn Trading Vol***	***Abn Retail Trading Vol***	***Retail Trading Consistency***
lnBarCharts	-0.0060**	-0.0030	0.0091**	0.0077***	-0.0011
	(-2.101)	(-0.412)	(1.756)	(2.800)	(-1.493)
lnLineGraphs	-0.0194***	-0.0367***	-0.0127	0.0005	0.0014
	(-5.188)	(-4.000)	(-1.592)	(0.145)	(1.283)
lnPieGraphs	0.0014	0.0064	0.0119**	-0.0025	0.0001
	(0.402)	(0.692)	(1.785)	(-0.765)	(0.125)
lnMaps	0.0021	-0.0055	0.0088	0.0003	-0.0014
	(0.565)	(-0.522)	(1.159)	(0.108)	(-0.963)
lnBarCharts = lnLineGraphs	6.490**	6.627**	3.874*	2.329	
lnBarCharts = lnPieGraphs	2.182	0.519	0.099	4.492**	
lnBarCharts = lnMaps	2.724	0.033	0.001	3.465*	
lnLineGraphs = lnPieGraphs	14.940***	10.780***	5.206**	0.452	
lnLineGraphs = lnMaps	14.230***	4.709**	3.349*	0.001	
lnPieGraphs = lnMaps	0.015	0.633	0.079	0.329	
Controls	Yes	Yes	Yes	Yes	Yes
Observations	29,337	30,332	29,137	30,193	29,370
Adj. R-squared	0.330	0.148	0.274	0.146	0.099

Panel B: Number of Data Visualizations per Slide					
	(1)	(2)	(3)	(4)	(5)
Dependent Variable =	***Abn BidAsk Spread***	***Abn Price Impact***	***Abn Trading Vol***	***Abn Retail Trading Vol***	***Retail Trading Consistency***
lnDataVis1	-0.0089***	-0.0112**	0.0022	0.0033*	-0.0004
	(-3.474)	(-1.737)	(0.449)	(1.410)	(-0.481)
lnDataVis2	-0.0109***	-0.0219***	0.0063	0.0049*	0.0006
	(-3.908)	(-2.823)	(1.105)	(1.657)	(0.572)
lnDataVis3	0.0044	0.0130	0.0077	0.0030	-0.0025**
	(1.338)	(1.262)	(1.098)	(0.994)	(-2.082)
lnDataVis4	-0.0018	-0.0094	0.0081	0.0006	0.0001
	(-0.383)	(-0.670)	(1.000)	(0.154)	(0.057)
lnDataVis1 = lnDataVis2	0.237	0.963	0.239	0.159	
lnDataVis1 = lnDataVis3	9.734***	3.553*	0.38	0.005	
lnDataVis1 = lnDataVis4	1.683	0.015	0.366	0.300	
lnDataVis2 = lnDataVis3	12.510***	7.123**	0.022	0.186	
lnDataVis2 = lnDataVis4	2.675	0.528	0.030	0.703	
lnDataVis3 = lnDataVis4	1.188	1.457	0.001	0.226	
Controls	Yes	Yes	Yes	Yes	Yes
Observations	29,322	30,324	29,126	30,189	29,374
Adj. R-squared	0.331	0.148	0.274	0.146	0.099

Figure 1: Proportion of Earnings Calls with Slides

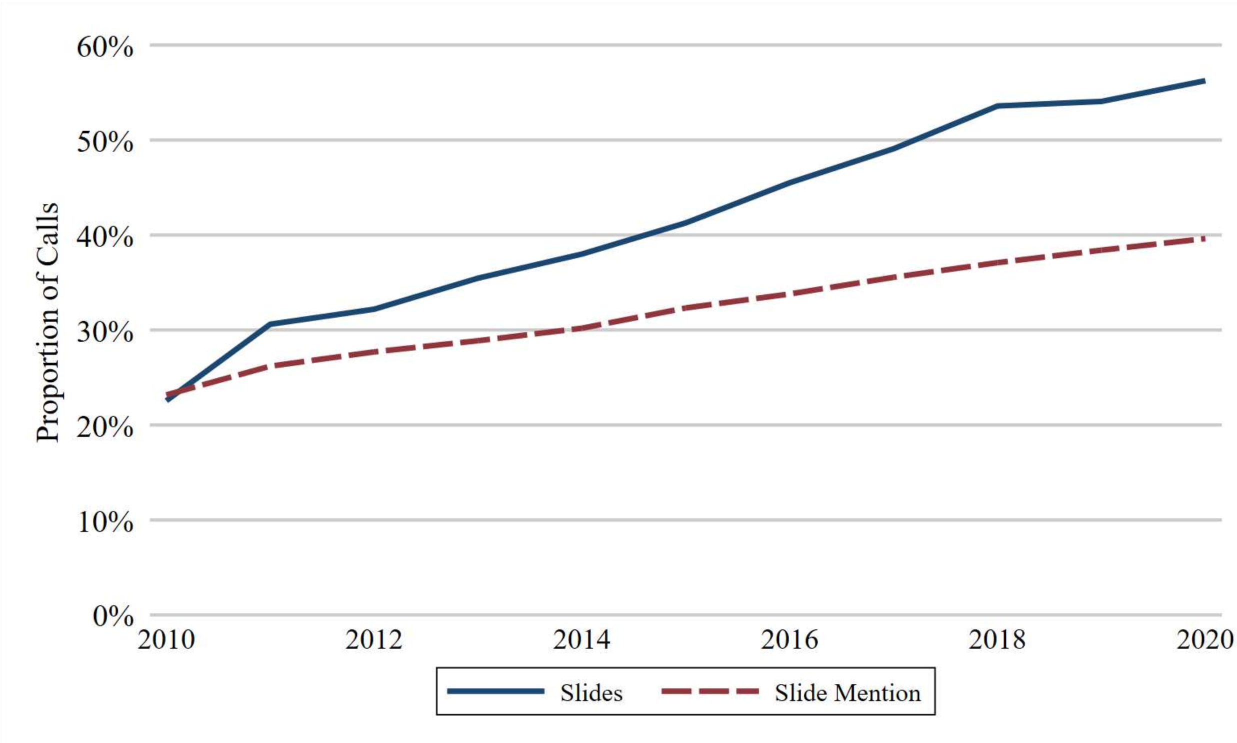


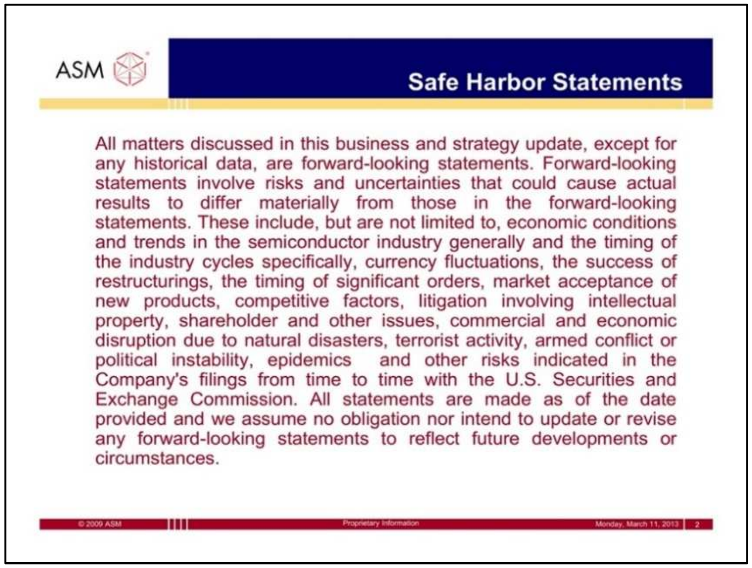
Figure 2: Slide Classification Word Clouds

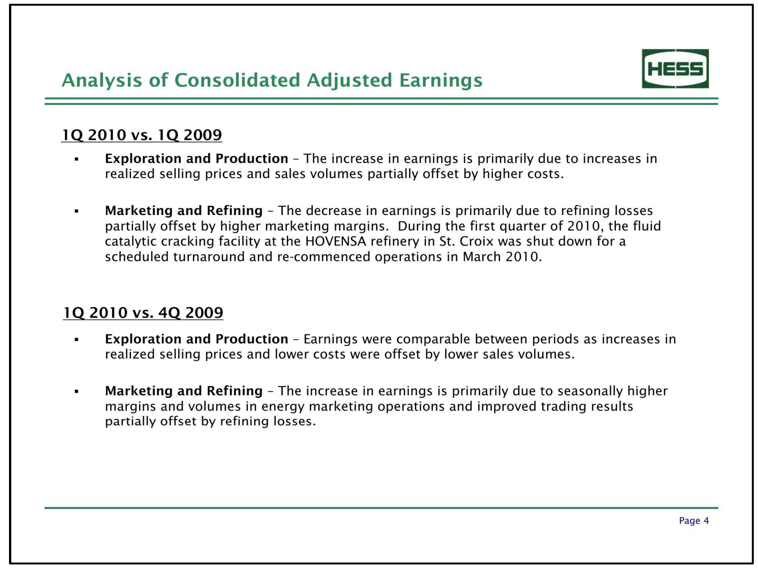
Data Visualizations

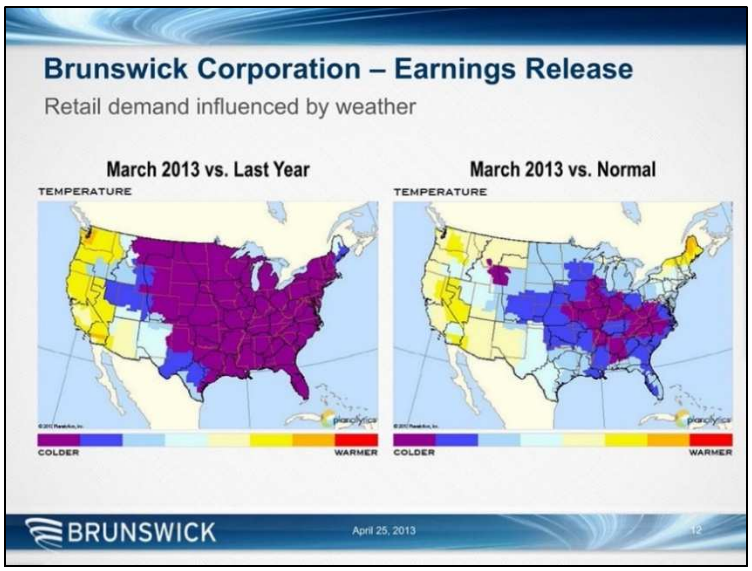
Tables

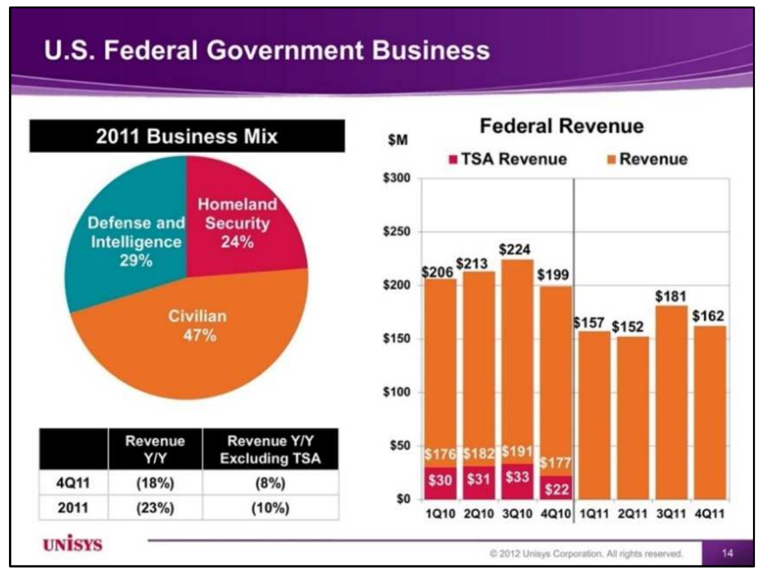
Text

Appendix A: Slide examples









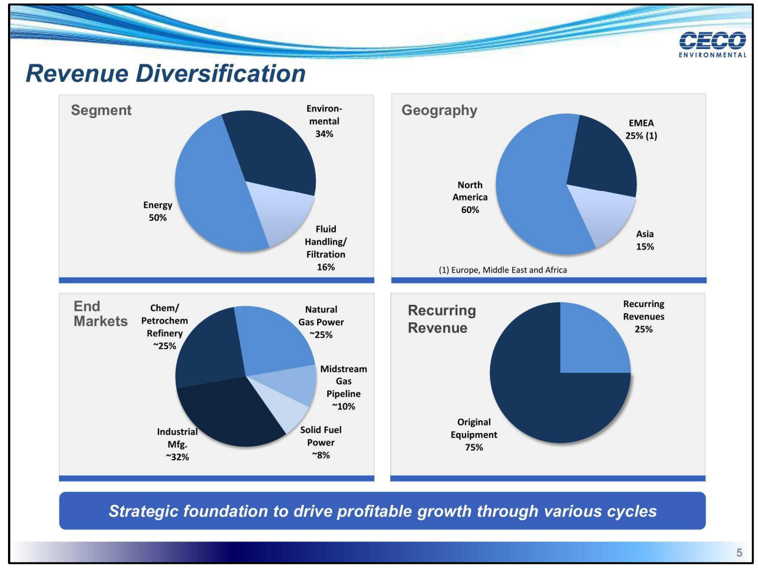
Unaudited Calculation of Adjusted EBITDA Leverage Ratio (in thousands)

LEVERAGE RATIO - Post Linc Acquisition		
(1) Outstanding borrowings *	$	452,461
(2) Adjusted EBITDA **	$	184,023
Ratio of (1) to (2)		2.46
LEVERAGE RATIO - October 31, 2011		
(1) Outstanding borrowings *	$	300,000
(2) Adjusted EBITDA **	$	184,023
Ratio of (1) to (2)		1.63

* Outstanding borrowings immediately following the acquisition of Linc on December 1, 2010.
** Adjusted EBITDA for trailing twelve months as of October 31, 2011 (for comparison purposes).

ABM Industries Incorporated

• 15 •



PLEXUS CORP.

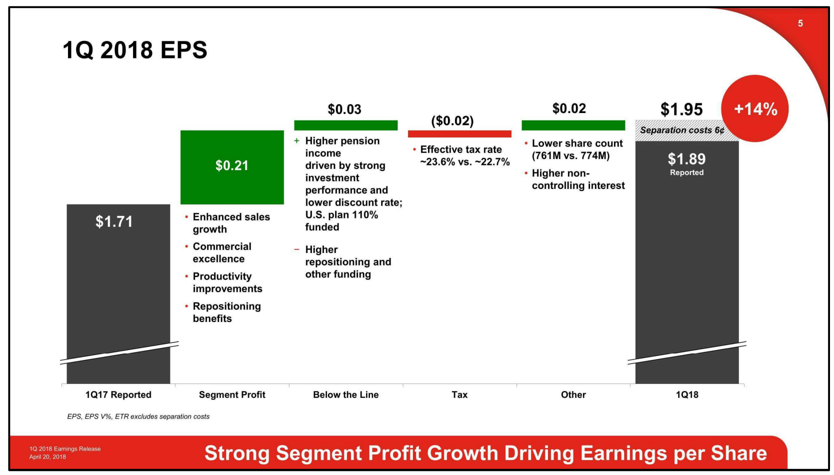
Fiscal fourth quarter results

	Q4F20 Results	Q4F20 Guidance	Q3F20 Results
Revenue ($ millions)	$913	$850 to $890	$857
GAAP Diluted EPS	$1.26(1)	$1.05 to $1.20(2)	$1.20(3)

- Record quarterly revenue representing 7% sequential growth and 13% y-o-y growth
- Strong results from I/C and HC/LS
- GAAP operating margin of 5.5% - best result in over a decade

(1) Includes $0.23 stock-based compensation expense
(2) Includes $0.21 stock-based compensation expense
(3) Includes $0.22 stock-based compensation expense

3



Appendix B: Variable Definitions

Variable	***Description***
Abnormal BidAsk Spread	is the log of (the weighted average daily spread over trading days [0,1] divided by the weighted average daily spread over days [-41, -11]), following Blankespoor et al. (2020). Daily spread is average percent effective spread. Daily spreads are weighted based on total number of trades during market hours.
Abnormal Price Impact	is the weighted average daily impact over trading days [0,1] divided by the weighted average daily impact over days [-41, -11]), following Blankespoor et al. (2020). Daily price impact is the average percent price impact of each trade. Daily impacts are weighted based on total number of trades during market hours.
Abnormal Trading Vol	is the log of (the average daily volume over trading days [0, 1] divided by one + the average daily volume over days [-41, -11]), following Blankespoor et al. (2020). Daily volume is the number of shares traded divided by total shares outstanding.
Abnormal Retail Trading Vol	is the log of ((one + the average daily retail volume over trading days [0,1]) divided by (one + the average daily retail volume over days [-41, -11]), multiplied by 100, following Blankespoor et al. (2020). Daily retail volume is the number of retail shares traded divided by total shares outstanding. Retail trades are identified using the method developed by Boehmer et al. (2021).
Retail Trading Consensus	is the absolute value of the average of (retail buy orders minus retail sell orders) divided by (retail buy order plus retail sell orders) over trading days [0, 1], following Boehmer et al. (2021).
DataVis	is the number of slides that contain at least one data visualization (i.e., bar charts, line graphs, pie graphs, and maps). Data visualizations are identified through a machine learning object detection image recognition process defined in Appendix C. I download slideshows from Capital IQ.
Tables	is the log of one plus the number of slides that contain at least one table, but do not contain any data visualizations. Tables are identified as any slide with the number to number plus words ratio above 33% and more than 75 words and numbers.
Text	is the log of one plus the number of slides that do not contain either data visualizations or tables.
|UnEarn|	is the firm's earnings per share less the median consensus, scaled by the price two days before the earnings announcement.
Loss	is an indicator if income before extraordinary items is negative.
|MgrTone|	is the absolute value of (positive tone words minus negative tone words) divided by (positive tone words plus negative tone words) of managers over the entirety of the call. Positive and negative tone words are determined using the dictionary from Loughran and McDonald (2014) with the exclusion of question from the negative dictionary.

|AnalystTone|	is the absolute value of (positive tone words minus negative tone words) divided by (positive tone words plus negative tone words) of analyst questions from the earnings call.
InfoMgr	is the first principal component of the two separate regressions of spoken complexity by managers during the presentation and responses on spoken complexity by analysts and other firm characteristics, following Bushee et al. (2018) columns 2 and 4 of table 2.
ObfuPres	is the residual from the regression of spoken complexity by managers during the presentation on spoken complexity by analysts and other firm characteristics, following Bushee et al. (2018) column 2 of table 2.
ObfuQA	is the residual from the regression of spoken complexity by managers during the responses on spoken complexity by analysts and other firm characteristics, following Bushee et al. (2018) column 4 of table 2.
PctNum	is the number of numbers divided by the (number of numbers plus the number of words) spoken during the earnings conference call.
Guid	is an indicator if the firm issues management guidance on the same day as their earnings announcement.
Bundled	is an indicator if the firm files their 10-K/Q on the same day as their earnings announcement.
ConcurEA	is the number of other firm earnings announcements that occur on the same day.
AnalystsFore	is the number of earnings per share forecasts issued by analysts over days [0, 1].
Articles	is the number of earnings-related news articles written over days [0, 1].
AftHrs	is an indicator if the firm holds the earnings call after the close of market (i.e., 4 PM EST).
Lag	is the number of days between the earnings conference call and fiscal quarter end date.
MVE	is the total common shares outstanding times the price.
Segs	is the number of business and geographic segments.
MTB	is the market value of equity by the book value of common equity divided.
MOM	is the buy-and-hold return over the previous quarter.
InstOwn	is the percentage of shares held by institutional owners at the end of the most recent calendar quarter.
AnalystFoll	is the number of analysts that issued at least one earnings per share forecast for the firm-quarter.
FirmAge	is the number of years since the firm first appeared on CRSP.
EarnVol	is the standard deviation of cash flows from operations scaled by total assets over the prior five years, requiring a minimum of three years.

Appendix C: Machine Learning Process

I begin collecting my sample of earnings call slide decks by downloading the PDF versions available on Capital IQ's website.[18] After downloading each PDF version, I use Python to convert each PDF page into individual slide images and extract the text from each page. Then, I use Python to examine the data visualization content of the earnings call slides. Specifically, I use PyTorch to create a custom object detection dataset rather than a standard image recognition model because object detection identifies the location of the specific category type on the image which allows the model to recognize multiple category types on the same slide. Thus, a slide with multiple pie graphs will be able to identify multiple different pie graphs rather than just the occurrence of a pie graph. My evidence suggests that the average number of data visualizations on a slide with a data visualization is approximately 1.72. To train my sample, I choose approximately 775 slides with a variety of styles (i.e., containing only text or tables or differing amounts, locations, colors, and mix of text/tables, bar charts, pie graphs, line graphs, and/or maps)[19,20]. One concern is the size of my training sample is relatively small. However, three aspects of object detection models alleviate this concern. First, I label 4,743 objects across the 775 images. Second, object detection requires a labelling application programming interface (API) which focuses the model on the relevant RGB pixels to identify the image content increasing the accuracy of the model. Third, object detection models use pre-trained models and leverage the information from those models to apply it to my

[18] To my knowledge, no other source exists with broad firm coverage of earnings call slide shows. I do not believe this introduces a sample selection problem. However, I cannot rule out the possibility that my results do not generalize to other earnings call slide show settings.

[19] I do not separately identify data maps vs. location maps as done in Christensen et al. (2021) because I cannot programmatically distinguish between a data map and location map. The main distinguishing factor between those two types of maps is the use of quantitative information within the graph. While I can extract the text on the slide show, it is more difficult to identify whether the numerical information is part of the map or a legend/table under the graph.

[20] I do not explore other types of data visualizations, namely qualitative data visualizations, as done in Christensen et al. (2021) because I cannot train a computer to identify different types of qualitative data visualizations based on RGB pixel values. They identify these different types of qualitative data visualizations because they manually code each image. Morover, extracting the text from the slides would not be fruitful due to the variety of information displayed on each slide and differing types of qualitative data visualization information type.

custom dataset. From these 775 images, I use a labelling API to draw boxes around each different type of image content (e.g., text, bar charts, or pie graphs) and then use these images to train my custom object detection model. I hold out a random sample of approximately 155 slides to avoid overfitting.

After the labelling I estimate my object detection model. I use a Faster Region Based Convolution Neural Network (Faster R-CNN) model as the baseline model that I customize to my dataset through the training and testing sample. The Faster R-CNN model goes through the following four-step process. First, using a region proposal network algorithm (RPN), the algorithm generates locations or bounding boxes that contain possible objects within the image. Second, a feature generation stage obtains features for the possible identified objects using a convolution neural network. Third, a classification layer predicts to which classification each identified object belongs. Finally, a regression layer makes coordinates around the object more precise.

Once the Faster R-CNN model is complete, I apply it to my slide images. Prior to testing the accuracy of the model, I extract the text from each slide and eliminate false positives based on the content of the slides (i.e., question and answer, executive, agenda, safe harbor, and cover slides). Overall, from a sample of 1915 randomly selected individual slides, my out of sample accuracy rate is 92%, meaning the custom object detection model perfectly identified everything on the slide (i.e., if the slide had multiple pie graphs, they were all identified). Moreover, within the sample of 1,915 slides, my model identified 1,183 data visualizations across 661 slides with an accuracy rate of 86%.[21]

[21] Type 1 and type 2 errors were 10% and 4%, respectively.