Received: 12 July 2021 | Accepted: 28 October 2023

DOI: 10.1111/1911-3846.12919

ARTICLE

CAAA ACPC CONTEMPORARY ACCOUNTING RESEARCH | RECHERCHE COMPTABLE CONTEMPORAINE

Financial statement adequacy and firms' MD&A disclosures

Stephen V. Brown[1] | **Lisa A. Hinson**[2] | **Jennifer Wu Tucker**[2]

[1]University of Connecticut, Storrs, Connecticut, USA

[2]Fisher School of Accounting, University of Florida, Gainesville, Florida, USA

Correspondence
Jennifer Wu Tucker, Fisher School of Accounting, University of Florida, Gainesville, FL, USA.
Email: jenny.tucker@warrington.ufl.edu

Abstract

Firms are required to provide financial information via the financial statements and the management discussion and analysis (MD&A), a narrative explanation of the financial statements. Our study examines how firms use the MD&A channel when their financial statement channel is inadequate. We focus on two textual attributes of the MD&A: non-GAAP disclosure and forward-looking statements. We find that firms with less adequate financial statements discuss non-GAAP measures more and provide a larger number of forward-looking statements. We then identify the topics, and therefore the context, in which non-GAAP and forward-looking disclosures are provided. Our study provides evidence on *how* managers use the MD&A, a relatively more flexible channel, to provide information when their financial statement channel is less adequate.

KEYWORDS

deep learning, disclosure, machine learning, MD&A, text segmentation, topic analysis

La pertinence des états financiers et la communication d'informations des entreprises dans le rapport de gestion

Résumé

Les entreprises sont tenues de fournir des informations financières au moyen des états financiers et du rapport de gestion, un texte présentant les états financiers. La présente étude examine comment les entreprises ont recours au rapport de gestion lorsque les états financiers sont non pertinents. Les auteurs se concentrent sur deux attributs textuels du rapport de gestion : les informations non conformes aux PCGR et les énoncés prospectifs. Ils constatent

Accepted by Peter Clarkson.

que les entreprises dont les états financiers sont moins pertinents discutent davantage des mesures non conformes aux PCGR et fournissent un plus grand nombre d'énoncés prospectifs. Ils identifient ensuite les thèmes, et donc le contexte, pour lesquels des informations non conformes aux PCGR et prospectives sont fournies. L'analyse montre *comment* les gestionnaires utilisent le rapport de gestion, un outil relativement plus souple, pour fournir des informations lorsque les états financiers sont moins pertinents.

MOTS-CLÉS

analyse thématique, apprentissage automatique, apprentissage profond, communication d'information, rapport de gestion, segmentation de texte

1 | INTRODUCTION

Firms are required to provide financial information to the public via two channels in their reports filed with the SEC. The financial statement channel includes the balance sheet, income statement, cash flow statement, statement of shareholders' equity, and the accompanying notes (hereinafter, "financial statements"), all presented in Item 8 of an annual report.[1] This channel is regulated by GAAP, which has specific rules for almost every financial statement item, and is audited for compliance with GAAP. The management discussion and analysis (MD&A) channel includes disclosure about liquidity and capital resources, results of operations, and other issues required by the SEC, as well as voluntary disclosure, all presented in Item 7 of the annual report. The MD&A is intended to "provide a *narrative* explanation of a company's financial statements and other statistical data" and "provide management with *flexibility* to describe the financial matters impacting the registrant" (SEC, 2008, emphasis added). Our study examines how managers use the MD&A channel when their financial statement channel is less adequate.[2]

This research question is important for two reasons. First, concerns about the adequacy of GAAP are widespread. Critics argue that accounting rules are more appropriate for a traditional fixed-assets-based economy and less adequate for a knowledge-based economy (Amir & Lev, 1996; S. Brown et al., 1999; Lev, 1989; Lev & Zarowin, 1999; Srivastava, 2014). As a result, the financial statements prepared under GAAP by some firms may not adequately communicate information about these firms' financial positions, operating results, and future prospects to investors. It is thus critical to investigate what managers do when their financial statement channel is constrained by GAAP.

Second, the MD&A is arguably the most widely read section of a financial report beyond the financial statement section (Li, 2010; Tavcar, 1998). Financial analysts acknowledge that they most frequently rely upon the MD&A over other disclosure sections of an annual report when performing their tasks (Knutson, 1993; Rogers & Grant, 1997). The MD&A was mandated by the SEC via an amendment to Regulation S-K in 1980 with three objectives: (1) providing a *narrative* explanation of a firm's financial statements that enables investors to see the firm through the eyes of management, (2) providing the *context* within which financial information should be analyzed, and (3) providing information to assist investors in predicting *future* performance (SEC, 2008, 2013). The SEC has conducted several reviews of MD&A practices and

[1]We use "financial statements" to refer to the four financial statements and their accompanying notes because the notes are (1) an integral part of the financial statements and (2) prepared under GAAP.

[2]We focus on the financial statements and the MD&A channels because both are mandatory and presented back-to-back in an annual report. Managers have other channels to communicate financial information with the public, for example, press releases and conference calls. Examining other channels is outside of the scope of this study.

issued multiple interpretive releases, suggesting a significant emphasis on the MD&A. Our evidence sheds light on *how* firms use this important communication channel.

If managers follow the spirit of MD&A regulation, they are expected to provide more information in the MD&A when the financial statement channel cannot adequately communicate financial information to investors. When the financial statement channel is in fact adequate, managers can save the effort of using the MD&A to communicate with investors. These expectations are consistent with the theoretical prediction of Einhorn (2005), who demonstrates that when one communication channel is constrained, managers make more use of another, less-constrained communication channel.

The above expectations may not materialize for two reasons. First, managers might view financial reporting as a compliance task and choose straightforward compliance over providing information in the best possible way (Dichev et al., 2013). For example, one CFO said, “Because at the end of the day, how should I spend my time? Do I want to spend my time working on this? Or do I want to spend my time working on strategy and driving the business? We’re not going to let the accounting wag the business here, so we’re just going to comply” (Dichev et al., 2013, pp. 23–24). In a compliance mindset, managers will meet the minimum disclosure requirements irrespective of the adequacy of the financial statement channel and save their firms’ resources for other activities. Second, managers might exercise their discretion opportunistically. Prior research finds that when firms report lower earnings or manipulate earnings, the MD&A is less readable (Li, 2008; Lo et al., 2017). Opportunistic managers are unlikely to prioritize providing information to investors in the best possible way.

Our primary proxy for the adequacy of the financial statement channel is the value relevance of earnings and book value of equity. “Value relevance” is defined as the ability of a firm’s financial statements to capture or summarize information that affects the firm’s stock value (Francis & Schipper, 1999). A large body of research uses the explanatory power of earnings and book value of equity for stock price or the association of earnings with stock price (or their respective changes forms) to measure value relevance (Barth et al., 2001). Following Banker et al. (2009), we use the R^2 from regressing a firm’s stock price on its EPS and book value of equity per share in a time series as the measure of value relevance for the sample year. We refer to one minus this measure as “financial statement inadequacy.” In supplementary analyses, we use alternative proxies constructed from a firm’s information in the sample year alone.

We focus on two textual attributes of the MD&A channel. The first is the intensity of non-GAAP disclosure. If managers intend to provide context within which their financial statements should be analyzed, as stated in the second regulatory objective of the MD&A, managers could discuss non-GAAP measures, which are customized to their firm’s situation. In other words, non-GAAP disclosure encapsulates the firm’s performance that managers tailor to reflect the firm’s particular situation. The second attribute is the number of forward-looking statements (FLS). If the financial statements, which summarize historical transactions, are inadequate for helping investors project future performance, managers may provide FLS to bridge the information gap and achieve the third regulatory objective of the MD&A.

We use two textual analysis approaches to measure MD&A textual attributes. To identify non-GAAP disclosure in the MD&A, we use the traditional keyword approach, which is simple and relatively accurate for this task (Jo & Yang, 2020). To classify FLS, we use an advanced approach because it performs substantially better than the keyword approach. Specifically, we use a deep learning model, implemented with a convolutional neural network (CNN), to identify FLS as well as the FLS that contain quantitative information (see Bochkay et al., 2023 for a review of deep learning models). We train the machine to recognize patterns of consecutive words, numbers, and symbols for predefined categories in a small hand-coded sample and then let the machine identify and classify unseen phrases and sentences in our full sample. Compared with the keyword approach, the deep learning approach can learn subtle relationships that the keyword approach cannot and is robust to coding errors as long as the training sample is large enough.

Using firm-years ended in 2004–2016, we document several findings. First, firms with less adequate financial statements discuss non-GAAP measures more in the MD&A. The proportional increase in the intensity of non-GAAP disclosure from firms in the most adequate group to firms in the least adequate group is 19.0%. This result suggests that firms with a less adequate financial statement channel turn to the MD&A channel to *customize* financial information for investors.[3]

Our second finding is that firms with less adequate financial statements provide more FLS. On average, our sample firms provide 30 forward-looking sentences in the MD&A, including 12 forward-looking sentences that contain quantitative and therefore hard information. From the most adequate group to the least adequate group, the number of FLS increases by 1.508 and the number of quantitative FLS increases by 0.613, equivalent to 5.1% and 4.9% of the unconditional intensity, respectively. These results suggest that firms with less adequate financial statements provide more forward-looking information for investors to project the future.

We then identify the topics, and therefore the context, in which non-GAAP and forward-looking disclosures appear. Using TopicTiling, a combination of topic modeling and text segmentation, we split an MD&A document into text segments and assign each text segment (i.e., consecutive sentences that discuss the same topic) a latent topic. Our topic model discovers 133 topics, which are sorted into 21 categories and then aggregated into four groups: required, traditional, intangibles, and other. We find that firms with less adequate financial statements discuss non-GAAP measures more in required topics, traditional topics, and intangibles topics and provide more FLS in required topics and intangibles topics. These results suggest that firms provide additional disclosures in areas where traditional financial reporting is likely the weakest.[4]

Our study contributes to the accounting literature by emphasizing what managers do when the boundaries of GAAP constrain the provision of adequate financial statement information to investors. Amir and Lev (1996) examine the wireless communications industry and find that earnings and book value of equity do not explain a firm's market value; however, the authors find that firms provide nonfinancial information to help investors. Our study extends this research by reporting that managers also use the MD&A to address financial statement deficiencies. Lang and Lundholm (1993) examine analysts' ratings of a firm's overall disclosure, which includes annual reports, quarterly reports, press releases, and investor relations. They find higher ratings for firms with weaker earnings-return correlations, suggesting that firms provide additional disclosure when their financial statements are inadequate. We provide direct evidence on the specific types of corporate disclosures supplied when financial statements are inadequate.

From a broader perspective, managers have a breadth of channels to provide information to investors. It is unclear whether managers adopt a mosaic approach, characterized by using multiple channels in combination, or a fragmented approach, characterized by isolated compliance with individual disclosure requirements. Few studies examine more than one managerial decision at a time and even fewer studies examine a less flexible channel in combination with a more flexible channel. Our findings are consistent with the mosaic approach.

[3]Given the paucity of evidence in the literature about non-GAAP disclosure in the MD&A, we hand-coded 200 firms randomly selected from those that provide non-GAAP disclosure in both the earnings announcement press release and the subsequent MD&A. We observe that firms provide a variety of non-GAAP measures beyond the traditionally examined non-GAAP EPS and that non-GAAP disclosure is sometimes more extensive in the earnings announcement than in the MD&A. Even if a non-GAAP measure has appeared in the earnings announcement, its appearance in the subsequent MD&A is meaningful because the latter is final and placed in close proximity to the audited financial statements.

[4]We also find significant differences in another MD&A textual attribute—the total amount of new information in a firm's MD&A. We measure the amount of new information using the degree of year-over-year MD&A modifications (S. V. Brown & Tucker, 2011). Firms with less adequate financial statements modify their MD&As to a greater extent and therefore provide more new information in the MD&A.

Our study is related to Hribar et al. (2022). Hribar et al. proxy for managers' constraints in GAAP reporting by the frequency of "shall," "should," and "must" in the applicable accounting standards. The authors examine the likelihood of non-GAAP EPS in earnings announcements and the frequency of management earnings forecasts—voluntary disclosure channels that we do not examine. The authors examine the MD&A but limit their examination to readability. Their study and our study are complementary in that the two studies examine different channels yet both conclude that managers use more flexible communication channels when their GAAP channel is constrained.

Our study also contributes to the textual analysis literature. So far, the literature has evolved in three phases. In the first phase, researchers analyzed simple textual attributes, such as length, tone, and readability, of narrative information available to market participants (Henry & Leone, 2016; Li, 2008). In the second phase, techniques were introduced to compare document similarity either from year to year within a firm or between a firm and its benchmark firms (S. V. Brown & Knechel, 2016; S. V. Brown & Tucker, 2011; Hoberg & Phillips, 2010; Peterson et al., 2015). In the third phase, researchers use machine learning approaches to capture more nuanced textual attributes (Campbell et al., 2014; N. C. Brown et al., 2020; Dyer et al., 2017; Frankel et al., 2016; Frankel et al., 2022; Huang et al., 2018; Li, 2010). We are one of the first in accounting and finance to use *deep learning*—an advanced form of machine learning—in textual analysis. Deep learning models are built on neural networks and can discover intricate structures in large data (Bochkay et al., 2023; LeCun et al., 2015). In addition, we introduce TopicTiling as an advanced technique for topic analysis. These tools can help researchers address a wide range of questions in the future.

2 | BACKGROUND AND HYPOTHESIS DEVELOPMENT

2.1 | Financial statement channel

We consider the financial statement channel adequate if financial statements convey to investors useful information about the economic activities of a firm. Most prior research evaluates the adequacy of the financial statement channel using value relevance. This approach implicitly assumes that a firm's stock price reflects the economic value of the firm (a stronger assumption) or investors' consensus beliefs (a weaker assumption) (Barth et al., 2001, pp. 94–95).

Lev (1989, p. 154) points out that, despite the statistical significance of the association between the news of reported earnings—"the premier product of financial disclosure regulation"—and stock returns, the explanatory power of earnings for stock returns is very low. One explanation he offers is that the accounting measurement and valuation principles under GAAP are ineffective. Amir and Lev (1996) report that earnings (which summarizes the income statement) and the book value of equity (which summarizes the balance sheet) of 14 wireless companies have no explanatory power for firms' share prices. The authors argue that the problem is attributable to GAAP requiring firms to expense their investments in R&D and advertising instead of capitalizing those costs. However, the authors find that corporate disclosure of nonfinancial information about growth prospects and market penetration is value relevant.

Lev and Zarowin (1999) find a decline in the value relevance of financial statements over time and argue that GAAP has become ineffective in reflecting the financial impact of business changes caused by innovation and competition. Facing these changes, firms incur costs up front, such as investment in innovation, employee training, restructuring, and product reengineering, but benefits are realized in subsequent periods. GAAP requires firms to expense those intangible costs immediately, resulting in a mismatch of costs and revenues and therefore less value relevant financial statements. Consistent with this idea, Srivastava (2014) concludes that the decline in earnings quality is mostly attributable to the increase of intangible-intensive firms.

Balachandran and Mohanram (2011) also find evidence of a decline in value relevance over time, but find no evidence that the decline is associated with the increased use of conservative accounting practices, such as expensing R&D and advertising and using LIFO for inventory valuation. Recent research reports no evidence of a decline in value relevance after adding accounting variables to the regression and allowing nonlinear relationships (Barth et al., 2023).[5]

Overall, the literature has expressed serious concerns that GAAP might have fallen behind the growth of a knowledge-based economy and that one-size-fits-all accounting rules might have resulted in inadequate financial statements for some firms. We examine how managers use the MD&A when their financial statements are inadequate.

2.2 | MD&A channel

Based on hand-coded small samples, early studies find that the MD&A conveys useful information to investors. Bryan (1997) examines the favorability (e.g., tone) of seven types of required MD&A disclosure and finds that it is positively associated with analysts' and investors' belief revisions of future performance. Cole and Jones (2004) focus on retail firms that provide comparable store sales, planned store openings, and planned capital expenditures, and find that such quantitative information in the MD&A is associated with future performance changes.

Prior studies also document evidence consistent with the flexibility of the MD&A channel. Clarkson et al. (1994) examine voluntary disclosure of forecasts in the MD&A and find that the disclosure decision depends on a firm's external financing needs and the barriers to entry in its product market. Examining firms with substantial inventory changes that are not justified by sales changes, Sun (2010) finds that only half of these firms explain their inventory changes. When explanations are provided, they appear to be credible.

Several studies use textual analysis to examine the MD&A. Li (2008) examines the readability of annual reports, including the MD&A, and concludes that managers structure the documents to hide unfavorable information from investors. This conclusion is confirmed by Lo et al. (2017) but questioned by Gee (2018). Li (2010) and Muslu et al. (2015) examine FLS, with the former focusing on the predictability of tone for future performance and the latter focusing on FLS enhancing the price informativeness of future earnings. Davis and Tama-Sweet (2012) examine the tone of the MD&A and find that it is less optimistic than the tone of the preceding earnings announcement. Feldman et al. (2010) find that the MD&A tone change is indicative of future performance. Bochkay and Levine (2019) report that traditional random-walk earnings prediction models can be improved by adding an MD&A word count matrix.

Recently, researchers have started to conduct topic analyses of the MD&A. Hoberg and Lewis (2017) find that, relative to non-fraudulent managers, fraudulent managers explain their firms' performance less and discuss the positive aspects more, suggesting that managers use their discretion in the MD&A to cover up their misuse of discretion in preparing financial statements. On the other hand, N. C. Brown et al. (2020) find that the systematic topical patterns of misreporting firms reveal, rather than cover, their reporting anomalies, and that these patterns are helpful in detecting misreporting.

[5]Beaver et al. (2018, 2020) document an increase in the information content of earnings *announcements* in the past two decades. Their evidence does not necessarily mean that the value relevance of earnings has increased. Earnings announcements increasingly contain information well beyond realized financials. For example, the majority of management earnings forecasts are provided at earnings announcements (Anilowski et al., 2007); the majority of analyst forecasts are released within a few days after an earnings announcement (Keskek et al., 2014); and conference calls, which occur soon after earnings announcements, have become one of the most important informational events for market participants (L. D. Brown et al., 2019).

2.3 | Hypothesis development

We first develop an overarching hypothesis on the relationship between the financial statement channel and the MD&A channel. We then operationalize this hypothesis into two testable relationships.

Managers can voluntarily disclose information to help capital markets participants with valuation when financial statements do not provide adequate information. For example, Healy and Palepu (1993) argue that financial statements may not reflect the benefits of investments, such as R&D, in a timely fashion and that managers may respond to this problem with voluntary disclosure. There are many types of and channels for voluntary disclosure; our study focuses on managers' discretionary disclosure choices in the MD&A.

To our knowledge, only two theoretical studies model how managers make two different but related reporting and disclosure decisions simultaneously. In Bagnoli and Watts (2007), managers must report two financial numbers and may voluntarily provide information about the precision of the first financial number. Note that all signals are quantitative and the voluntary signal would be irrelevant if the related financial number is not reported. The authors find that managers opportunistically provide the voluntary signal to complement the mandatory signal. Einhorn (2005) examines when a manager provides a voluntary signal to accompany a mandatory signal, both of which are value relevant and quantitative. In Proposition 6, she shows that when the manager has limited discretion in mandatory disclosure, the probability of voluntary disclosure increases. This scenario is similar to our setting, even though the "voluntary" signal in our study is not totally voluntary but is a signal over which managers have substantial discretion. Thus, we expect managers to use the MD&A channel to a greater extent when the financial reporting channel is more constrained.

A cynical view is that financial reporting is a compliance task.[6] Under this view, financial reporting is largely a deadweight cost and not a means to provide the best information to investors (Dichev et al., 2013). Some of the CFOs interviewed in Dichev et al. (2013) view implementation guidance as futile, yet still prefer detailed rules to reduce uncertainty in compliance. One CFO said, "There are so many things that are ridiculous, but rather than saying oh this is ridiculous, we say OK. We just want to get it right" (Dichev et al., 2013, p. 23).[7] In a compliance mindset, managers would put in the minimum effort necessary to prepare the MD&A simply to avoid securities enforcements and then save their resources for other activities, such as strategy, product development, and business growth. In this mindset, managers are unlikely to consider the financial statement channel and the MD&A channel together. Thus, the adequacy of financial statements may be unrelated to the way in which managers prepare the MD&A.

In addition, prior research reports evidence of opportunistic managerial disclosure (Li, 2008; Lo et al., 2017). Opportunistic managers are unlikely to prioritize providing information to investors in the best possible way. On the contrary, they are likely to manipulate financial statements, reducing their value relevance, and use the MD&A to cover up, resulting in uninformative disclosure.

Furthermore, additional disclosure in the MD&A might reveal information that could reduce the firm's competitiveness. Managers with this concern may not provide discretionary disclosure in the MD&A when the financial statement channel is constrained. The extent to which concerns about proprietary disclosure costs mute or eliminate any relation between the use of the two communication channels depends on whether additional disclosures would reveal *actionable* information relevant to the dimension of competition that concerns managers (Cao et al., 2018).

[6]Some companies even refer to their department in charge of external financial reporting as the "financial compliance" division. For example, Monga (2017) wrote, "Health-products company Johnson & Johnson took 6 months last year to fill an open position for a junior-level accountant in its financial-compliance department."

[7]One example of the compliance view is the use of eXtensible Business Reporting Language (XBRL). Since 2009, the SEC has been requiring firms to tag their financial data using XBRL; more than 7,600 corporate annual reports were filed in XBRL in 2014 (Chasan, 2015). Even though the tags are designed to make financial reports easier to compare across firms and therefore provide investors a clearer picture of a firm's financials, managers "still view it as a compliance exercise" without any benefits for them or their firms (Chasan, 2015).

We state our overarching hypothesis in the alternative form as follows:

Hypothesis 1. Firms with less adequate financial statements use the MD&A channel for communication to a larger extent.

2.3.1 | Non-GAAP disclosure

One regulatory objective of the MD&A is "to enhance the overall financial disclosure and provide the context within which financial information should be analyzed" (SEC, 2008). While GAAP measures result from one-size-fits-all accounting rules, non-GAAP measures are customized to each firm's situation. Thus, non-GAAP disclosure provides context for investors to understand GAAP measures.[8] We expect managers to discuss non-GAAP measures more when the financial statement channel is less adequate. In our first testable hypothesis, we specify non-GAAP disclosure as the MD&A attribute and state Hypothesis 1a in the alternative form:

Hypothesis 1a. Firms with less adequate financial statements provide more non-GAAP discussion in the MD&A.

Note that this hypothesis differs from prior non-GAAP research by focusing on the intensity of qualitative non-GAAP discussion rather than the provision of a non-GAAP measure. Moreover, prior non-GAAP research examines non-GAAP EPS disclosure in the earnings announcement. We examine non-GAAP disclosure in the MD&A with a scope beyond non-GAAP EPS.

2.3.2 | Forward-looking statements

Another regulatory objective of the MD&A is "to provide information about the quality of, and potential variability of, a company's earnings and cash flow so that investors can ascertain the likelihood that past performance is indicative of *future* performance" (SEC, 2008, emphasis added). For this reason, the SEC encourages FLS in the MD&A and even provides a safe harbor for such disclosure. Thus, we specify FLS as the MD&A attribute in our second testable hypothesis and state it in the alternative form:

Hypothesis 1b. Firms with less adequate financial statements provide more forward-looking statements in the MD&A.

3 | SAMPLE SELECTION, KEY MEASUREMENTS, AND EMPIRICAL FRAMEWORK

3.1 | Sample selection

Our data analysis requires electronic 10-K reports filed with the SEC available on EDGAR. Because the SEC underwent a major formatting upgrade from plain text to HTML in 2003, we start our sample collection with fiscal year 2003 to reduce extraction errors. We end the

[8] Academic research finds that investors perceive non-GAAP measures in earnings announcements to be more informative than GAAP measures and that non-GAAP measures are highly predictive of future performance, especially for loss firms and firms with low informativeness of GAAP earnings (Bhattacharya et al., 2003; Leung & Veenman, 2018; Lougee & Marquardt, 2004).

sample collection with fiscal year 2016. The initial sample of 157,085 firm-year observations collected from Compustat experiences a few major attritions. We exclude 36,888 very small firms with total assets less than $1 million at the fiscal year-end. We cannot find the 10-K filings on EDGAR for another 48,425 firm-years. For 8,937 observations, we cannot extract valid MD&As largely due to incorporation by reference. We drop a firm's first year from the sample because we use the previous year's MD&A for comparison in untabulated analysis (see Footnote 4), resulting in a loss of 10,404 observations. The data requirement for estimating the value relevance of financial statements results in a loss of 8,425 observations. Our test sample has 36,552 firm-year observations. Table 1 summarizes our sample collection.

3.2 | Key measurements

3.2.1 | Identifying non-GAAP disclosure

We use keyword search in the MD&A to identify non-GAAP disclosure. The set of keywords includes "non-gaap," "non gaap," and "nongaap," tokens beginning with "EBIT" (common matches include "EBIT," "EBITA," "EBITDA," "EBITDAR," "EBITDAS," and "EBITDAX"), "adjusted" + (0, 1, or 2 other words) + "earnings"/"income"/"eps" (e.g., "adjusted net income" and "adjusted basic EPS"), and "free cash flow." In our sample, 40.6% of the firm-years provide non-GAAP disclosure in the MD&A and we code these observations as 1 for *NonGAAP Dummy*. We proxy for the intensity of non-GAAP discussion using the occurrences of the keywords, *NonGAAP Count*. For all sample firm-years, the variable has a mean value of five. For the subsample of firm-years that provide non-GAAP disclosure (that is, *NonGAAP Dummy* equal to 1), the variable has a mean of 12. Appendix 1 summarizes all our variable definitions.

Our study is the first to examine non-GAAP disclosure in the MD&A. As such, we compare non-GAAP disclosure in the MD&A with non-GAAP disclosure in the preceding earnings announcement (EA). Following Bentley et al. (2018), we collected the EA press releases for the fiscal quarters of the sample year from Item 2.02 (Item 12 before August 23, 2004) of 8-K reports. We use the same set of keywords to identify non-GAAP disclosure in the EA as in the

TABLE 1 Sample selection.

	Attrition	Total
US firm-year observations in Compustat for fiscal years 2003–2016		157,085
Drop observations:		
Total assets at year-end are less than $1 million	(36,888)	
There is a change in the fiscal year-end date during the year	(4,543)	
Corresponding 10-K filing is unavailable in EDGAR	(48,425)	
MD&A could not be extracted or is less than 50 sentences	(8,937)	
The six-digit GICS industry has fewer than five firms	(166)	
Previous year's MD&A is unavailable (we lose sample year 2003)	(10,404)	
There is insufficient data for calculating the *Inadequacy* variable	(8,425)	
Book value of equity at year-end is less than or equal to zero	(2,727)	
Control variables are missing	(18)	
Sample firm-years		36,552

Note: The table summarizes our sample selection process. The final sample is 36,552 firm-years from 5,928 unique firms during 2004–2016. *Inadequacy* is our primary explanatory variable. See Appendix 1 for the variable definition.

MD&A. *EA NonGAAP Count* is the occurrences of the keywords in the fourth fiscal quarter EA.[9] Panel A of Figure 1 presents the distribution of *NonGAAP Count* versus *EA NonGAAP Count* for the 31,282 firm-years for which we can locate both the fourth fiscal quarter EA and the fiscal year's MD&A.

We observe a variety of non-GAAP disclosure in the MD&A that are not about EPS, such as aggregate earnings measures (e.g., EBITDA), revenues, expenses, and free cash flows. For example, Sirius XM Holding Inc. stated in its MD&A for fiscal year 2010:

> These Non-GAAP financial measures include: average monthly revenue per subscriber, or ARPU; subscriber acquisition cost, or SAC, per gross subscriber addition; customer service and billing expenses, per average subscriber; free cash flow; adjusted total revenue; and adjusted EBITDA. . . . We use these Non-GAAP financial measures to manage our business, set operational goals and as a basis for determining performance-based compensation for our employees.

To better understand the types and intensity of non-GAAP disclosure in the EA versus the MD&A, we randomly select 200 firm-years from the firm-years that provide non-GAAP disclosure in both the fourth fiscal quarter EA and the MD&A. After examining a pilot sample, we create 10 types of non-GAAP measures and code the 200 firms. Panel B of Figure 1 presents the frequency of non-GAAP measures belonging to a given type in the EA versus the MD&A.[10] Even though non-GAAP earnings on a per share or aggregate basis are most common, we observe a large number of other types of non-GAAP measures. Almost every type of non-GAAP measure is slightly more prevalent in the EA than in the MD&A. A typical EA contains three types of non-GAAP measures and 37.5% of the firms provide four to seven types. A typical MD&A contains two types of non-GAAP measures and 38% of the firms provide three to six types (untabulated).

3.2.2 | Identifying forward-looking statements

We use deep learning, an advanced machine learning approach, to identify forward-looking disclosure. A machine learning model uses a training sample to learn the mapping from an input to an output (e.g., tone) and then applies the mapping to unseen data. Traditional machine learning can model a large variety of data relations but has difficulties as the relations become complex. Deep learning can replicate the basic capabilities of naturally occurring neural networks, allow fewer interventions by the researcher, and capture complex relations. Deep learning models are on the cutting edge of textual analysis (Bochkay et al., 2023).

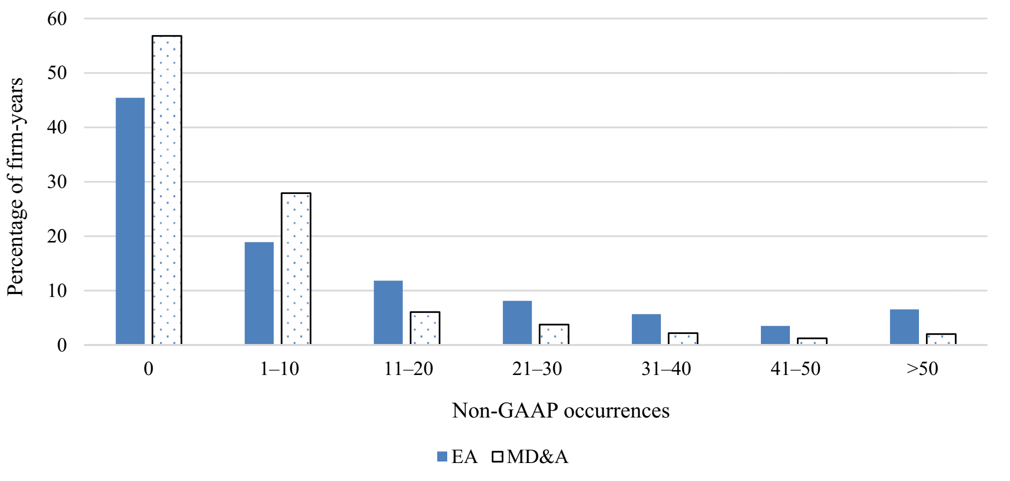
We implement deep learning in Spacy, a natural language processing library with accompanying pre-trained models.[11] Before training the machine to recognize FLS, we manually classify (label) each sentence in the training data as (1) an FLS about the company (e.g., "capital expenditures will rise to at least $10 million"), (2) a future-oriented statement that can be made by many companies (e.g., "actual results could differ from those estimates"), or (3) not future-oriented. We distinguish these three types of statements to isolate the first type, which is the

[9] We alternatively construct an indicator variable, which equals one if a firm provides non-GAAP disclosure in the EA for any of the fiscal quarters of the current year, and zero otherwise. Among the 31,282 firm-years that have both EAs and MD&As, 34.9% provide non-GAAP disclosure in both the EAs and the MD&A, 29.4% do so in the EAs but not in the MD&A, 8.3% do so in the MD&A but not in the EAs, and 27.4% do not provide non-GAAP disclosure in either.

[10] There could be more than one measure in each type. For example, both non-GAAP basic EPS and non-GAAP diluted EPS belong to the "non-GAAP EPS" type.

[11] The word "pre-trained" means that Spacy has already trained the model on documents obtained from the web. Thus, users do not need to reinvent the wheel but only need to fine-tune the model for their own settings. Moreover, the model can learn from each batch of the user's training sample and continue to improve performance.

(A) The intensity of non-GAAP disclosure in the EA versus the MD&A



(B) Types of non-GAAP measures in the EA versus the MD&A for our hand-coded sample

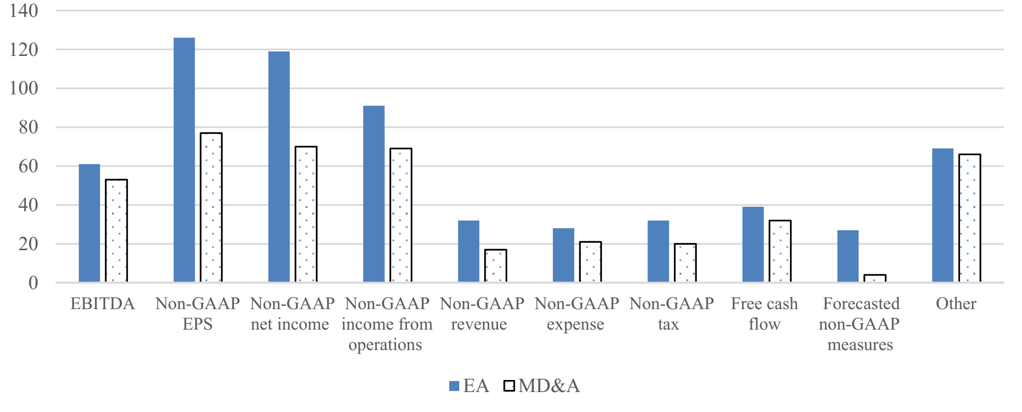


FIGURE 1 Non-GAAP disclosure. We collected the fourth fiscal quarter earnings announcement press releases (EAs) for 31,282 sample firm-years from 8-K reports. We use the same set of keywords to identify non-GAAP disclosure in the EAs as in the MD&As. (A) presents the occurrences of keywords for non-GAAP disclosure on the x-axis and the percentage of firm-years on the y-axis for the EA versus the MD&A. For (B), we randomly selected 200 firm-years from the firm-years that provide non-GAAP disclosure in both the EA and the MD&A and then hand-coded their non-GAAP measures into 10 types: EBITDA, non-GAAP EPS, non-GAAP net income, non-GAAP income from operations, non-GAAP revenue, non-GAAP expense, non-GAAP tax, free cash flow, forecasted non-GAAP, and other. (B) presents the number of firm-years with a given type of non-GAAP measure in the EA versus the MD&A.

type we are interested in.[12] We load the training data (which includes the training sample and the validation sample) into Prodigy, a web-based interface that works with Spacy and improves the efficiency of labeling. The deep learning model establishes the mapping between the textual input of the training sample and the provided labels and then predicts the classification of unseen observations in the validation sample. Spacy reports model performance statistics,

[12]The SEC appears to have a broader definition of "forward-looking" content in the MD&A and includes any content that is not historic, for example, discussions of liquidity, capital expenditures, commitments, known trends, and risk factors (see Regulation S-K 303(c)). Our definition is narrower, and we do not count risk disclosure and cautionary language as FLS.

and we repeat the training process until the model performance statistics reach desired levels. After that, we let the machine identify FLS in our full sample.[13] Appendix 2 provides technical details.

FLS is the number of forward-looking sentences in the firm's MD&A for fiscal year t and captures the intensity of forward-looking disclosure. *FLS* has a mean (median) value of 29.854 (27) for our sample (see Panel A of Table 2). We are also interested in the number of forward-looking sentences that contain hard information. "Hard information" means verifiable information and is proxied by quantitative information on the grounds that investors can verify it with realizations in future periods (Dyer et al., 2017). We use the named entity recognition (NER) technique in Spacy to identify quantitative information (see Appendix 2). *Quantitative FLS* is the number of forward-looking sentences in the firm's MD&A for fiscal year t that contain quantitative information. The variable has a mean (median) value of 12.434 (10) for our sample.

3.2.3 | Measuring financial statement adequacy

We use value relevance as our primary proxy for financial statement adequacy and measure it as the R^2 of the time-series regression in Equation (1) using a firm's 20 past quarters:

$$P_q = a_0 + a_1\, EPS_q + a_2\, BPS_q + e_q, \tag{1}$$

where P_q is the firm's stock price at the end of fiscal period q, EPS_q is its diluted EPS reported for period q, and BPS_q is its book value of equity divided by the number of common shares outstanding at the end of period q. Ohlson (1995) demonstrates that the value of a firm's stock can be expressed as a linear combination of its earnings and book value of equity if (1) value is determined by the discounted future cash flows, (2) the clean surplus relation holds, and (3) future abnormal earnings follow a linear stochastic process. With the additional assumption that the observed stock price is a fair representation of the firm's value, researchers replace value with stock price and assess the degree to which accounting earnings and book value of equity track a firm's economic value.

Following Banker et al. (2009), we use Equation (1) on a firm's time series of data and construct a measure of value relevance for each firm-year. We deviate from Banker et al. (2009) in two respects. Banker et al. use a firm's most recent 10 years' annual report data and require a minimum of 8 years. Given the growth and changes in today's economy, many firms do not exist as public firms for a long time before being acquired or delisted. To mitigate the survivorship concern, we use 5 years of data. To increase the number of observations for the time-series regression estimation, we obtain data at the end of each fiscal quarter in the past 20 quarters and require a minimum of 16 quarterly observations. Banker et al. measure stock price 3 months after the fiscal year-end. We follow Francis and Schipper (1999) and Lev and Zarowin (1999) and measure stock price at the end of the fiscal period.

We refer to one minus the value relevance measure as *Inadequacy*, which has a mean (median) of 0.527 (0.525) for our sample. To facilitate interpretation, in empirical analyses we use the decile-ranked form of *Inadequacy*. The highest decile, valued at 1, is the group with the least adequate financial statements; the lowest decile, valued at 0, is the group with the most adequate financial statements.[14]

[13]We observe a variety of FLS and these statements are much broader than traditional management earnings forecasts. For example, "DuPont and Unifi will continue to own and operate their respective sites and employees will remain with their respective employers" and "As a result, we expect the term loan and revolving credit facility to be funded and the construction bridge loan to be repaid during the first quarter of fiscal 2004."

[14]Firms are sorted into deciles ranging from 1 to 10. The highest decile is assigned the value of 1 (i.e., (10 − 1)/9 = 1) and the lowest decile is assigned the value of 0 (i.e., (1 − 1)/9 = 0).

TABLE 2 Descriptive statistics.

Panel A: Summary statistics

	N	Mean	SD	25th	Median	75th
Explanatory variables						
Inadequacy (*raw*)	36,552	0.527	0.266	0.302	0.525	0.755
Business Change	36,552	0.536	0.499	0	1	1
Loss Firm	36,552	0.307	0.461	0	0	1
MD&A variables						
Descriptive						
MDA Sentences	36,552	362	171	244	333	445
MDA Words	36,552	10,187	5,790	6,472	9,129	12,493
MDA Numbers	36,552	572	481	275	438	702
MDA Words to Numbers	36,552	23	11	16	21	27
MDA Tables	36,552	10	9	4	7	13
MDA Text Segments	36,552	90	48	59	82	110
NonGAAP Dummy	36,552	0.406	0.491	0	0	1
Content						
NonGAAP Count	36,552	4.898	11.385	0	0	3
FLS	36,552	29.854	16.794	18	27	39
Quantitative FLS	36,552	12.434	9.112	6	10	17
Required Topics	36,552	64.863	50.860	29	51	84
Traditional Topics	36,552	172.218	134.667	85	134	209
Intangibles Topics	36,552	37.513	58.081	2	15	46
Firm characteristics						
Total Assets	36,552	4,029	10,740	141	694	2,679
MB	36,552	3.222	4.699	1.151	1.887	3.326
ROA	36,552	−0.024	0.212	−0.018	0.021	0.065
StdRet (%)	36,552	12.434	8.296	6.905	10.223	15.296
Analysts	36,552	10.028	11.264	1	7	15
M&A	36,552	0.191	0.393	0	0	0
Company Age	36,552	21.551	13.995	11	17	27
Segments	36,552	4.390	3.843	1	3	6
Compustat Items	36,552	288.989	37.271	280	297	312
SGA Intensity	36,552	0.280	0.237	0.088	0.238	0.421
R&D Intensity	36,552	0.062	0.139	0	0	0.054
PP&E	36,552	0.195	0.231	0.024	0.100	0.278

Panel B: Mean values for firm-years in the highest versus lowest *Inadequacy* deciles

	Highest *Inadequacy* decile	Lowest *Inadequacy* decile
Explanatory variables		
Inadequacy (*raw*)	**0.948**	**0.106**
Business Change	**0.533**	**0.471**
Loss Firm	**0.318**	**0.218**

TABLE 2 (Continued)

Panel B: Mean values for firm-years in the highest versus lowest *Inadequacy* deciles

	Highest *Inadequacy* decile	Lowest *Inadequacy* decile
MD&A variables		
Descriptive		
MDA Sentences	**352**	**370**
MDA Words	**9,924**	**10,608**
MDA Numbers	**530**	**641**
MDA Words to Numbers	**23**	**22**
MDA Tables	**9**	**11**
MDA Text Segments	**88**	**94**
NonGAAP Dummy	**0.410**	**0.381**
Content		
NonGAAP Count	**4.894**	**4.211**
FLS	**29.787**	**28.567**
Quantitative FLS	**12.255**	**11.842**
Firm characteristics		
Total Assets	**3,702**	**5,341**
MB	3.302	3.424
ROA	**−0.029**	**0.004**
StdRet (%)	**12.028**	**11.033**
Analysts	**9.668**	**10.195**
M&A	**0.180**	**0.224**
Company Age	**21.124**	**22.304**
Segments	**4.456**	**4.264**
Compustat Items	**290.310**	**282.790**
SGA Intensity	**0.289**	**0.269**
R&D Intensity	**0.068**	**0.044**
PP&E	**0.214**	**0.167**

Note: Panel A presents the descriptive statistics for the variables used in our analyses. See variable definitions in Appendix 1. We provide *SGA Intensity* (the selling, general, and administrative expenses for fiscal year *t* divided by total expenses for fiscal year *t*), *R&D Intensity* (the research and development expenses for fiscal year *t* divided by total expenses for fiscal year *t*), and *PP&E* (net property, plant, and equipment at the end of fiscal year *t* divided by total assets at the end of fiscal year *t*) for descriptive purposes. *NonGAAP Count*, *FLS*, *Quantitative FLS*, *Total Assets*, *MB*, *StdRet*, *Analysts*, *Segments*, *Compustat Items*, *SGA Intensity*, *PP&E*, *Required Topics*, *Traditional Topics*, and *Intangibles Topics* are winsorized at 99%. *R&D Intensity* and *ROA* are winsorized at 1% and 99%. Panel B presents the mean values for the 3,655 firm-years in the highest decile of *Inadequacy* and the 3,655 firm-years in the lowest decile of *Inadequacy*. The means that are significantly different at the 5% level between the subgroups are in boldface.

3.3 | Empirical framework

To test Hypothesis 1a, we examine the associations of a firm's *Inadequacy* with *NonGAAP Count* for fiscal year *t*. We estimate Equation (2):

$$\text{Log}\,(1 + NonGAAP\ Count) = b_0 + b_1\ Inadequacy + \sum \gamma_m Control_m + fixed\ effects + \xi. \tag{2}$$

The set of control variables should be confounders, which affect the dependent variable directly as well as indirectly through the explanatory variable (Gow et al., 2016, pp. 483–484).

We heed the advice of Whited et al. (2022) and include only confounders. We control for firm size, growth (*MB*), GAAP performance (*ROA*), and return volatility (*StdRet*). Early non-GAAP studies report several firm characteristics related to the provision of non-GAAP EPS. Large firms, firms in growing industries such as service and technology, firms with poor GAAP performance, and volatile firms are more likely to provide non-GAAP EPS (Bhattacharya et al., 2003; Black & Christensen, 2009). These firm characteristics are also related to financial statement adequacy (Collins et al., 1997; Srivastava, 2014).

To test Hypothesis 1b, we examine the associations of *Inadequacy* with *FLS* and alternatively with *Quantitative FLS* in Equations (3) and (4) and consider (3) our primary model:

$$FLS = c_0 + c_1\, Inadequacy + \sum \rho_n Control_n + fixed\ effects + \mu. \tag{3}$$

$$Quantitative\ FLS = d_0 + d_1\, Inadequacy + \sum \pi_n Control_n + fixed\ effects + \nu. \tag{4}$$

We follow Merkley (2014) and use the number of sentences instead of the percentage of sentences that are forward looking. Our inferences are robust to additionally controlling for the total number of sentences in the MD&A (untabulated). To identify confounders, we start with the determinant model of Muslu et al. (2015). They find that managers tend to provide forward-looking disclosure when the firm is large and growing, the performance is poor, the business environment is uncertain, and analysts demand information. On the other hand, firms engaging in mergers and acquisitions and complex firms refrain from forward-looking disclosure. Of these determinants, size, growth, performance, and return volatility are also related to financial statement adequacy. Thus, we control for firm size, growth, performance, and return volatility.

For transparency to readers, we also present an expanded model by adding variables that exhibit significant explanatory power in Muslu et al. (2015) but are uncorrelated with financial statement adequacy according to prior research. The variables are analyst following (*Analysts*), the indicator of mergers and acquisitions (*M&A*), company age (*Company Age*), operating complexity (*Segments*), and reporting complexity (*Compustat Items*).

In all the above models, we add year fixed effects to control for shocks to the economy (e.g., the financial crisis) that affect both financial statement adequacy and MD&A disclosures. We are primarily interested in the relation between the financial statement channel and the MD&A channel across firms regardless of whether these firms are from the same industry. So, our primary models do not include industry fixed effects. In supplementary models, we add industry (six-digit GICS) fixed effects to examine the relation across firms from the same industry.

4 | TEST RESULTS

4.1 | Descriptive statistics

Panel A of Table 2 presents the descriptive statistics of our sample. The average MD&A in our sample has 362 sentences, 10,187 words, 572 numbers, 23 times as many words as numbers, and 10 tables. Panel B contrasts the highest decile of *Inadequacy* with the lowest decile. The former is smaller in size and younger and has lower profitability and greater return volatility as well as higher operating complexity (*Segments*) and reporting complexity (*Compustat Items*) than the latter. Even though the former has shorter MD&As, it discusses non-GAAP measures more and provides more FLS, including more quantitative FLS, than the latter.

Panels A and B of Table 3 provide pairwise correlations of the variables used in our non-GAAP and forward-looking disclosure regression analyses. The correlations are consistent with

TABLE 3 Pairwise correlations.

Panel A: Pairwise correlations for non-GAAP disclosure analysis

	Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	Log (1 + *NonGAAP Count*)		0.009	**0.158**	**−0.077**	**0.291**	0.005	**0.120**	**−0.105**
(2)	*Inadequacy*	0.008		**0.043**	**0.070**	**−0.074**	−0.006	**−0.053**	**0.035**
(3)	*Business Change*	**0.155**	**0.043**		**0.136**	**0.084**	**0.012**	**−0.051**	**0.074**
(4)	*Loss Firm*	**−0.092**	**0.070**	**0.136**		**−0.370**	**0.072**	**−0.614**	**−0.408**
(5)	*Size*	**0.304**	**−0.076**	**0.078**	**−0.366**		**−0.139**	**0.392**	**−0.408**
(6)	*MB*	**0.015**	**−0.060**	**0.054**	**−0.088**	**−0.026**		**−0.166**	**0.073**
(7)	*ROA*	**0.078**	**−0.100**	**−0.058**	**−0.799**	**0.250**	**0.268**		**−0.367**
(8)	*StdRet*	**−0.114**	**0.072**	**0.112**	**0.439**	**−0.448**	**−0.053**	**−0.325**	

Panel B: Pairwise correlations for forward-looking disclosure analysis

	Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1)	*FLS*		**0.839**	**0.011**	**0.113**	**−0.032**	**0.392**	**−0.020**	**0.058**	**−0.056**	**0.268**	**0.056**	**0.064**	**0.158**	**0.060**
(2)	*Quantitative FLS*	**0.823**		0.002	**0.091**	**−0.102**	**0.412**	**−0.049**	**0.135**	**−0.098**	**0.219**	**0.057**	**0.136**	**0.171**	**0.026**
(3)	*Inadequacy*	0.007	0.000		**0.043**	**0.070**	**−0.074**	−0.006	**−0.053**	**0.035**	**−0.015**	**−0.027**	**−0.025**	**0.019**	**0.066**
(4)	*Business Change*	**0.125**	**0.108**	**0.043**		**0.136**	**0.084**	**0.012**	**−0.051**	**0.074**	**0.148**	**0.196**	**0.104**	**0.114**	**0.414**
(5)	*Loss Firm*	**−0.040**	**−0.119**	**0.070**	**0.136**		**−0.370**	**0.072**	**−0.614**	**0.414**	**−0.143**	**−0.082**	**−0.171**	**−0.115**	**0.101**
(6)	*Size*	**0.391**	**0.428**	**−0.076**	**0.078**	**−0.366**		**−0.139**	**0.392**	**−0.408**	**0.567**	**0.098**	**0.265**	**0.194**	**−0.064**
(7)	*MB*	**0.031**	−0.006	**−0.060**	**0.054**	**−0.088**	**−0.026**		**−0.166**	**0.073**	**0.089**	−0.003	**−0.020**	**−0.037**	**0.110**
(8)	*ROA*	**0.015**	**0.091**	**−0.100**	**−0.058**	**−0.799**	**0.250**	**0.268**		**−0.367**	**0.150**	**0.089**	**0.173**	**0.137**	**−0.031**
(9)	*StdRet*	**−0.054**	**−0.112**	**0.072**	**0.112**	**0.439**	**−0.448**	**−0.053**	**−0.325**		**−0.172**	**−0.087**	**−0.171**	**−0.093**	**0.120**
(10)	*Analysts*	**0.307**	**0.258**	**−0.024**	**0.158**	**−0.150**	**0.562**	**0.243**	**0.192**	**−0.158**		**0.112**	**0.107**	**0.086**	**0.221**
(11)	*M&A*	**0.068**	**0.070**	**−0.027**	**0.196**	**−0.082**	**0.096**	**0.099**	**0.116**	**−0.058**	**0.132**		**0.024**	**0.131**	**0.218**
(12)	*Company Age*	−0.005	**0.090**	**−0.025**	**0.090**	**−0.174**	**0.228**	**0.023**	**0.217**	**−0.188**	**0.054**	**0.020**		**0.236**	**0.209**
(13)	*Segments*	**0.126**	**0.148**	**0.030**	**0.140**	**−0.084**	**0.069**	**0.066**	**0.185**	**−0.020**	**0.079**	**0.148**	**0.192**		**0.282**
(14)	*Compustat Items*	**0.077**	**0.046**	**0.039**	**0.441**	**0.053**	**0.038**	**0.184**	**0.113**	**0.125**	**0.259**	**0.220**	**0.213**	**0.285**	

Note: This table presents correlation coefficients for the variables used in our analyses. Pearson correlations are above the diagonal and Spearman correlations are below the diagonal. The correlations that are significantly different from zero at the 5% level are in boldface. See variable definitions in Appendix 1.

our expectations except for the insignificant correlation between *Inadequacy* and Log (1 + *NonGAAP Count*).

4.2 | Financial statement adequacy and non-GAAP disclosure

We report the OLS estimation of Equation (2) in Table 4. In Column 1, the coefficient on *Inadequacy* is significantly positive at 0.143 with a *t*-statistic of 5.04, suggesting that the intensity of non-GAAP disclosure in the MD&A increases significantly from firms in the most adequate group to firms in the least adequate group. Because the dependent variable is a logarithm, the economic effect is a 19.0% proportional increase in non-GAAP intensity.[15]

TABLE 4 Financial statement adequacy and non-GAAP disclosure in the MD&A.

Models:
$\text{Log}\,(1 + NonGAAP\ Count) = b_0 + b_1\,Inadequacy + \sum \rho_n Control_n + fixed\ effects + \xi$.

	(1)	(2)	(3)	(4)
Intercept	−0.720***	−0.826***	−0.604***	−0.544***
	(−12.89)	(−5.83)	(−9.58)	(−3.93)
Inadequacy	0.143***	0.110***	0.093***	0.089***
	(5.04)	(4.06)	(3.09)	(3.10)
Log (1 + *EA NonGAAP Count*)			0.231***	0.220***
			(23.30)	(21.95)
Size	0.159***	0.180***	0.116***	0.113***
	(21.33)	(22.97)	(14.09)	(12.70)
MB	0.010***	0.005**	0.006**	0.004*
	(4.25)	(2.35)	(2.12)	(1.66)
ROA	0.191***	−0.145***	0.130***	−0.220***
	(5.07)	(−3.64)	(2.99)	(−4.55)
StdRet	0.005***	0.003***	0.004***	0.004***
	(4.48)	(2.91)	(3.13)	(2.88)
Year FE	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes
Adjusted R^2	13.5%	20.5%	20.9%	26.5%
N	36,552	36,552	31,282	31,282

Note: The table reports the OLS estimations with Log (1 + *NonGAAP Count*) as the dependent variable. *NonGAAP Count* is the occurrences of case-insensitive keywords for non-GAAP disclosure in fiscal year *t*'s MD&A. *EA NonGAAP Count* is the occurrences of case-insensitive keywords for non-GAAP disclosure in the earnings announcement for fiscal year *t*'s fourth fiscal quarter. The set of keywords includes "non-gaap," "non gaap," and "nongaap," tokens beginning with "EBIT" (common matches include "EBIT," "EBITA," "EBITDA," "EBITDAR," "EBITDAS," and "EBITDAX"), "adjusted" + (0, 1, or 2 other words) + "earnings"/"income"/"eps" (e.g., "adjusted net income" and "adjusted basic EPS"), and "free cash flow." See other variable definitions in Appendix 1. The decile-ranked variable of *Inadequacy* is used in the regression analysis. *MB* and *StdRet* are winsorized at 99%. *ROA* is winsorized at 1% and 99%. The estimations are robust to heteroskedasticity and within-firm error correlations. We present the estimated coefficients with *t*-statistics in parentheses.
***, **, and * represent statistical significance in two-tailed tests at the 1%, 5%, and 10% levels, respectively.

[15] Let us denote the non-GAAP intensity of firms in the lowest decile of *Inadequacy* as Y_0 and that of firms in the highest decile as Y_1. The coefficient of 0.143 means that $\text{Ln}\,(Y_1 + 1) - \text{Ln}\,(Y_0 + 1) = 0.143 \times (1\text{–}0) = 0.143$. Thus, $(Y_1 + 1)/(Y_0 + 1) = \exp(0.143)$. After some algebra, $(Y_1 - Y_0)/Y_0 = [\exp(0.143) - 1] \times [(Y_0 + 1)/Y_0]$. For the average firm in the lowest decile, $Y_0 = 4.211$ (see Panel B of Table 2). Thus, the proportional increase, $(Y_1 - Y_0)/Y_0$, is 19.0%.

If we additionally control for industry fixed effects in Column 2, the results are similar. In Columns 3 and 4, we control for *EA NonGAAP Count* and the results are also similar.[16]

The test results in Table 4 suggest that firms with less adequate financial statements use the MD&A to provide more discussion of customized financial information. This finding is consistent with Hypothesis 1a.

4.3 | Financial statement adequacy and forward-looking statements

Panel A of Table 5 presents the OLS estimation results of Equations (3) and (4). We present *FLS* in the left columns and *Quantitative FLS* in the right columns, each for the primary model and then the expanded model. In Column 1, the coefficient on *Inadequacy* is significantly positive at 1.508 with a *t*-statistic of 4.19, suggesting that firms with less adequate financial statements provide significantly more forward-looking information in the MD&A. The increase of 1.508 forward-looking sentences from firms in the most adequate group to firms in the least adequate group represents 5.1% of the unconditional mean.

In Column 5, the coefficient on *Inadequacy* is significantly positive at 0.613 with a *t*-statistic of 3.18, suggesting that firms with less adequate financial statements provide significantly more quantitative and therefore verifiable forward-looking information. The increase of 0.613 quantitative forward-looking sentences from firms in the most adequate group to firms in the least adequate group represents 4.9% of the unconditional mean.

The above results are robust to controlling for industry fixed effects and to estimating the expanded models. In Panel B, we replace *FLS* with *FOS*, a variable that measures the number of future-oriented sentences in the MD&A that are not FLS—the second type of sentences described in Section 3.2.2. Because these statements can be made by almost any company, we do not expect them to help investors predict the sample firm's future performance and therefore do not expect *FOS* to be associated with financial statement adequacy. Indeed, we do not observe any statistically significant association between *Inadequacy* and *FOS* in Panel B. This placebo analysis bolsters our confidence in our primary results. The test results in Table 5 suggest that firms with less adequate financial statements use the MD&A to provide more forward-looking information to investors. This finding is consistent with Hypothesis 1b.

5 | ADDITIONAL ANALYSES

5.1 | Context of non-GAAP and forward-looking disclosures

After documenting increased non-GAAP and forward-looking disclosures by firms whose financial statement channel is inadequate, we investigate the context of the disclosures. By "context" we mean where in the MD&A and in which topics such disclosures appear.

We use TopicTiling to identify latent topics in the MD&A and their locations in the document. TopicTiling is a text segmentation technique introduced in computational linguistics that combines the topic model of latent Dirichlet allocation (LDA) with the text segmentation model of TextTiling (Riedl & Biemann, 2012a, 2012b). In LDA, a topic is a vector of weights, where each weight corresponds to one unique word. LDA discovers latent topics in a corpus, assuming that managers choose a topic mix for a document and then choose a word mix for each topic (Blei et al., 2003). Our model produces a document-topic matrix and a topic-word matrix. We use these matrices to determine the probability distribution of

[16]In untabulated analysis, we control for the indicator variable of whether the firm has provided non-GAAP disclosure in any of the earnings announcements for fiscal year *t*'s fiscal quarters and the results are similar.

TABLE 5 Financial statement adequacy and forward-looking disclosure in the MD&A.

Models:

$FLS = c_0 + c_1\, Inadequacy + \sum \rho_n Control_n + fixed\ effects + \mu,$

$Quantitative\ FLS = d_0 + d_1\, Inadequacy + \sum \pi_n Control_n + fixed\ effects + \nu.$

Panel A: Using forward-looking statements and quantitative forward-looking statements

	FLS				*Quantitative FLS*			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Intercept	7.356***	7.976***	−2.099	−6.434*	−0.411	0.740	−3.989***	−6.185***
	(9.14)	(4.71)	(−1.16)	(−1.96)	(−1.02)	(0.82)	(−4.13)	(−3.64)
Inadequacy	1.508***	0.991***	1.183***	0.891**	0.613***	0.409**	0.571***	0.399**
	(4.19)	(2.84)	(3.29)	(2.56)	(3.18)	(2.21)	(2.99)	(2.17)
Size	3.787***	3.871***	3.733***	3.639***	1.959***	1.914***	2.029***	1.860***
	(37.26)	(36.57)	(29.80)	(25.96)	(36.70)	(34.42)	(30.41)	(25.11)
MB	0.123***	0.036	0.084***	0.037	0.039***	0.020	0.043***	0.026*
	(4.57)	(1.40)	(3.19)	(1.45)	(2.65)	(1.42)	(2.97)	(1.87)
ROA	−7.807***	−6.641***	−7.935***	−6.160***	−1.236***	−2.083***	−1.629***	−2.013***
	(−13.14)	(−10.77)	(−13.28)	(−9.94)	(−4.38)	(−7.05)	(−5.72)	(−6.74)
StdRet	0.167***	0.148***	0.146***	0.144***	0.054***	0.045***	0.057***	0.049***
	(10.09)	(9.20)	(9.03)	(9.05)	(6.30)	(5.50)	(6.69)	(5.98)
Analysts			0.020	0.014			−0.038***	−0.021*
			(0.86)	(0.58)			(−3.05)	(−1.65)
M&A			−0.073	0.134			0.207	0.370**
			(−0.24)	(0.47)			(1.30)	(2.43)
Company Age			−0.039**	−0.051***			0.028***	0.013
			(−2.36)	(−3.16)			(3.18)	(1.45)
Segments			0.161***	0.159***			0.067**	0.059**
			(2.89)	(2.90)			(2.20)	(1.98)
Compustat Items			0.037***	0.057***			0.010***	0.025***
			(6.22)	(5.21)			(2.98)	(4.47)

TABLE 5 (Continued)

Panel A: Using forward-looking statements and quantitative forward-looking statements

	FLS				*Quantitative FLS*			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes	No	Yes	No	Yes
Adjusted R^2	22.1%	25.9%	23.0%	26.4%	23.6%	28.4%	24.3%	28.8%
N	36,552	36,552	36,552	36,552	36,552	36,552	36,552	36,552

Panel B: Using future-oriented statements that are not forward-looking statements (a placebo test)

Dependent variable	*FOS*			
	(1)	**(2)**	**(3)**	**(4)**
Intercept	13.874***	16.692***	7.087***	12.343***
	(16.17)	(12.63)	(5.01)	(5.06)
Inadequacy	0.120	−0.003	−0.236	−0.080
	(0.44)	(−0.01)	(−0.88)	(−0.33)
Size	2.442***	2.441***	2.282***	2.285***
	(26.23)	(31.27)	(20.16)	(21.00)
MB	−0.040**	−0.051***	−0.095***	−0.061***
	(−2.11)	(−2.98)	(−4.72)	(−3.55)
ROA	−5.299***	−4.266***	−4.841***	−3.880***
	(−11.10)	(−9.11)	(−10.04)	(−8.24)
StdRet	0.139***	0.124***	0.108***	0.115***
	(10.43)	(10.01)	(8.24)	(9.32)
Analysts			0.081***	0.049***
			(4.98)	(2.94)
M&A			−0.847***	−0.483**
			(−3.51)	(−2.28)

(Continues)

TABLE 5 (Continued)

Panel B: Using future-oriented statements that are not forward-looking statements (a placebo test)

Dependent variable	***FOS***			
	(1)	**(2)**	**(3)**	**(4)**
Company Age			−0.088***	−0.064***
			(−7.61)	(−5.98)
Segments			0.088**	0.020
			(2.05)	(0.51)
Compustat Items			0.033***	0.022***
			(8.50)	(2.63)
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	No	Yes	No	Yes
Adjusted R^2	19.1%	28.3%	21.0%	28.8%
N	36,552	36,552	36,552	36,552

Note: Panel A reports the OLS estimations and uses the number of forward-looking statements (*FLS*) and the number of quantitative forward-looking statements (*Quantitative FLS*) as the dependent variables. Panel B is a placebo test using the number of future-oriented statements (*FOS*) that are not forward-looking statements as the dependent variable. Such statements are coded as *FLS (not specific)* in Appendix 2. Examples of *FOS* are cautionary language and risk disclosures. *FOS* has a mean (median) of 18 (15) for our full sample. See variable definitions in Appendix 1. The decile-ranked variable of *Inadequacy* is used in the regression analysis. *FLS*, *Quantitative FLS*, *FOS*, *MB*, *StdRet*, *Analysts*, *Segments*, and *Compustat Items* are winsorized at 99%. *ROA* is winsorized at 1% and 99%. The estimations are robust to heteroskedasticity and within-firm error correlations. We present the estimated coefficients with *t*-statistics in parentheses.

***, **, and * represent statistical significance in two-tailed tests at the 1%, 5%, and 10% levels, respectively.

topics in an MD&A and the probabilistic topic assignments for each word in that document.

A text segmentation model splits a document into text segments, each of which contains *topically similar* consecutive sentences. Our model groups consecutive sentences into a text segment based on topic similarity of the sentences and sustained topic shifts from multiple preceding and subsequent sentences, identifies the location of each text segment in the document, and assigns one topic to each text segment. See technical details in Supporting Information Appendix A. To increase estimation stability and interpretation, we group related topics into 21 topic categories. Figure 2 provides graphical presentations of Idera Pharmaceuticals and Home Depot. The two companies differ substantially in the topic categories, space allocated to a topic category, and location of discussion. For example, Idera allocates a substantial proportion of its MD&A to research and development, whereas Home Depot allocates a substantial proportion to liquidity and capital resources and results of operations.

(A) Example 1

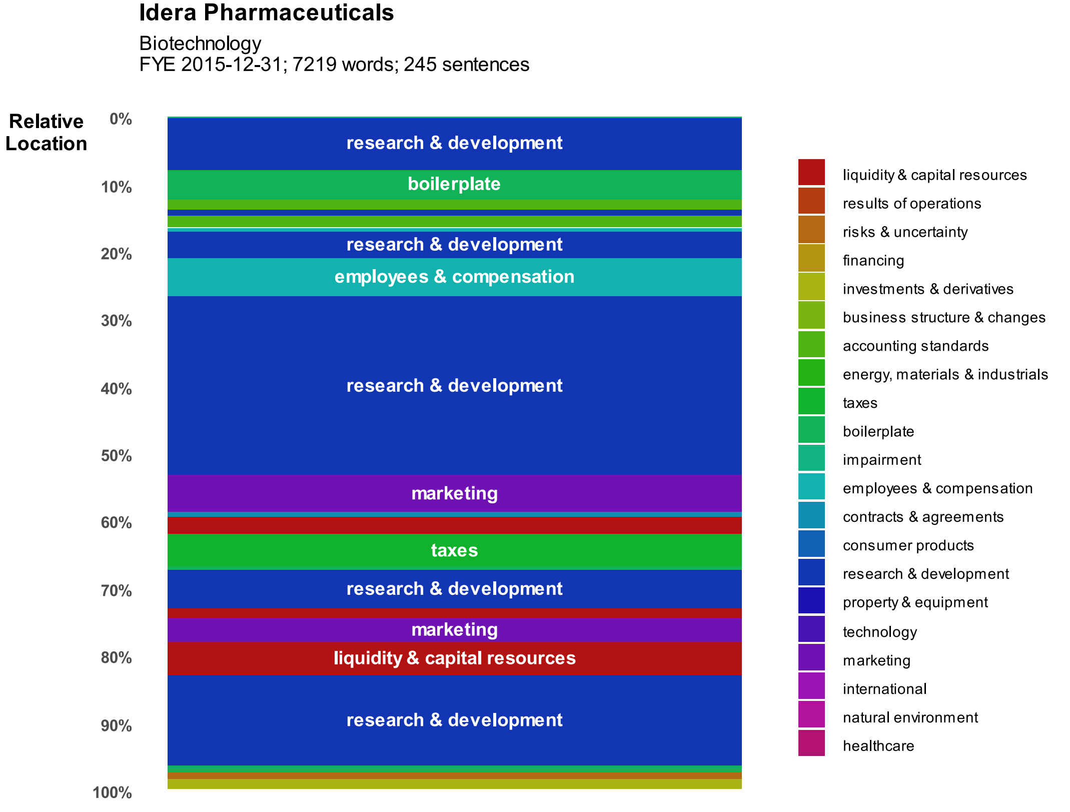


FIGURE 2 Graphical presentations of the MD&A after TopicTiling. TopicTiling splits an MD&A document into text segments, which are classified into the 21 topic categories presented in the legends on the right side. See details in Supporting Information Appendices A and B. We use a color to represent a topic category. The legends on the right-hand side of each graph indicate the color scheme of the topic categories in the decreasing order of document frequencies for that category. The colored area represents the proportion of MD&A space allocated to that topic category. The relative location of the colored area represents the location of discussion in the MD&A, where the beginning and end of the document are marked by 0% and 100%, respectively. We add category names in the graphs for the relatively large colored areas. The graphs for Idera and Home Depot are quite different, suggesting that the two companies differ in the content, amount of space, and location of discussion in the MD&A. For example, Idera allocates a substantial proportion of the MD&A to research & development, whereas Home Depot allocates a substantial proportion to liquidity & capital resources.

(B) Example 2

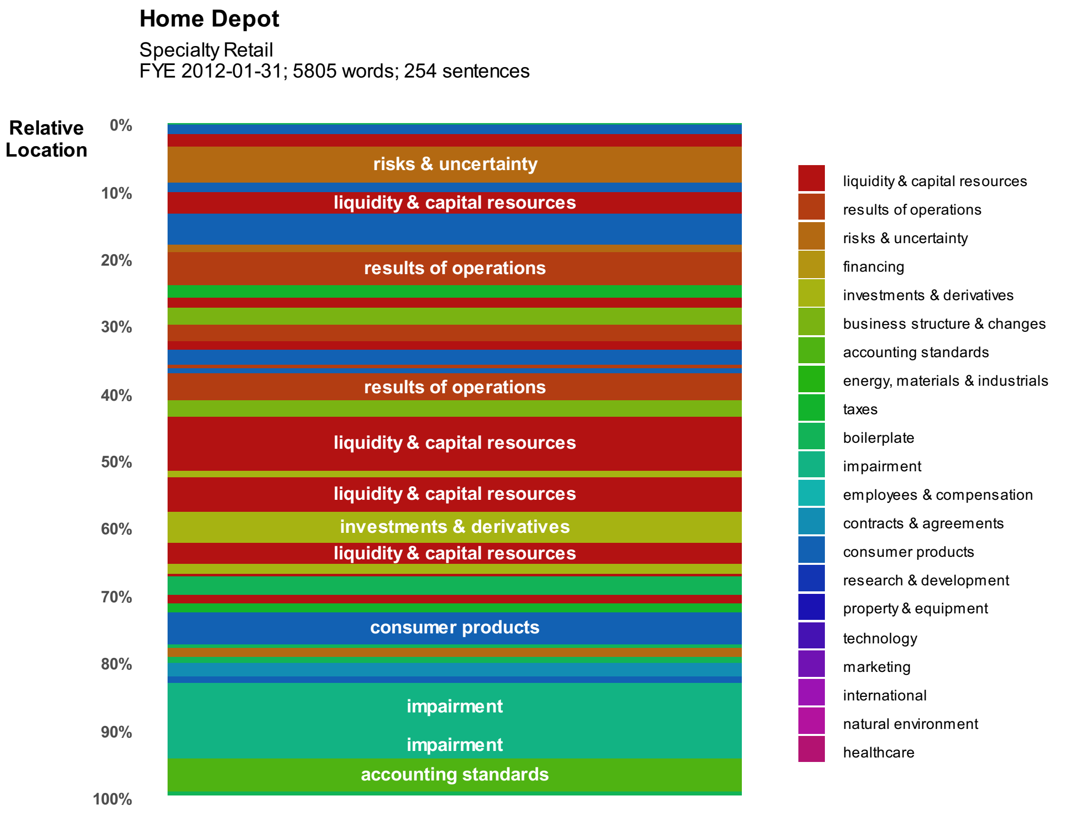


FIGURE 2 (Continued)

For our disclosure contextual analysis, we further aggregate the 21 topic categories into four topic groups: required, traditional, intangibles, and other. "Required topics" include those required by the SEC, such as liquidity and capital resources, results of operations, risks and uncertainty, new accounting standards, and boilerplate language.[17] "Traditional topics" include those often provided by firms even though not required. "Intangibles topics" include those related to business changes, investments that are expensed rather than capitalized under GAAP, and long-term commitments that are not reflected in the balance sheet. An example is from Allscripts Healthcare Solutions for fiscal year 2014: "To supplement our statement of operations, the table below presents a non-GAAP measure of research and development-related expenses that we believe is a useful metric for evaluating how we are investing in research and development." "Other topics" include geographic or industry-specific discussion.[18] We create variables *Required Topics*, *Traditional Topics*, and *Intangibles Topics* to measure the number of

[17]A common example of boilerplate language relates to the safe harbor protection—language that is required to receive the protection.

[18]Even though the discussion can be related to required, traditional, or intangibles content, the topic model likely generates new clusters because of the geographic and industry-specific phrases. LDA automatically determines topics and does not always lead to topics that are assigned as desired, such as into one of the first three topic groups. For example, Allergan states in its MD&A for fiscal year 2005, "We provide global marketing strategy teams to ensure development and execution of a consistent marketing strategy for our products in all geographic regions that share similar distribution channels and customers. Management evaluates its various global product portfolios on a revenue basis, which is presented below. We also report sales performance using the non-GAAP financial measure of constant currency sales." Our topic model assigns this disclosure to "International," which belongs to "other topics," whereas human coding would assign the disclosure to "marketing," which belongs to "intangibles topics."

TABLE 6 Context of non-GAAP and forward-looking disclosures.

Panel A: Mean values for firm-years in the highest versus lowest *Inadequacy* deciles		
	Highest *Inadequacy* decile	Lowest *Inadequacy* decile
Number of sentences in the entire MD&A for:		
Required topics	63	63
Traditional topics	**160**	**189**
Intangibles topics	**42**	**32**
Count of keywords for non-GAAP disclosure in:		
Required topics	0.161	0.137
Traditional topics	**1.055**	**0.886**
Intangibles topics	**0.123**	**0.076**
Number of forward-looking statements in:		
Required topics	**4.542**	**4.275**
Traditional topics	**13.131**	**14.014**
Intangibles topics	**4.077**	**2.947**

Panel B: Context of non-GAAP disclosure in the MD&A			
	Count of keywords for non-GAAP disclosure in		
	Required topics (1)	Traditional topics (2)	Intangibles topics (3)
Intercept	−0.152***	−0.331***	−0.156***
	(−8.74)	(−10.08)	(−9.04)
Inadequacy	0.025***	0.071***	0.019***
	(3.11)	(4.32)	(2.66)
Size	0.026***	0.078***	0.022***
	(11.75)	(17.99)	(10.55)
MB	0.002***	0.002	0.002***
	(2.58)	(1.63)	(3.30)
ROA	−0.002	0.086***	−0.001
	(−0.27)	(4.14)	(−0.12)
StdRet	0.001***	0.003***	0.001***
	(3.24)	(4.60)	(3.86)
Year FE	Yes	Yes	Yes
Industry FE	No	No	No
Adjusted R^2	3.3%	8.0%	3.1%
N	36,552	36,552	36,552

Panel C: Context of forward-looking disclosure in the MD&A			
	Number of forward-looking statements in		
	Required topics (1)	Traditional topics (2)	Intangibles topics (3)
Intercept	2.355***	−2.836***	−0.669
	(9.62)	(−6.05)	(−2.21)
Inadequacy	0.402***	−0.258	0.790***
	(3.33)	(−1.16)	(5.11)

(Continues)

TABLE 6 (Continued)

Panel C: Context of forward-looking disclosure in the MD&A

	Number of forward-looking statements in		
	Required topics (1)	**Traditional topics (2)**	**Intangibles topics (3)**
Size	0.383***	2.418***	0.427***
	(11.79)	(38.30)	(10.22)
MB	0.040***	−0.061***	0.129***
	(4.57)	(−3.70)	(9.72)
ROA	1.094***	−4.666***	−5.001***
	(6.34)	(−14.00)	(−14.58)
StdRet	0.022***	0.074***	0.056***
	(4.31)	(7.12)	(8.21)
Year FE	Yes	Yes	Yes
Industry FE	No	No	No
Adjusted R^2	7.0%	21.9%	5.4%
N	36,552	36,552	36,552

Note: We use TopicTiling to break down our sample MD&A documents into text segments and assign a topic category to each text segment. We focus on three groups of topic categories: required topics, traditional topics, and intangibles topics. See Supporting Information Appendix A for the technical details of topic discovery. See Supporting Information Appendix B for the list of 21 topic categories, their groupings, and descriptive statistics. Panel A presents the number of sentences in each of the three topic groups for firm-years in the highest versus lowest financial statement inadequacy deciles; the count of keywords for non-GAAP disclosure in required, traditional, and intangibles topics; and the number of forward-looking statements in required, traditional, and intangibles topics. We present in boldface the textual content that is significantly different between the highest and lowest *Inadequacy* deciles at the 5% level. Panel B examines the role of financial statement adequacy in a firm's intensity of non-GAAP disclosure when it discusses required, traditional, and intangibles topics. Panel C examines the role of financial statement adequacy in a firm's intensity of forward-looking disclosure when it discusses required, traditional, and intangibles topics. Panels B and C estimate OLS regressions and the estimations are robust to heteroskedasticity and within-firm error correlations. We present the estimated coefficients with *t*-statistics in parentheses. See Appendix 1 for variable definitions. The decile-ranked variable of *Inadequacy* is used in the regression analysis.
***, **, and * represent statistical significance in two-tailed tests at the 1%, 5%, and 10% levels, respectively.

sentences in an MD&A that belong to a given topic group. On average, the three topic groups include 65, 172, and 38 sentences, respectively (Panel A of Table 2). Firms with the lowest financial statement adequacy provide significantly more sentences on intangibles topics than firms with the highest financial statement adequacy, even though the former group has shorter MD&As (Panel A of Table 6).

We identify the topic groups in which non-GAAP disclosure appears. The univariate analysis in Panel A of Table 6 shows that firms with the lowest financial statement adequacy discuss non-GAAP measures in traditional topics and intangibles topics significantly more than firms with the highest financial statement adequacy. The multivariate analysis in Panel B indicates that the former firms discuss non-GAAP measures significantly more than the latter firms in all the three topic groups.

We count the number of FLS in required, traditional, and intangibles topics and provide the univariate comparisons in Panel A of Table 6. On average, firms with the lowest financial statement adequacy provide 4.5, 13.1, and 4.1 FLS in the required, traditional, and intangibles topics, respectively. These firms provide significantly more FLS in required and intangibles topics and fewer FLS in traditional topics than firms with the highest financial statement adequacy. The univariate comparisons of required and intangibles topics are confirmed by the multivariate analysis in Panel C. Thus, firms with less adequate financial statements provide more FLS when discussing required topics and intangibles. The analyses

TABLE 7 Alternative measures of financial statement adequacy.

Panel A: Mean values of MD&A textual content for subsamples

	Business Change		*Loss Firm*	
	Yes	No	Yes	No
Number of firm-years	19,578	16,974	11,218	25,334
NonGAAP Count	**6.332**	**3.243**	**3.986**	**5.301**
FLS	**31.623**	**27.814**	**29.056**	**30.207**
Quantitative FLS	**13.204**	**11.546**	**11.043**	**13.050**
Required Topics	**75.320**	**52.802**	**63.738**	**65.362**
Traditional Topics	**163.747**	**181.988**	**143.652**	**184.867**
Intangibles Topics	**40.661**	**33.882**	**43.876**	**34.696**

Panel B: Significant business changes

Dependent variable	Log (1 + *NonGAAP Count*)		*FLS*	
	(1)	(2)	(3)	(4)
Intercept	−0.698***	−0.768***	7.801***	8.498***
	(−13.14)	(−5.50)	(10.09)	(5.06)
Business Change	0.315***	0.210***	2.257***	1.840***
	(16.06)	(11.16)	(8.57)	(7.69)
Size	0.146***	0.165***	3.689***	3.739***
	(19.63)	(20.60)	(36.03)	(34.82)
MB	0.010***	0.006***	0.118***	0.041
	(4.07)	(2.61)	(4.44)	(1.59)
ROA	0.233***	−0.091**	−7.535***	−6.168***
	(6.23)	(−2.27)	(−12.61)	(−9.93)
StdRet	0.003**	0.002*	0.150***	0.139***
	(2.54)	(1.94)	(9.22)	(8.70)
Year FE	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes
Adjusted R^2	15.0%	21.0%	22.4%	26.2%
N	36,552	36,552	36,552	36,552

Panel C: Loss firms

Dependent variable	Log (1 + *NonGAAP Count*)		*FLS*	
	(1)	(2)	(3)	(4)
Intercept	−0.676***	−0.798***	7.402***	7.833***
	(−12.43)	(−5.68)	(9.43)	(4.67)
Loss Firm	0.142***	0.122***	2.985***	2.426***
	(5.84)	(5.29)	(9.67)	(8.10)
Size	0.161***	0.181***	3.835***	3.905***
	(21.48)	(23.13)	(37.70)	(36.94)
MB	0.011***	0.006**	0.133***	0.047*
	(4.40)	(2.54)	(4.94)	(1.81)
ROA	0.340***	−0.017	−4.573***	−4.025***
	(7.97)	(−0.39)	(−7.00)	(−6.04)

(Continues)

TABLE 7 (Continued)

Panel C: Loss firms

Dependent variable	**Log (1 + *NonGAAP Count*)**		***FLS***	
	(1)	**(2)**	**(3)**	**(4)**
StdRet	0.004***	0.002*	0.135***	0.124***
	(3.16)	(1.71)	(8.45)	(7.94)
Year FE	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes
Adjusted R^2	13.5%	20.5%	22.4%	26.2%
N	36,552	36,552	36,552	36,552

Note: We use alternative measures of financial statement adequacy to re-estimate our primary models of non-GAAP and forward-looking disclosures. Panel A presents the mean values of MD&A textual content variables for (1) firms that have experienced significant business changes during fiscal year *t* versus those without such changes and (2) firms that report losses for fiscal year *t* versus those without losses. We present in boldface the textual content that is significantly different between the subsamples at the 5% level. Panel B uses *Business Change* and Panel C uses *Loss Firm* as the alternative measure of financial statement adequacy, respectively. See variable definitions in Appendix 1. The OLS estimations are robust to heteroskedasticity and within-firm error correlations. We present the estimated coefficients with *t*-statistics in parentheses.

***, **, and * represent statistical significance in two-tailed tests at the 1%, 5%, and 10% levels, respectively.

in this subsection suggest that managers provide more narrative disclosures about issues for which GAAP reporting is deficient to help investors understand the reported financials and predict future performance.

5.2 | Alternative measures of financial statement adequacy

Our primary measure of financial statement adequacy uses a rolling window of 5 years of data.[19] In this subsection, we use alternative measures that require information only about a firm's sample year *t*.[20] The first measure is *Business Change*, which equals one if the firm reports any restructuring charges for fiscal year *t* or reports special items in a magnitude larger than the sample median and zero otherwise.[21] Restructuring represents organizational changes; special items with a large magnitude represent large operational changes. *Business Change* captures the main reason articulated by Lev and Zarowin (1999) for deficient GAAP reporting and therefore the value of one for this variable represents low financial statement adequacy. The second measure is *Loss Firm*, which equals one if the firm reports negative net income before extraordinary items and zero otherwise. Financial information of loss firms is less informative to investors and therefore the value of one represents low financial statement adequacy (Collins et al., 1997; Hayn, 1995).

Panel A of Table 7 contrasts MD&A textual attributes for firms with versus without business changes. Firms with business changes discuss non-GAAP measures more, provide more FLS as well as more quantitative FLS, and discuss required topics and intangibles to a greater extent than firms without business changes. Panel B confirms in multivariate analysis that firms with significant business changes indeed provide more non-GAAP and FLS disclosures than

[19]One concern about the measure is that it might reflect financial statement adequacy for the 5-year period as a whole, not precisely for year *t*. In untabulated analysis, we estimate Equation (1) for each year-quarter cross-sectionally using firms in the lowest *Inadequacy* decile and then firms in the highest decile. The mean value of R^2 from 52 quarterly regressions for the former firms is 62.7%, significantly higher than the value of 50.4% for the latter firms. This comparison suggests that our measure behaves as expected for year *t*.

[20]We use earnings persistence as an alternative measure of financial statement adequacy and find that firms with lower earnings persistence provide more non-GAAP and FLS disclosures in the MD&A (untabulated).

[21]Among the 36,552 firm-years, 9,781 have nonzero restructuring charges and 23,790 have nonzero special items.

CAAA ACPC CONTEMPORARY ACCOUNTING RESEARCH RECHERCHE COMPTABLE CONTEMPORAINE

TABLE 8 Changes analyses using alternative measures of financial statement adequacy.

Panel A: Significant business changes

Dependent variable	Δ*NonGAAP Count* (1)	Δ*NonGAAP Count* (2)	Δ*FLS* (3)	Δ*FLS* (4)
Intercept	0.400***	0.376***	6.486***	6.691***
	(23.11)	(16.57)	(32.74)	(21.43)
Δ*Business Change*	0.028***	0.028***	0.219**	0.218**
	(4.34)	(4.32)	(2.36)	(2.34)
Δ*Size*	0.142***	0.144***	2.524***	2.493***
	(8.59)	(8.69)	(10.50)	(10.32)
Δ*MB*	0.003**	0.003*	0.034	0.034
	(2.10)	(1.92)	(1.54)	(1.51)
Δ*ROA*	−0.105***	−0.112***	−3.130***	−3.133***
	(−3.92)	(−4.17)	(−7.88)	(−7.87)
Δ*StdRet*	0.001*	0.001*	0.013*	0.011
	(1.82)	(1.68)	(1.69)	(1.40)
Year FE	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes
Adjusted R^2	2.8%	3.0%	6.9%	6.9%
N	32,030	32,030	32,030	32,030

Panel B: Loss firms

Dependent variable	Δ*NonGAAP Count* (1)	Δ*NonGAAP Count* (2)	Δ*FLS* (3)	Δ*FLS* (4)
Intercept	0.402***	0.377***	6.504***	6.699***
	(23.20)	(16.63)	(32.85)	(21.43)
Δ*Loss Firm*	0.051***	0.050***	0.504***	0.490***
	(5.09)	(4.99)	(3.55)	(3.44)
Δ*Size*	0.140***	0.143***	2.505***	2.475***
	(8.49)	(8.59)	(10.44)	(10.27)
Δ*MB*	0.003**	0.003*	0.035	0.035
	(2.13)	(1.94)	(1.57)	(1.54)
Δ*ROA*	−0.057**	−0.065**	−2.629***	−2.647***
	(−2.01)	(−2.29)	(−6.28)	(−6.31)
Δ*StdRet*	0.001	0.001	0.012	0.010
	(1.64)	(1.52)	(1.57)	(1.29)
Year FE	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes
Adjusted R^2	2.8%	3.0%	7.0%	7.2%
N	32,030	32,030	32,030	32,030

Note: The table reports the OLS estimation of the changes regressions. Except for the fixed effect variables, each variable in the changes model is measured as the difference between the variable for fiscal year t and the variable for fiscal year $t-1$ as defined in the levels models in Tables 4 and 5. Panel A (B) uses *Business Change* (*Loss Firm*) as the measure of financial statement adequacy. See other notes in Table 7.

firms without business changes. The changes analysis in Panel A of Table 8 conveys the same message.

Panel A of Table 7 also contrasts loss firms with profit firms. Loss firms provide fewer non-GAAP and FLS disclosures but discuss intangibles more often than do profit firms. In multivariate analysis reported in Panel C, however, loss firms provide more non-GAAP and FLS disclosures in the MD&A than profit firms, suggesting that the univariate comparisons were driven by firm characteristics that are effectively controlled in the multivariate analysis. The changes analysis in Panel B of Table 8 yields results consistent with Panel C of Table 7.

6 | CONCLUSION

Our study examines how firms use the MD&A channel when their financial statement channel is inadequate. We find that firms with less adequate financial statements discuss non-GAAP measures more and provide more forward-looking statements, including those with verifiable quantitative information. These firms' non-GAAP discussions appear significantly in required topics, traditional topics, and intangibles topics and their forward-looking statements appear significantly in required topics and intangibles topics. These findings suggest that managers use the MD&A, a relatively more flexible channel, to a greater extent to convey information when their financial statement channel is less adequate.

Our study has two caveats. First, we cannot entirely rule out endogeneity as an alternative explanation for our findings. As in Li et al. (2013) and Merkley (2014), our primary results are obtained from levels analyses. We have attempted to address the omitted correlated variable problem by using changes analyses. Second, we do not examine the credibility of narrative disclosure. For example, we do not verify firms' non-GAAP and forward-looking disclosures with realizations. Despite these caveats, our study contributes to the literature by (1) addressing an important research question that should be of interest to investors, managers, and regulators and (2) introducing textual analysis techniques that may help future researchers.

ACKNOWLEDGMENTS

We thank Pawel Bilinski, Khrystyna Bochkay, Mark Bradshaw, Peter Clarkson (editor), Ilia Dichev, Eduardo Fuste (discussant), Olivia Huseman (discussant), Jay Heon Jung, Tim Martens, Greg Miller, Karen Nelson (discussant), Pat O'Brien, Matthew Phillips, Shiva Rajgopal, Mani Sethuraman (discussant), Steve Utke, Eric Weisbrod, Mark Zakota, two anonymous reviewers, and workshop participants at the University of Connecticut, University of Georgia, Hong Kong University of Science and Technology, University of Florida, City University of London, and Office of Economic and Risk Analysis of the PCAOB. We thank conference participants of the 2018 Florida Accounting Symposium, 2019 McGill University Inaugural Accounting Conference, and 2019 American Accounting Association Annual Meeting. We thank Paige Bruman for able research assistance. Hinson thanks the Fisher School of Accounting and the Ernst & Young Professorship for financial support. Tucker thanks the J. Michael Cook/Deloitte Professorship for financial support.

ORCID

Lisa A. Hinson https://orcid.org/0000-0002-6575-3009
Jennifer Wu Tucker https://orcid.org/0000-0003-3553-4531

REFERENCES

Amir, E., & Lev, B. (1996). Value-relevance of nonfinancial information: The wireless communications industry. *Journal of Accounting and Economics*, *22*(1–3), 3–30.

Anilowski, C., Feng, M., & Skinner, D. J. (2007). Does earnings guidance affect market returns? The nature and information content of aggregate earnings guidance. *Journal of Accounting and Economics*, *44*(1–2), 36–63.
Bagnoli, M., & Watts, S. G. (2007). Financial reporting and supplemental voluntary disclosures. *Journal of Accounting Research*, *45*(5), 885–913.
Balachandran, S., & Mohanram, P. (2011). Is the decline in the value relevance of accounting driven by increased conservatism? *Review of Accounting Studies*, *16*, 272–301.
Banker, R. D., Huang, R., & Natarajan, R. (2009). Incentive contracting and value relevance of earnings and cash flows. *Journal of Accounting Research*, *47*(3), 647–678.
Barth, M. E., Beaver, W. H., & Landsman, W. R. (2001). The relevance of the value relevance literature for financial accounting standard setting: Another view. *Journal of Accounting and Economics*, *31*(1–3), 77–104.
Barth, M. E., Li, K., & McClure, C. G. (2023). Evolution in value relevance of accounting information. *The Accounting Review*, *98*(1), 1–28.
Beaver, W. H., McNichols, M. F., & Wang, Z. Z. (2018). The information content of earnings announcements: New insights from intertemporal and cross-sectional behavior. *Review of Accounting Studies*, *23*, 95–135.
Beaver, W. H., McNichols, M. F., & Wang, Z. Z. (2020). Increased market response to earnings announcements in the 21st century: An empirical investigation. *Journal of Accounting and Economics*, *69*(1), 1–21.
Bentley, J. W., Christensen, T. E., Gee, K. H., & Whipple, B. C. (2018). Disentangling managers' and analysts' non-GAAP reporting. *Journal of Accounting Research*, *56*(4), 1039–1080.
Bhattacharya, N., Black, E. L., Christensen, T. E., & Larson, C. R. (2003). Assessing the relative informativeness and permanence of pro forma earnings and GAAP operating earnings. *Journal of Accounting and Economics*, *36*(1–3), 285–319.
Black, D. E., & Christensen, T. E. (2009). US managers' use of "Pro Forma" adjustments to meet strategic earnings targets. *Journal of Business, Finance & Accounting*, *36*(3–4), 297–326.
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.
Bochkay, K., Brown, S. V., Leone, A., & Tucker, J. W. (2023). Textual analysis in accounting: What's next? *Contemporary Accounting Research*, *40*(2), 765–805.
Bochkay, K., & Levine, C. B. (2019). Using MD&A to improve earnings forecasts. *Journal of Accounting, Auditing & Finance*, *34*(3), 458–482.
Bozanic, Z., Roulstone, D. T., & Van Buskirk, A. (2018). Management earnings forecasts and other forward-looking statements. *Journal of Accounting and Economics*, *65*(1), 1–20.
Brown, L. D., Call, A. C., Clement, M. B., & Sharp, N. Y. (2019). Managing the narrative: Investor relations officers and corporate disclosure. *Journal of Accounting and Economics*, *67*(1), 58–79.
Brown, N. C., Crowley, R. M., & Elliott, W. B. (2020). What are you saying? Using *topic* to detect financial misreporting. *Journal of Accounting Research*, *58*(1), 237–291.
Brown, S., Lo, K., & Lys, T. (1999). Use of *R*-squared in accounting research: Measuring changes in value relevance over the last four decades. *Journal of Accounting and Economics*, *28*(2), 83–115.
Brown, S. V., & Knechel, W. R. (2016). Auditor-client compatibility and audit firm selection. *Journal of Accounting Research*, *54*(3), 725–775.
Brown, S. V., & Tucker, J. W. (2011). Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research*, *49*(2), 309–346.
Bryan, S. H. (1997). Incremental information content of required disclosures contained in management discussion and analysis. *The Accounting Review*, *72*(2), 285–301.
Campbell, J., Chen, H., Dhaliwal, D., Lu, H., & Steele, L. (2014). The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies*, *19*, 396–455.
Cao, S. S., Guang, M., Tucker, J. W., & Wan, C. (2018). Technological peer pressure and product disclosure. *The Accounting Review*, *93*(6), 95–126.
Chasan, E. (2015, November 17). CFO journal: Big number. *Wall Street Journal*, B4.
Clarkson, P. M., Kao, J. L., & Richardson, G. D. (1994). The voluntary inclusion of forecasts in the MD&A section of annual reports. *Contemporary Accounting Research*, *11*(1), 423–450.
Cole, C. J., & Jones, C. L. (2004). The usefulness of MD&A disclosures in the retail industry. *Journal of Accounting, Auditing & Finance*, *19*(4), 361–388.
Collins, D. W., Maydew, E. L., & Weiss, I. S. (1997). Changes in the value-relevance of earnings and book values over the past forty years. *Journal of Accounting and Economics*, *24*(1), 39–67.
Davis, A. K., & Tama-Sweet, I. (2012). Managers' use of language across alternative disclosure outlets: Earnings press releases versus MD&A. *Contemporary Accounting Research*, *29*(3), 804–837.
Dichev, I. D., Graham, J. R., Harvey, C. R., & Rajgopal, S. (2013). Earnings quality: Evidence from the field. *Journal of Accounting and Economics*, *56*(2–3), 1–33.
Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from latent Dirichlet allocation. *Journal of Accounting and Economics*, *64*(2–3), 221–245.

Einhorn, E. (2005). The nature of the interaction between mandatory and voluntary disclosures. *Journal of Accounting Research*, *43*(4), 593–621.

Feldman, R., Govindaraj, S., Livnat, J., & Segal, B. (2010). Management's tone change, post announcement drift and accruals. *Review of Accounting Studies*, *15*, 915–953.

Francis, J., & Schipper, K. (1999). Have financial statements lost their value relevance? *Journal of Accounting Research*, *37*(2), 319–352.

Frankel, R., Jennings, J., & Lee, J. (2016). Using unstructured and qualitative disclosures to explain accruals. *Journal of Accounting and Economics*, *45*(2), 209–227.

Frankel, R., Jennings, J., & Lee, J. (2022). Disclosure sentiment: Machine learning vs. dictionary methods. *Management Science*, *68*(7), 5514–5532.

Gee, K. H. (2018). Readability, profitability, and discretionary MD&A text. Working paper, The Ohio State University.

Gow, I., Larcker, D., & Reiss, P. (2016). Causal inference in accounting research. *Journal of Accounting Research*, *54*(2), 477–523.

Hayn, C. (1995). The information content of losses. *Journal of Accounting and Economics*, *20*(2), 125–153.

Healy, P. M., & Palepu, K. G. (1993). The effect of firms' financial disclosure strategies on stock prices. *Accounting Horizons*, *7*(1), 1–11.

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

Hoberg, G., & Lewis, C. (2017). Do fraudulent firms produce abnormal disclosure? *Journal of Corporate Finance*, *43*, 58–85.

Hoberg, G., & Phillips, G. (2010). Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies*, *23*(10), 3773–3811.

Hribar, P., Mergenthaler, R., Roeschley, A., Young, S., & Zhao, C. X. (2022). Do managers issue more voluntary disclosure when GAAP limits their reporting discretion in financial statements? *Journal of Accounting Research*, *60*(1), 299–351.

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst information discovery and interpretation roles: A topic modelling approach. *Management Science*, *64*(6), 2833–2855.

Jo, K. M., & Yang, S. (2020). SEC comment letters on firms' use of non-GAAP measures: The determinants and firms' responses. *Accounting Horizons*, *34*(2), 167–184.

Keskek, S., Tse, S., & Tucker, J. W. (2014). Analyst information production and the timing of annual earnings forecasts. *Review of Accounting Studies*, *19*(4), 1504–1531.

Knutson, P. H. (1993). *Financial reporting in the 1990s and beyond: A position paper*. Association for Investment Management and Research.

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

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.

Leung, E., & Veenman, D. (2018). Non-GAAP earnings disclosure in loss firms. *Journal of Accounting Research*, *56*(4), 1083–1137.

Lev, B. (1989). On the usefulness of earnings and earnings research: Lessons and directions from two decades of empirical research. *Journal of Accounting Research*, *27*(Suppl), 153–192.

Lev, B., & Zarowin, P. (1999). The boundaries of financial reporting and how to extend them? *Journal of Accounting Research*, *37*(2), 353–385.

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

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

Li, F., Lundholm, R., & Minnis, M. (2013). A measure of competition based on 10-K filings. *Journal of Accounting Research*, *51*(2), 399–436.

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

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

Merkley, K. J. (2014). Narrative disclosure and earnings performance: Evidence from R&D disclosures. *The Accounting Review*, *89*(2), 725–757.

Monga, V. (2017, March 7). CFO journal: Firms hit by lack of accountants. *Wall Street Journal*, B6.

Muslu, V., Radhakrishnan, S., & Subramanyam, K. R. (2015). Forward-looking MD&A disclosures and the information environment. *Management Science*, *61*(5), 931–948.

Ohlson, J. (1995). Earnings, book values, and dividends in security valuation. *Contemporary Accounting Research*, *11*(2), 661–687.

Peterson, K., Schmardebeck, R., & Wilks, T. J. (2015). The earnings quality and information processing effects of accounting consistency. *The Accounting Review*, *90*(6), 2483–2514.
Riedl, M., & Biemann, C. (2012a). TopicTiling: A text segmentation algorithm based on LDA. In *Proceedings of ACL 2012 Student Research Workshop* (pp. 37–42). Association for Computational Linguistics.
Riedl, M., & Biemann, C. (2012b). Text segmentation with topic models. *Journal for Language Technology and Computational Linguistics*, *27*(1), 47–69.
Rogers, R. K., & Grant, J. (1997). Content analysis of information cited in reports of sell-side financial analysts. *Journal of Financial Statement Analysis*, *3*(1), 17–30.
SEC. (2008). *Financial Reporting Manual*. https://www.sec.gov/corpfin/cf-manual/topic-9
SEC. (2013). *Report on review of disclosure requirements in Regulation S-K*. https://www.sec.gov/files/reg-sk-disclosure-requirements-review.pdf
Srivastava, A. (2014). Why have measures of earnings quality changed over time? *Journal of Accounting and Economics*, *57*(2–3), 196–217.
Sun, Y. (2010). Do MD&A disclosures help users interpret disproportionate inventory increases? *The Accounting Review*, *85*(4), 1411–1440.
Tavcar, L. R. (1998). Make the MD&A more readable. *The CPA Journal*. http://archives.cpajournal.com/1998/0198/newsViews/0198NV7.htm
Whited, R., Swanquist, Q., Shipman, J., & Moon, J. (2022). Out of control: The (over) use of controls in accounting research. *The Accounting Review*, *97*(3), 395–413.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Brown, S. V., Hinson, L. A., & Tucker, J. W. (2024). Financial statement adequacy and firms' MD&A disclosures. *Contemporary Accounting Research*, *41*(1), 126–162. https://doi.org/10.1111/1911-3846.12919

APPENDIX 1: VARIABLE DEFINITIONS

Variable	Definition
Explanatory variables	
Inadequacy	One minus the R^2 from regressing the firm's stock price on its earnings per share (ibq/cshoq) and book value per share (ceqq/cshoq) using 20 quarters ending with the 4th quarter of fiscal year *t*. We require at least 16 quarterly observations to estimate a firm's time-series regression. We use the decile-ranked variable in regression analyses with zero for the lowest decile and one for the highest decile
Business Change	Indicator variable proxying for significant business changes at the sample firm during fiscal year *t*. The variable is one if the firm reports any restructuring charges (rcp) for fiscal year *t* or special items scaled by total assets (spi/at) in a magnitude larger than the sample median, and zero otherwise
Loss Firm	One if the firm reports negative net income before extraordinary items (ib) for fiscal year *t*, and zero otherwise
MD&A variables	
Descriptive variables	
MDA Sentences	Number of sentences in the firm's MD&A for fiscal year *t*
MDA Words	Number of words in the firm's MD&A for fiscal year *t*
MDA Numbers	Number of numbers in the firm's MD&A for fiscal year *t*. Here, "numbers" include any named entity recognition (NER) in the categories of *MONEY*, *PERCENT*, and

(Continues)

APPENDIX 1 (Continued)

Variable	Definition
	CARDINAL. See technical details of NER in Appendix 2. The variable includes the numbers in tables
MDA Words to Numbers	Number of words divided by the number of numbers in the firm's MD&A for fiscal year *t*
MDA Tables	Number of tables in the firm's MD&A for fiscal year *t*. A table is identified as at least two sets of two or more consecutive numbers with no other types of text in between. For example, "$1,934,222 22.3%" is a set of two consecutive numbers, but "45% and 65%" is not
MDA Text Segments	Number of text segments in the firm's MD&A for fiscal year *t*. A text segment is a group of consecutive sentences that discuss the same topic category. We identify text segments using a text segmentation technique called TopicTiling. See Supporting Information Appendix A for technical details
NonGAAP Dummy	One if the firm uses case-insensitive keywords for non-GAAP disclosure in the firm's MD&A for fiscal year *t*, and zero otherwise. The set of keywords includes "non-gaap," "non gaap," and "nongaap," tokens beginning with "EBIT" (common matches include "EBIT," "EBITA," "EBITDA," "EBITDAR," "EBITDAS," and "EBITDAX"), "adjusted" + (0, 1, or 2 other words) + "earnings"/"income"/"eps" (e.g., "adjusted net income" and "adjusted basic EPS"), and "free cash flow"
Content variables	
NonGAAP Count	Occurrences of case-insensitive keywords for non-GAAP disclosure in the firm's MD&A for fiscal year *t*. See the set of keywords in the definition for *NonGAAP Dummy*
FLS	Number of forward-looking sentences in the firm's MD&A for fiscal year *t*. We use deep learning to identify forward-looking sentences. See Appendix 2 for technical details
Quantitative FLS	Number of forward-looking sentences in the firm's MD&A for fiscal year *t* that contain quantitative information. We use deep learning to identify forward-looking sentences as well as the existence of quantitative information in a sentence. See Appendix 2 for technical details
Required Topics	Number of sentences in the firm's MD&A for fiscal year *t* that relate to topics required by Regulation S-K. These sentences belong to the topic categories of liquidity and capital resources, results of operations, risks & uncertainty, accounting standards, and boilerplate. See Supporting Information Appendix A for the technical details of topic discovery and Appendix B for the list of 21 topic categories
Traditional Topics	Number of sentences in the firm's MD&A for fiscal year *t* that relate to disclosure traditionally provided by firms outside of the required topics. These sentences belong to the topic categories of taxes, investments and derivatives, impairment, employees and compensation, property and equipment, and financing. See Supporting Information Appendix A for the technical details of topic discovery and Appendix B for the list of 21 topic categories
Intangibles Topics	Number of sentences in the firm's MD&A for fiscal year *t* that relate to intangibles. These sentences belong to the topic categories of business structure and changes, contracts and agreements, research and development, marketing, and technology. See Supporting Information Appendix A for technical details and Appendix B for the list of 21 topic categories
Control variables	
Total Assets	Firm's total assets (at) in millions of dollars at the end of fiscal year *t*. *Size* is the natural logarithm of *Total Assets*
MB	Firm's market value of equity divided by its book value of equity (prcc_f × csho/ceq) at the end of fiscal year *t*
ROA	Firm's net income before extraordinary items divided by its total assets (ib/at) at the end of fiscal year *t*

APPENDIX 1 (Continued)

Variable	Definition
StdRet	Standard deviation of the firm's monthly raw stock returns during fiscal year *t*
Analysts	Number of unique analysts from the I/B/E/S detail dataset who issue forecasts of annual EPS for the firm's fiscal year *t*
M&A	One if the acquisition-sale contribution (aqs) divided by sales (sale) is greater than 1% or if the value of acquisitions (aqc) divided by total assets (at) is greater than 2% in fiscal year *t*, and zero otherwise. See Muslu et al. (2015)
Company Age	Number of years between fiscal year *t* and the year of the firm's first appearance in Compustat
Segments	Number of business segments with nonzero net sales reported by the firm for fiscal year *t*, according to Compustat Segments database
Compustat Items	Number of nonzero and nonmissing items in Compustat for the firm for fiscal year *t*

Note: We include Compustat variable names in lower case in parentheses.

APPENDIX 2: DEEP LEARNING APPROACH TO IDENTIFY FORWARD-LOOKING DISCLOSURE

We use deep learning to identify forward-looking disclosure as well as forward-looking disclosure that contains quantitative information. Specifically, we use named entity recognition (NER) to identify quantitative information and use the Classification task to identify forward-looking sentences.

Named entity recognition

NER is a task of information extraction that locates and classifies a group of one or more consecutive words, numbers, or symbols into predefined categories. We designate three categories to identify quantitative information: *MONEY* (e.g., $1 million), *PERCENT* (e.g., 5%), and *CARDINAL* (e.g., 1 million). The last category ignores years, dates, list items, and numeric-looking but not quantitative symbols (e.g., Phase 3 trials and FIN 48).

Classification

Classification is a task of determining whether a unit of text (e.g., words, sentences, and paragraphs) belongs to a predefined category. We perform this task at the sentence level and use the multi-class categories of *FLS* (*specific*), *FLS* (*not specific*), and *Not-FLS*. An example of *FLS* (*specific*) is, "We expect research and development costs to increase in 2009, due to clinical testing of our lead product candidates." An example of *FLS* (*not specific*) is, "If the company's plans or assumptions change or prove to be inaccurate, the foregoing sources of funds may prove to be insufficient." Most sentences in *FLS* (*not specific*) are risk disclosure or cautionary language. Our algorithm classifies a sentence into one and only one of the three categories. For our empirical analyses, we count only *FLS* (*specific*) as forward-looking statements and refer to a sentence as a quantitative forward-looking statement if it is *FLS* (*specific*) and contains an NER of either *MONEY*, *PERCENT*, or *CARDINAL*.

Training data

For NER, we use 3,150 paragraphs and 350 tables as our training sample and 900 paragraphs and 100 tables as our validation sample. We include both paragraphs and tables because an NER is likely to occur in either type. Because tables are fairly uniform in structure for the machine to parse, tables account for only 10% of our training and validation samples.

For Classification, we use 2,250 sentences as the training sample and 500 sentences as the validation sample.[22] Because forward-looking statements account for a small percentage of all sentences in 10-K reports, we use stratified random sampling to select our training and validation observations: half of the observations are randomly selected from sentences that contain at least one future-oriented keyword as commonly identified by prior research and the other half are randomly selected from sentences that do not contain any such keyword. Stratified random sampling ensures that our training and validation observations are representative and provide enough variation for the machine to learn patterns and test them.

Each item (e.g., a paragraph or table for NER and a sentence for Classification) in the training and validation samples is hand-coded. This process discovers various ways in which a unit of consecutive words, numbers, or symbols of variable length fits our NER categories and whether a sentence is a forward-looking statement. We then use the hand-coded samples to train a deep learning model and generate out-of-sample predictions for the validation sample so that we can assess model performance by comparing the predictions with human classifications. We implement a deep learning model called CNN. The model has been pre-trained on a wide variety of English text and users can fine-tune the model for their own settings.[23] The model learns to recognize the patterns of word, number, and symbol combinations in unseen text that resemble the human-coded units and then labels or classifies unseen units.

Validation of NER

At the end of our iterative process of training and evaluating the machine, our model demonstrates the precision rate, recall rate, and *F*1 score for NER (see Table 9).

TABLE 9 Validation performance statistics for NER.

	Precision (%)	Recall (%)	*F*1 score (%)
MONEY	98.8	98.4	98.6
PERCENT	98.4	99.3	98.8
CARDINAL	98.9	98.8	98.8

Validation of forward-looking statement classifications and comparison with the keyword approach

We use the more disaggregated categories of *FLS* (*specific*), *FLS* (*not specific*), and *Not-FLS* for flexibility, but eventually combine the latter two categories for our empirical analyses and therefore collapse three categories into two categories of *FLS* versus *NFLS*. Here, a sentence is *FLS* if the machine classifies it as *FLS* (*specific*); the sentence is *NFLS* otherwise. In Table 10, we

[22]The final sizes of the training and validation samples are determined iteratively by the machine learning *F*1 score, which is the harmonic mean of the precision rate (i.e., what proportion of positive identifications was actually correct?) and the recall rate (i.e., what proportion of actual positives was identified correctly?). We started with 500 training items, calculated the *F*1 score, and added 250 additional items if the *F*1 score did not reach 90% for NER and was still rising for Classification. This process iterated until our *F*1 score goal was achieved.

[23]We use the en_core_web_lg model from Spacy 3.3.1. See details at https://spacy.io/usage/training. The idea behind CNN came from image processing. The input for CNN is words, numbers, symbols, or sentences, which are represented by a matrix. The model uses many layers, each of which takes the output from the previous layer as input and estimates weights for the input in generating the output before proceeding to the next layer.

present the classifications of human coder versus machine for the 500 observations in our validation sample, which contains 81 *FLS* and 419 *NFLS* observations (i.e., 16.2% *FLS*). We report the results under two alternative decision rules: (1) default decision rule—a sentence is assigned to one of the three categories with the highest estimated likelihood and (2) 50% cutoff rule—a sentence is assigned to *FLS* (*specific*) if the estimated likelihood for this category is at least 50%.

TABLE 10 Forward-looking statement classifications by keyword and deep learning approaches versus human coding.

		Deep learning (DL) approach				Keyword approach	
		Default decision rule		50% cutoff rule			
		FLS	*NFLS*	*FLS*	*NFLS*	*FLS*	*NFLS*
Human coding	*FLS*	57	24	54	27	76	5
	NFLS	41	378	35	384	128	291

We calculate the precision, recall, and *F*1 score for our binary classifications of *FLS* versus *NFLS*. We add a performance metric "accuracy" because the previous three metrics ignore true negatives (i.e., negatives that have been correctly identified). The recall rate is one minus Type II error rate, but the precision rate is not a function of Type I error (i.e., false positives as a percentage of negatives, including true negatives and false alarms) (see footnote 29 of Bochkay et al., 2023). In our opinion, the error of falsely calling a sentence forward-looking when it is not is serious because the vast majority of sentences in a 10-K report are not forward looking and therefore subject to this error. The accuracy rate takes into account both Type I and Type II errors.[24] These metrics are as shown in Table 11.

TABLE 11 Validation performance statistics for forward-looking statement classifications.

	FLS (%)	Type I error (%)	Type II error (%)	Accuracy (%)	Precision (%)	Recall (%)	*F*1 score (%)
Human coding	16.2						
DL (default)	19.6	9.8	29.6	87.0	58.2	70.4	63.7
DL (50% cutoff)	17.8	8.3	33.3	87.6	60.7	66.7	63.5
Keyword approach	40.8	30.5	6.2	73.4	37.3	93.8	53.3

Abbreviation: DL, deep learning.

The deep learning approach using the 50% cutoff yields an accuracy rate of 87.6% and a precision of 60.7%, both higher than that using the default decision rule. In untabulated analyses, we observe that the 50% cutoff rule also results in a higher *F*1 score than the alternatives of 75% and 90% cutoffs. Therefore, we decide to use the 50% cutoff rule in classifying our full sample.

We compare our deep learning classification approach with the keyword search approach. Li (2010) is the first study using keywords to identify forward-looking statements. His word list is subsequently improved and updated by Muslu et al. (2015) and Bozanic et al. (2018). We use the word list provided by Bozanic et al. to implement the keyword approach and we add its

[24]The program we use does not report the accuracy rate or Type I error rate. This is why we rely upon the *F*1 score instead of the accuracy rate in the iterative process of training and evaluating the model.

performance metrics to Tables 10 and 11. The keyword approach yields a low Type II error rate but a high Type I error rate, with an overall accuracy of 73.4%. Both our deep learning applications produce substantially higher accuracy than the keyword approach. Because our full sample contains a greater percentage of sentences that are not forward looking (about 92%) than the percentage of our validation sample (about 84%) and our approach has a much lower Type I error rate than the keyword approach, our approach's advantage of classification accuracy over the keyword approach should be even greater for our full sample than for the validation sample.

The above accuracy improvement over the keyword approach is attributable to two strengths of deep learning. First, deep learning can learn subtle relationships that the keyword approach cannot. For example, "Because of inherent uncertainties in estimating costs and revenues, it is at least reasonably possible that the estimates used will change" and "To date, we have only generated limited revenue from our new strategic focus, and we do not know if we will ever generate significant revenue from our new products" are not forward looking but the keyword approach would identify them as forward looking. Second, deep learning is robust to coding errors because there are many other correctly coded items for the adaptive model to outweigh those errors.

Software implementation

We use (1) Spacy, a deep-learning-based natural language processing free and open-source library that provides a broad set of production-quality functionality out of the box, and (2) Prodigy, a web-based interface for efficient NER and Classification tagging to be used as input for Spacy modeling. Within Spacy, we use CNN-based models and word representations.[25]

[25]For an earlier version of our manuscript, we used Spacy 2.3, which classified 13.6% of the training data observations as FLS. For our current manuscript, we use Spacy 3.3.1, which classifies 17.8% of the training data observations as FLS. The difference between the two percentages is 4.2%. The average firm in our previous manuscript had 15 FLS; the average firm in our current manuscript has 30 FLS. The difference can be roughly explained by $362 \times 4.2\% = 15.2$, where "362" is the average number of sentences in a sample MD&A document.