Qualitative Disclosure and Changes in Sell-Side Financial Analysts' Information Environment*

ZAHN BOZANIC, *The Ohio State University*

MAYA THEVENOT, *Florida Atlantic University*

1. Introduction

Firms communicate with analysts and other investors through a variety of media. Following Regulation Fair Disclosure, the medium for communication can take such public forms as press releases, television interviews, conference calls, or mandatory financial reporting. For several decades, researchers have traditionally measured the information content of disclosure using a *quantitative* summary earnings figure, usually benchmarked against the consensus analyst forecast of earnings. Recently, a literature has evolved which uses tools from computational linguistics to measure information in disclosure *qualitatively*. To date, several papers have documented that such nonquantitative information is incrementally informative to quantitative earnings news (e.g., Davis, Piger, and Sedor 2012).

While the informativeness of quantitative earnings and management forecasts has been extensively studied and documented, our understanding of qualitative disclosure has only just begun. Beyond sentiment or readability, researchers know very little about the qualitative information signals analysts use. We extend the nascent qualitative disclosure literature by examining whether and how several novel aspects of firms' disclosures affect sell-side financial analysts' information environment. In doing so, we further our understanding of the qualitative information signals that analysts rely upon to form and update their beliefs. Consequently, our paper is related to a recent paper by Brown, Call, Clement, and Sharp (2014) who also attempt to open up the "black box" of sell-side financial analysts' inputs. In contrast to Brown et al.'s (2014) survey approach, we study how qualitative information contained in a routine and timely interim disclosure, earnings press releases, impacts analysts' information use and production.

We examine three characteristics of qualitative disclosure in this study. We first investigate how disclosure readability, originally pioneered in the accounting literature by Li (2008), affects analysts' information environment. If analysts add value through their information-processing skills, particularly as related to qualitative information, we expect disclosure to be more informative when a clear message is conveyed as processing costs are reduced (Bloomfield 2002). Our next measure of qualitative disclosure is quarter-over-quarter textual similarity, first introduced in the accounting literature by Brown and Tucker (2010). The predicted effect of this measure is ambiguous ex ante. On one hand, if managers provide textually similar disclosure quarter-over-quarter, to the extent that analysts' priors are confirmed by such disclosure, some of the analysts' uncertainty about the future may be resolved, which should manifest itself through an improved information environment. On the other hand, to the extent that more similar disclosure merely repeats

* Accepted by Jeffrey Callen. We thank Orie Barron, Jeffrey Callen, Darren Roulstone, Oded Rozenbaum, Tzachi Zach, Yong Yu, and two anonymous referees for many helpful comments and suggestions. The paper has also benefited from the research assistance of David Caban and Bret Johnson.

doi:10.1111/1911-3846.12123

information that had been previously disclosed, and is therefore redundant, we would expect analysts' information environment to remain unchanged. Our third qualitative measure of disclosure is *within-document* lexical diversity (Goel, Gangolly, Faerman, and Uzuner 2010; Humpherys, Moffitt, Burns, Burgoon, and Felix 2011), which measures the number of unique words (types) relative to the total number of words found within the disclosure. While more disclosure may be regarded by some as better for investors, if the disclosure is repetitive or unnecessarily verbose, that assumption may not hold (Levitt 1997). As such, lexical diversity can be characterized as capturing the level of information intensity, that is, the amount of unique information conveyed per unit of processing cost. Therefore, to the extent analysts use qualitative disclosure to infer useful information, we expect that the wider the standardized variety of information being disclosed, the larger the improvement in analysts' information environment.

To assess how qualitative characteristics of disclosure influence analysts' inputs, we consider five measures of analysts' information environment following Barron, Kim, Lim, and Stevens (1998, BKLS hereafter)—namely, overall and common uncertainty, consensus, and public and private information precision. These metrics are well suited for our study because they are comprised of analysts' two primary sources of information—private and common information—and capture distinct features of analysts' information environment. For example, if a particular disclosure cue is associated with decreases in analyst uncertainty, while the result may suggest the cue has information content, it does not provide evidence on how the cue alters analysts' relative use of public and private information. Further, while consensus can be employed to gain insights into analyst's use of public relative to total information, it does not inform us on the specific types of information being altered, which is why we also examine the precision of private and public information.[1] Moreover, we implement Sheng and Thevenot's (2012) modification of the BKLS proxies, which provides more accurate and reliable measures of analysts' information environment.[2]

The empirical results support the conclusion that the qualitative elements of disclosure we examine are informative as they affect analysts' information environment in a predictable manner. Specifically, we find that more readable, similar, and diverse disclosures are associated with decreases in both overall and common uncertainty. This suggests that similar disclosures over time reinforce a message and lead to an improvement in analysts' information environment rather than providing redundant information. The magnitudes of the decreases are strongest for diversity, which suggests that analysts' reduction in uncertainty is largely a result of informational parsimony. Further, with respect to the components of readability, and corroborating the diversity results, the association is driven by sentence length, rather than complex words, indicating that shorter sentences reduce uncertainty.

Even more interestingly, we find that higher similarity is associated with decreases in consensus and increases in private information precision, whereas more readable and diverse releases do not affect consensus, but are related to increases in both public and private information precision. This suggests that similarity is associated with higher increases in private information precision relative to increases in public information precision and analysts seem to use similar disclosures to develop primarily private information. As before, sentence length is driving the association between readability and information precision, which again corroborates our initial findings with respect to uncertainty. Lastly, we show that our results are stronger for large and heavily followed firms, which is not only

1. Per BKLS, consensus is defined as the ratio of common to total information.
2. While the GARCH-based (generalized autoregressive conditional heteroscedasticity) proxies used in Sheng and Thevenot (2012) allow us to alleviate concerns regarding potentially spurious inferences, we note as a caveat that such an estimation restricts our inferences to large, stable, and heavily followed firms.

consistent with large firms operating in rich information environments but also with analysts being attracted to firms with such environments, given the higher potential benefits (Bhushan 1989).

The most pertinent antecedent paper to ours can be found in Lehavy, Li, and Merkley (2011), who focus on the readability of annual reports, a proxy for analysts' information-processing costs. We extend the work of Lehavy et al. (2011) along three important dimensions. First, to gain a more complete picture of the types of qualitative information analysts use beyond those traditionally found in the literature, we consider two additional novel qualitative disclosure characteristics of quarterly earnings press releases, quarter-over-quarter document similarity and within-document diversity. Since the effect of these disclosure characteristics on analysts' information use is not clear ex ante, their examination may lead to insights into analysts' ability to process latent information cues. Moreover, since Loughran and McDonald (2013) have called into question the precise role of readability, we decompose readability into its underlying components to better understand which element of readability affects analysts. Second, the features of analysts' information environments we consider allow us to obtain a more comprehensive picture of how qualitative disclosures affect analysts' information use. For instance, we decompose consensus into its public and private information precision constituents to shed light on the type of information analysts are cultivating from the qualitative aspects of quarterly earnings press releases. Third, our research question asks how qualitative disclosures modify analysts' information environment around quarterly earnings releases. Examining quarterly earnings releases allows us to establish how analysts extract relevant firm information on a more timely basis, which is critical when analysts are updating their assessments of annual earnings (Francis, Schipper, and Vincent 2002).[3]

The paper's findings contribute to the literatures on the function analysts serve in the capital markets and the informativeness of qualitative disclosure. First, we add to research on the value-relevance of analysts as information intermediaries by examining how several novel elements of qualitative disclosure impact analysts' information processing. It is well understood that analysts use quantitative information in their forecasting activities. However, analysts may also be superior processors of qualitative information who, upon receipt of public information signals, develop new and/or unique information and update their forecasts differentially (Barron, Byard, and Kim 2002).

Second, while the literature which examines the value-relevance of analysts' information-processing capabilities finds analysts to be superior in predicting earnings, relative to, say, time-series models (e.g., Brown, Griffin, Hagerman, and Zmijewski 1987), it is not clear whether this superiority is due to analysts' generation of new private information that relies, at least partially, on qualitative information and adjustments. Indeed, a recent paper by Bradshaw, Drake, Myers, and Myers (2012) not only demonstrates renewed interest in this debate but also cautions against staid overgeneralizations within the literature and provides insights into the circumstances under which analysts' forecasts are deemed to be superior. The aforementioned research does not, however, explicitly identify, nor make an attempt to empirically measure, the source of the information from which analysts make their *qualitative* adjustments. We take up this challenge by operationalizing a routine and timely source of qualitative information from which analysts can make their adjustments and test whether or not this form of information influences analysts' information processing.

3. From a practitioner's point of view, in conversations with several Fortune 500 chief accounting officers (preparers) and partners at three of the "Big 4" accounting firms (auditors), the view that the 10-K was untimely and contains substantial boilerplate was widely held. Both groups of individuals expressed that interim disclosures—and notably earnings press releases—are more likely to contain new and timely information. See also Francis et al. (2002) who discuss the increasing usefulness of such releases.

The rest of the paper proceeds as follows. Section 2 lays out our research question in light of prior research. Section 3 details our theoretically motived empirical implementation and the sample employed. Section 4 discusses our empirical evidence prior to concluding in section 5.

2. Research question and predictions

A nascent body of research has developed in recent years which focuses on the effect of qualitative features of disclosure on market participants' behavior. For example, Li (2008) shows that more difficult to read annual reports are associated with lower earnings. Lehavy et al. (2011) find that the level of annual report readability is associated with several properties of analyst forecasts, such as dispersion, accuracy, and uncertainty. Although some of the first papers to examine corporate disclosures through the lens of computational linguistics focused on readability in the context of annual reports, the next wave shifted toward an examination of the tone, or net optimism, contained within earnings press releases. For example, Davis et al. (2012) find that the net optimism of earnings press releases is associated with short-window market responses and future return on assets. While the vast majority of extant research on qualitative disclosure has focused exclusively on readability and tone, little has been done to extend the literature to other features of qualitative disclosure.

Building on the research on qualitative characteristics of disclosure, the central research question we pose is: How do qualitative features of timely, interim corporate disclosures modify sell-side financial analysts' information environment? In doing so, we are implicitly conjecturing a joint hypothesis, that is, (a) the qualitative elements of disclosure we consider are informative and, if so, (b) analysts, purportedly serving in the capacity of skilled information processors, infer meaningful signals from these qualitative disclosures, such that the information contained within the signals confirm and/or update the analysts' priors about the firm. While (a) is a necessary condition, it is not sufficient; if either one of the criteria is not met, we expect to find no relation between the qualitative elements of disclosure and changes in analysts' information environment. However, on the presumption that both criteria are met, we anticipate observing the following, depending upon the qualitative disclosure proxy under consideration.

In the case of readability, we expect qualitative disclosure to be less informative, on average, when an opaque message is conveyed, all else equal. It, therefore, follows that we should observe an increase in total and common uncertainty. Since the predicted change in consensus depends on the relative changes in public and private information precisions, or public and total information (BKLS), we do not provide an ex ante prediction for this information environment proxy. As the earnings press release represents a public disclosure, we predict that higher readability increases the precision of public information. However, similar to consensus, ex ante, we could expect to observe either a decrease or increase in private information precision. Here, the direction of change in private information precision depends on the relation between private and public information. If the two types of information are substitutes, private information precision will likely increase following opaque qualitative disclosure as public information precision decreases. However, prior research in Kim and Verrecchia (1997) and Barron et al. (2002) suggests that private and public information are complements and, as a result, one can expect that opaque qualitative disclosures would result in a decrease in the precision of private information. For this reason, we expect to observe a decrease in both public and private information precision for less readable earnings press releases.

In the case of document similarity, or the degree to which the current earnings press release is textually similar to the preceding release, the anticipated effects are not obvious. On one hand, greater quarter-over-quarter disclosure similarity could imply a less informative current-period disclosure if the information content of that disclosure has been previously disclosed in the prior release. In this case, we would expect there to be no relation between textual similarity and analysts' information environment. On the other hand, greater similarity of the current earnings press release relative to the prior release may serve to resolve analysts' uncertainty regarding the prior release and, as a result, may reinforce that information. If this is the case, we expect a decrease in common and total uncertainty and an increase in both public and private information precision for more similar disclosures. We form no ex ante prediction for consensus for the reasons discussed above.

Finally, we examine lexical diversity, that is, the degree to which the within-document content of the earnings press release possesses an undiluted or intense variety of information types. Prior literature typically proxies for processing costs by way of length or readability and informativeness by way of tone or similarity. The novelty of this proxy is that it intuitively combines both features. That is, when management communicates greater standardized variety in their disclosures, we expect analysts' information environment to improve on average since a greater amount of unique information per unit of processing cost is being disclosed. Therefore, as with similarity, we expect a decrease in common and total uncertainty and an increase in both public and private information precision for more diverse disclosures. As before, we do not make a directional prediction for consensus. A summary of our predictions can be found in Table 1.

While it may seem counterintuitive to have identical directional expectations for similarity and diversity, the key difference is that similarity measures *across-document* commonality through time, whereas diversity measures *within-document* variety at a point in time. As such, a document can simultaneously possess high similarity and high diversity and vice versa. On one hand, *more* similar quarter-over-quarter earnings press releases may confirm analysts' prior beliefs, while *more* diverse point-in-time earnings press releases may provide analysts with greater standardized informational variety upon which to form their beliefs, which would lead to an improvement in analysts' information environment. On the other hand, *less* similar quarter-over-quarter earnings press releases may contain a large amount of new information, while *less* diverse point-in-time earnings press releases may be unnecessarily verbose, which would arguably lead to a negligible improvement in analysts' information environments if analysts cannot sift through and identify value-added information in the disclosure in a timely manner.[4]

3. Empirical implementation and sample selection

Empirical implementation

To examine the relations between qualitative characteristics of disclosure and the changes in firms' information environment around an earnings press release, we employ the following regression model:

$$\begin{aligned} InfoEnvir_{it} = {} & \beta_0 + \beta_1 AbsCAR_{it} + \beta_2 LMktvalue_{it} + \beta_3 LBM_{it} + \beta_4 Badnews_{it} + \beta_5 Guide_{it} \\ & + \beta_6 Readability_{it} + \beta_7 Similarity_{it} + \beta_8 Diversity_{it} + \varepsilon_{it}, \end{aligned} \tag{1}$$

where $InfoEnvir_{it}$ is the change in either overall uncertainty ($Uncert_{it}$), common uncertainty ($CommUncert_{it}$), consensus ($Cons_{it}$), precision of public information (H_{it}), or precision of private information (S_{it}) in the 30 days following, relative to the 90 days preceding, the

4. The above-mentioned qualitative disclosure proxies may reflect a combination of how managers choose to communicate their firms' economic conditions as well as the underlying economic conditions themselves. However, we expect that firms' operations and economic states are primarily conveyed quantitatively and therefore our qualitative measures and results would represent information incremental to the quantitative information conveyed.

TABLE 1
Predictions

Qualitative characteristic	Information environment measure				
	Overall uncertainty	Common uncertainty	Consensus	Precision of public information	Precision of private information
Readability	+	+	+/−	−	−
Similarity	−	−	+/−	+	+
Diversity	−	−	+/−	+	+

Notes:

This table contains a summary of our predictions. + (−) denotes a predicted positive (inverse) relation.

earnings press release in quarter t for firm i, following BKLS as modified by Sheng and Thevenot (2012) and described in detail below.

We include several control variables in our regression specification. To capture the overall informativeness of the press release, we include $AbsCAR_{it}$, which is the absolute value of the five-day abnormal return centered on the earnings announcement date for firm i in quarter t. $LMktvalue_{it}$, which is the logarithmic transformation of market value of equity as of the beginning of the current quarter for firm i in quarter t, controls for size effects, while LBM_{it}, the logarithmic transformation of book-to-market ratio as of the beginning of the current quarter for firm i in quarter t, captures value versus growth firms. The variable $Badnews_{it}$ is an indicator variable equal to one if the five-day abnormal return centered on the earnings announcement is negative, zero otherwise, for firm i in quarter t, to control for the possibility that analysts react differentially to good versus bad news. The last control, $Guide_{it}$, is an indicator variable equal to one if managers provided guidance in the five-day window centered on the earnings announcement, zero otherwise, for firm i in quarter t.

As discussed above, the qualitative characteristics of disclosure we consider are readability, textual similarity, and lexical diversity. $Readability_{it}$ is the calculated Fog Index for the earnings press release, as in Li (2008), using the Lingua::EN::Fathom module in Perl for firm i in quarter t, which can further be decomposed into its underlying components, the percent of complex words (*ComplexWords*) and words per sentence (*WordsSentence*). $Similarity_{it}$ is the textual similarity score between the current and prior quarter's earnings press release, following Brown and Tucker (2010), for firm i in quarter t. Intuitively, the measure represents the degree of overlap between two "buckets" of words after stopwords (e.g., "of," "the," "an") have been removed and the remaining words have been pared down to their stems (e.g., "helpful" becomes "help").[5] It can be interpreted analogously to a correlation coefficient insofar as it is bounded by [0,1] where a higher score reflects greater similarity across documents. $Diversity_{it}$ is the lexical diversity score of the earnings press release and is defined as the number of unique words (or types) divided by the total number of nonunique words (or tokens) for firm i in quarter t. As such, it is commonly referred to as the *type-token ratio* (Goel et al. 2010; Humpherys et al. 2011). To arrive at unique words, as with the calculation of $Similarity_{it}$, stopwords are first removed and the word stems themselves represent the unique word types. As before, the diversity score can

5. Please refer to pages 315–16 of Brown and Tucker (2010) for more technical detail.

be interpreted analogously to a correlation coefficient insofar as it is bounded by [0,1] where a higher score reflects greater standardized diversity (more unique word types per token) within the document.

To obtain empirical proxies for uncertainty, consensus, and the precision of public and private information, we use the BKLS measures, as modified by Sheng and Thevenot (2012). BKLS show that one can express uncertainty, consensus, and the precision of public and private information as follows:

$$U_{it} = SE_{it} + \left(1 - \frac{1}{N}\right) D_{it}, \tag{2}$$

$$\rho_{it} = \frac{SE_{it} - \frac{D_{it}}{N}}{SE_{it} + \left(1 - \frac{1}{N}\right) D_{it}}, \tag{3}$$

$$h_{it} = \frac{SE_{it} - \frac{D_{it}}{N}}{\left[SE_{it} + \left(1 - \frac{1}{N}\right) D_{it}\right]^2}, \tag{4}$$

and

$$s_{it} = \frac{D_{it}}{\left[SE_{it} + \left(1 - \frac{1}{N}\right) D_{it}\right]^2}, \tag{5}$$

where U_{it} is overall uncertainty for firm i in time t, ρ_{it} is consensus for firm i in time t, h_{it} is precision of public information for firm i in time t, s_{it} is precision of private information for firm i in time t, SE_{it} is the expected squared error of the mean forecast for firm i in time t, D_{it} is the expected forecast dispersion for firm i in time t and N is the number of analysts for firm i in time t. The interpretation of the theoretical constructs captured by these variables can be summarized as follows. Overall uncertainty measures analysts' uncertainty about the implications of both their common and private information for predicting future earnings. Common uncertainty is a subset of overall uncertainty as it refers solely to analysts' common information. Consensus captures the portion of analysts' information that is common. As consensus approaches one, analysts rely only on information that is common and their beliefs are identical, whereas when consensus approaches zero, analysts rely entirely on idiosyncratic information. Finally, public and private information precision refer to the quality of the respective information types.

BKLS suggest that one can use the sample variance of individual analyst forecasts and the squared difference between actual earnings and the consensus forecast as proxies for D_{it} and SE_{it} to empirically estimate the constructs in equations (2)–(5). Since it is derived from information available to analysts at the time forecasts are made, observed forecast dispersion is a feasible proxy for its expected counterpart, D_{it}. However, using actual earnings to estimate SE_{it} may lead to spurious inferences especially at long horizons because forecast errors are known to analysts only after the announcement of actual earnings. To alleviate the issues associated with using an ex post estimate of SE_{it}, Sheng and Thevenot (2012) employ a GARCH model to estimate the squared error of the mean forecast using historical data only. Specifically, the method uses the time series of errors in the mean forecast to provide an estimate of their variance, σ^2_{it}. After estimating the GARCH model, one can obtain the conditional variance, $\hat{\sigma}^2_{it}$, which is then used as an estimate of SE_{it} in the expressions above. We calculate the change in the above measures in the 30 days following from the 90 days prior to the earnings press release. In addition, we also examine changes in *common* uncertainty, SE_{it}, which is defined as analysts' uncertainty about common information (BKLS). This type of uncertainty is of interest because an earnings press release represents public information, which, intuitively, should affect primarily analysts' use of public information, that is, analysts' common uncertainty.

Empirically, the inputs to the information environment variables are obtained in the windows before and after a given earnings announcement as follows. Dispersion, D_{it}, is the sample variance of individual analyst forecasts. The conditional variance, $\hat{\sigma}_{it}^2$, which proxies for SE_{it} in equations (2)–(5), is obtained using Sheng and Thevenot's (2012) GARCH procedure. Finally, N_{it} is the number of analysts who revise their forecast in the period following the earnings announcement.

It should be noted that the GARCH approach has known shortcomings discussed in detail in Sheng and Thevenot (2012). First, it assumes that the past is a good predictor of the future and there is always a risk that structural changes to earnings may alter their inherent predictability. Second, the GARCH model estimation requires a long time series of data without missing values, which restricts our sample to large, stable, and heavily followed firms which therefore affects the generalizability of our results to other firms. These limitations, however, are not unique to our approach or setting (see, e.g., Lang 1991; Davis, Fama, and French 2000). Further, as Bradshaw et al. (2012) note, the autoregressive integrated moving average (ARIMA) models commonly used in prior literature often require decades of data to estimate the required time-series parameters.

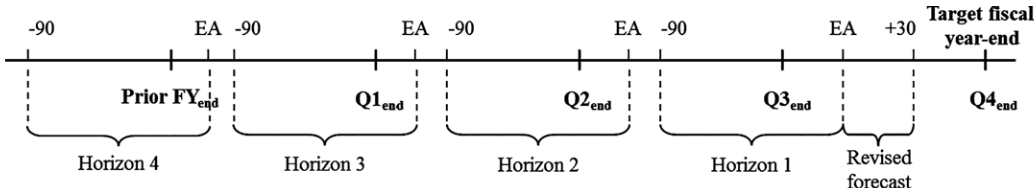
To avoid econometric problems in our regression analysis arising from outlying observations and non-normality of the changes in the information environment variables, following Rees and Thomas (2010), we sort firms with decreases (increases) in those variables into even groups 0–4 (5–9) in each year and horizon. The resulting ranks are divided by 9 to create ranked variables ranging from zero to one. As a result, $rUncert_{it}$, $rCommUncert_{it}$, $rCons_{it}$, rH_{it}, and rS_{it} denote ranked overall uncertainty, ranked common uncertainty, ranked consensus, ranked public information precision, and ranked private information precision, respectively, for firm i in quarter t. In addition, in the cases where quarterly market values or book-to-market ratios are missing, we obtain these variables as of the fiscal year-end immediately prior. We also include both firm and year fixed effects in all model specifications.

Sample selection

The initial sample includes all U.S. firms in the I/B/E/S Detail file for the period 1984–2012. Following Sheng and Thevenot (2012), we require that forecasts are made within 90 days before the earnings announcement and revised forecasts are issued no later than 30 days after the earnings announcement. If an analyst makes more than one forecast in the period before or after the announcement, we include only the forecasts that are closest to the earnings announcement date. We require that there are at least two forecasters in each firm/year/horizon in the 90 days before the earnings announcement and those same analysts revise their forecasts in the 30 days following.

Figure 1 provides a general timeline of events related to our empirical analysis. We split the data based on the length of the forecast horizon and examine forecasts of annual earnings that analysts issue before and after four prior earnings announcements. Horizons 1, 2, and 3 forecasts are made around the earnings announcements of third, second, and first quarter earnings for the current year, respectively, and Horizon 4 forecasts are made around the prior fiscal year's earnings announcement.

Figure 1 Time-line



Description:

This figure describes the time-line of events for our empirical analysis. The change in the information environment variables is measured using forecasts issued in the 30 days following and the 90 days preceding the current quarter's earnings announcement (EA). Horizons 1, 2, and 3 forecasts are issued from 90 days before to 30 days after the third, second, and first quarter's earnings announcement, respectively. Horizon 4 forecasts are made around the prior fiscal year's EA.

Since GARCH estimation requires a fairly lengthy time series of data, we require that sample firms have no missing forecasts or actual values and have data available for at least 24 years in each horizon.[6] The sample is then restricted to years following 2004, as this is the initial year in which earnings press releases were required to be furnished to the SEC. We further eliminate firms for which disclosure characteristics are unavailable. We are then left with a final sample of 1,838 observations for 160 firms. Table 2 provides information about the distribution of our sample observations across horizons (panel A) and the direction of change in the information environment variables (panel B). Panel B of Table 2 shows that uncertainty decreases and public information precision increases following the vast majority of earnings announcements. Private information precision also tends to increase, although not in as many instances as for public information precision.

4. Empirical results

Univariate analysis

Table 3 provides descriptive statistics and univariate results for the overall changes in firms' information environment around earnings press releases and other variables of interest. Panel A shows that *Uncert* and *CommUncert* have negative means and medians, suggesting that uncertainty decreases on average following the disclosure. *Cons*, *H*, and *S*, on the other hand, have positive means and medians, implying increases in these variables on average as a result of the earnings press release. It should be noted that, under certain assumptions, consensus can be expressed as the ratio of public information to total information (BKLS). Therefore, an increase (decrease) in consensus suggests an increase (decrease) in publicly available information relative to idiosyncratic (public) information. The means and medians of all information variables are statistically different from zero, which indicates that there is a significant change in investors' information environment following the disclosure, although the change in consensus is only marginally significant at the 10 percent level.

Panel A further shows that our sample is comprised of large firms that trade at a significant premium above book value. The average earnings press release requires 16.9 years of education to comprehend. In terms of quarter-over-quarter textual similarity, the cur-

6. While this estimation requirement may seem strict, as noted previously, a similar requirement is ordinarily imposed in parameter estimation of structural models found in the finance and accounting literatures (see, e.g., Lang 1991; Davis et al. 2000; Bradshaw et al. 2012). Further, depending upon the particular GARCH-based measure being examined, our sample observations represent roughly 5–10 percent of the population.

TABLE 2
Sample selection

Panel A: Across horizons

	Total	Firms	Horizon (1)	(2)	(3)	(4)
Number of observations with available data to calculate information environment variables after 2004	3,056	183	1,068	775	831	382
Less observations with unavailable disclosure data	1,220	23	408	288	352	170
Final number of observations	1,838	160	660	487	479	212

Panel B: The direction of change in the information environment around earnings releases

Variable	Total	Increase	Percent	Decrease	Percent
Uncertainty	1,838	260	14	1,578	86
Common uncertainty	1,838	274	15	1,564	85
Consensus	1,811	946	52	865	48
Public information precision (*H*)	1,812	1,495	83	317	17
Private information precision (*S*)	1,837	1,258	68	579	32

Notes:

The initial sample includes all U.S. firms in the I/B/E/S Detail file for the period 1984–2012. Following Sheng and Thevenot (2012), we require that analysts make forecasts within 90 days before the earnings announcement and those same analysts revise their forecasts no later than 30 days after the earnings announcement. If an analyst makes more than one forecast in the period before or after the announcement, we include only the forecasts that are closest to the earnings announcement date. We require that there are at least two forecasters in each firm/year/horizon in the 90 days before the earnings announcement. The sample is further restricted to observations after year 2004 because this is when 8-K press releases first became available from the SEC.

Panel A provides the distribution of sample observations across horizons and panel B shows the direction of changes in information environment characteristics. Horizons are defined in Figure 1.

rent quarter's press release is fairly similar to the preceding release with a score of 0.83, where 0 would imply perfect textual dissimilarity and 1 perfect textual similarity. As for lexical diversity, the current quarter's press release reflects an average type-token ratio of 0.346, where 0 would imply perfect homogeneity and 1 perfect heterogeneity.[7]

Panel B presents the percent changes by horizon. The results show that the changes in the uncertainty and precision variables are larger when the horizon is shorter, suggesting that timelier information is more relevant and effective at altering analysts' expectations.

7. Note that this sample average for *Diversity* represents a lower bound, as "stop words" (e.g., "the" "of" "and" "it" "for," etc.) have not been removed from the denominator. The inclusion of such words may not be informative and therefore will inflate the denominator. Analogously, the sample average for *Similarity* represents an upper bound.

TABLE 3
Descriptive statistics and univariate results

Panel A: Descriptive statistics

Variable	N	Mean	Median	SD	$Q1$	$Q3$
Uncert	1,838	−23.408	−37.385	211.548	−56.498	−15.696
CommUncert	1,838	−26.073	−37.534	78.395	−59.215	−14.985
Cons	1,812	12.949	0.192	309.950	−7.776	6.614
H	1,812	178.704	57.627	2579.160	16.171	132.149
S	1,837	425.510	51.048	2902.500	−15.585	194.804
Readability	1,838	16.898	16.668	2.272	15.425	18.211
Similarity	1,049	0.831	0.862	0.128	0.790	0.915
Diversity	1,837	0.344	0.343	0.090	0.282	0.410
ComplexWords	1,838	25.739	25.717	3.064	23.905	27.581
WordsSentence	1,838	16.505	16.103	4.048	13.904	18.813
AbsCAR	1,793	0.050	0.036	0.051	0.016	0.069
Mktvalue	1,838	38,155	14,034	59,596	5,721	40,597
BM	1,838	0.445	0.416	1.510	0.271	0.615

Panel B: Changes in the information environment around earnings announcements by horizon

	Uncertainty	Sign test	Common uncertainty	Sign test	Consensus	Sign test	H	Sign test	S	Sign test
Horizon 4	−19.609	<0.0001	−15.823	<0.0001	0.752	0.011	30.118	<0.0001	10.733	0.149
Horizon 3	−28.707	<0.0001	−28.220	<0.0001	0.352	0.061	40.750	<0.0001	28.808	<0.0001
Horizon 2	−39.602	<0.0001	−39.031	<0.0001	0.494	0.062	66.530	<0.0001	54.544	<0.0001
Horizon 1	−49.605	<0.0001	−55.659	<0.0001	−0.807	0.114	99.623	<0.0001	93.476	<0.0001

Notes:

Panel A presents percentage changes in information environment characteristics and summary statistics for the remaining variables of interest (without a logarithmic transformation for ease of interpretation). All variables are defined in the Appendix.

Panel B presents median percentage changes in information environment characteristics by horizons. The sign test columns present p-values for median tests of differences from zero. All variables are defined in the Appendix. The horizons are defined in Figure 1.

Interestingly, consensus seems to increase at long horizons and decrease in Horizon 1, although the median is not statistically different from zero in Horizon 1.

While the univariate results for uncertainty and precision are not surprising, the increase in consensus following earnings announcements is in contrast to prior results reported in Barron et al. (2002). The authors use BKLS consensus and find that it decreases around announcements in all horizons and argue that this is due to the production of more private information by financial analysts. Barron et al. (2002) report that, while the increase in public information precision exceeds the increase in private information precision, the *percentage* increases in the precision of private information are larger than the percentage increases in the precision of public information, causing consensus to decline around previous earnings announcements. However, we find the opposite result for the percentage changes when we use the GARCH measures; namely, the percentage changes in public information precision exceed those of idiosyncratic information. To investigate the discrepancy in the result further, we replicate the analysis presented in Table 3 in Barron et al. (2002) and calculate changes in consensus using the BKLS and GARCH measures of consensus.[8] The results (untabulated) for BKLS consensus are qualitatively similar to those of Barron et al. (2002) showing a decrease in consensus over all horizons with the decreases statistically different from zero in Horizons 1 and 2. In addition, the decrease in consensus over all horizons is substantial at 28 percent. The results for the GARCH measure of consensus are quite different: consensus increases significantly in Horizon 3 (p-value of 0.03) and the overall decrease in consensus over all horizons is only 5 percent. This casts doubt on Barron et al.'s (2002) results for consensus, especially in the long horizons, that is Horizons 3 and 4, because Sheng and Thevenot (2012) show that the BKLS measures are especially noisy and unstable in long horizons.

Multivariate analysis

Tables 4–6 present our main results. We examine each disclosure element's individual and joint effect on the characteristics of the information environment. Table 4 shows that all of the qualitative disclosure elements we consider (disclosure readability, textual similarity, and lexical diversity) have information content as they are all significantly associated with the change in overall and common uncertainty. The evidence suggests that less readable language in earnings press releases leads to increases in overall (panel A) and common (panel B) uncertainty. In addition, in decomposing the constituents of readability, the relation appears to be driven by longer sentences rather than complex words. Conversely, more textually similar earnings press releases are associated with decreases in uncertainty, suggesting that the similarity of the current disclosure to the prior one likely reinforces previously disclosed news, rather than providing redundant information. More lexically diverse earnings press releases are also found to be associated with decreases in uncertainty, which suggests that more diverse earnings press releases provide analysts with greater standardized informational variety upon which to form their beliefs. Interestingly, this proxy is the most significant, which demonstrates its usefulness to analysts in extracting the maximum amount of information at minimum cost. Based on the results found in the fourth column of panel A, the coefficient estimates suggest that a one standard deviation increase in readability, similarity, and diversity is associated with a 9, 6, and 11 percent larger decrease in overall uncertainty, respectively.[9] Therefore, qualitative disclosure has a significant economic effect on revisions of analysts' expectations, especially when it comes to readability and information parsimony.

8. For the purposes of this analysis, we use the complete I/B/E/S sample, not restricted by the availability of disclosure characteristics, and keep firms with available data in Horizons 1, 2, and 3 in a given year (if we require available data in Horizon 4 as well, the results are qualitatively similar but the sample is much smaller). This provides a maximum of 1,886 observations in each horizon.
9. All other variables are held at their means when computing economic significance.

TABLE 4
Overall and common uncertainty

Panel A: Overall uncertainty

	Expected sign	Dependent variable = *rUncert*				
		(1)	(2)	(3)	(4)	(5)
AbsCAR		−0.236*	−0.114	−0.238**	−0.066	−0.061
		(−1.959)	(−0.726)	(−1.974)	(−0.423)	(−0.390)
LMktvalue		−0.074***	−0.096***	−0.074***	−0.083***	−0.085***
		(−3.738)	(−3.688)	(−3.745)	(−3.203)	(−3.281)
LBM		−0.078***	−0.088***	−0.077***	−0.091***	−0.090***
		(−3.631)	(−3.190)	(−3.581)	(−3.312)	(−3.277)
Badnews		0.012	0.014	0.012	0.014	0.015
		(1.061)	(0.950)	(1.066)	(0.977)	(1.014)
Guide		−0.007	−0.013	−0.005	−0.012	−0.012
		(−0.338)	(−0.414)	(−0.219)	(−0.390)	(−0.399)
Readability	+	0.012***			0.013**	
		(2.948)			(2.261)	
Similarity	−		−0.130*		−0.134**	−0.113*
			(−1.948)		(−2.029)	(−1.672)
Diversity	−			−0.293***	−0.386***	−0.361***
				(−2.977)	(−2.990)	(−2.775)
ComplexWords	+					0.000
						(0.030)
WordsSentence	+					0.009***
						(2.665)
Constant		0.816***	1.297***	1.034***	1.078***	1.097***
		(3.916)	(4.906)	(5.211)	(3.846)	(3.915)
Observations		1,777	1,012	1,776	1,011	1,011
R^2		0.279	0.283	0.279	0.296	0.298

Panel B: Common uncertainty

	Expected sign	Dependent variable = *rCommUncert*				
		(1)	(2)	(3)	(4)	(5)
AbsCAR		−0.125	−0.129	−0.128	−0.088	−0.084
		(−1.043)	(−0.825)	(−1.070)	(−0.564)	(−0.537)
LMktvalue		−0.062***	−0.088***	−0.063***	−0.077***	−0.078***
		(−3.172)	(−3.380)	(−3.186)	(−2.950)	(−3.009)
LBM		−0.078***	−0.087***	−0.077***	−0.089***	−0.089***
		(−3.664)	(−3.153)	(−3.615)	(−3.254)	(−3.225)
Badnews		0.011	0.021	0.011	0.021	0.022
		(0.958)	(1.450)	(0.952)	(1.474)	(1.502)
Guide		0.010	0.001	0.013	0.001	0.001
		(0.497)	(0.025)	(0.603)	(0.049)	(0.042)
Readability	+	0.011***			0.010*	
		(2.783)			(1.729)	
Similarity	−		−0.157**		−0.162**	−0.145**
			(−2.361)		(−2.440)	(−2.143)

(The table is continued on the next page.)

TABLE 4 (continued)

Panel B: Common uncertainty

	Expected sign	Dependent variable = *rCommUncert*				
		(1)	(2)	(3)	(4)	(5)
Diversity	–			−0.252**	−0.371***	−0.351***
				(−2.579)	(−2.876)	(−2.702)
ComplexWords	+					−0.000
						(−0.000)
WordsSentence	+					0.007**
						(2.057)
Constant		0.719***	1.239***	0.922***	1.073***	1.088***
		(3.468)	(4.701)	(4.667)	(3.832)	(3.883)
Observations		1,777	1,012	1,776	1,011	1,011
R^2		0.303	0.333	0.302	0.342	0.343

Notes:

This table provides regression results for ranked change in overall uncertainty (panel A) and common uncertainty (panel B). All variables are defined in the Appendix. All models include year and firm fixed effects. ***, **, * denotes significance at the 1 percent, 5 percent, and 10 percent levels, respectively. *t*-statistics are provided in parentheses.

Table 5 provides the results for consensus, which can be interpreted as the ratio of common to total information. Hence, a decrease in consensus suggests production of more private relative to public information, while an increase implies that analysts produce more commonly held information. Interestingly, only textual similarity is statistically significant in this specification and the evidence indicates that more similar press releases are associated with decreases in consensus, that is, analysts use the reinforced disclosures to produce more private relative to public information. The strength of this relation suggests that the "newness" of the current disclosure's information relative to the prior disclosure plays a particularly important role in analysts' generation of private information, rather than decreasing their uncertainty. Since the other disclosure elements appear to be unrelated to consensus, we next examine the underlying theoretical components of consensus. Specifically, under certain assumptions provided in BKLS, consensus is equal to the ratio of public information precision to the precision of public plus the precision of private information, and to the extent that readability and diversity affect both public and private information precision, their effect on consensus will be obfuscated. The results are provided in Table 6.

Panel A of Table 6 shows that more readable and diverse disclosure is associated with a larger increase in the precision of public information, whereas similarity does not seem to affect public information precision. Examining the economic significance of this result suggests that a one standard deviation increase in readability and diversity is associated with approximately a 4 percent higher increase in the precision of public information. Further, panel B shows that all qualitative disclosure elements are related to the precision of private information. Less readable disclosures are associated with decreases in private information precision, while textually similar and lexically diverse disclosures are associated with increases in the precision of private information. In addition, in decomposing the constituents of readability, the relation appears to be dri-

TABLE 5
Consensus

	Expected sign	Dependent variable = *rCons*				
		(1)	(2)	(3)	(4)	(5)
AbsCAR		−0.053	−0.272	−0.066	−0.266	−0.268
		(−0.314)	(−1.200)	(−0.389)	(−1.167)	(−1.175)
LMktvalue		−0.038	−0.051	−0.039	−0.050	−0.049
		(−1.355)	(−1.351)	(−1.421)	(−1.307)	(−1.290)
LBM		−0.061**	−0.053	−0.059**	−0.054	−0.055
		(−2.028)	(−1.339)	(−1.969)	(−1.360)	(−1.366)
Badnews		0.034**	0.056***	0.034**	0.056***	0.056***
		(2.176)	(2.668)	(2.161)	(2.651)	(2.643)
Guide		0.070**	0.097**	0.070**	0.097**	0.097**
		(2.346)	(2.161)	(2.370)	(2.158)	(2.158)
Readability	±	0.008			0.006	
		(1.371)			(0.779)	
Similarity	±		−0.243**		−0.241**	−0.248**
			(−2.508)		(−2.488)	(−2.503)
Diversity	±			0.040	0.066	0.058
				(0.289)	(0.353)	(0.307)
ComplexWords	±					0.004
						(0.712)
WordsSentence	±					0.001
						(0.291)
Constant		0.899***	1.274***	1.017***	1.163***	1.157***
		(3.071)	(3.324)	(3.643)	(2.841)	(2.823)
Observations		1,750	995	1,749	994	994
R^2		0.143	0.152	0.141	0.152	0.152

Notes:

This table provides regression results for ranked change in consensus. All variables are defined in the Appendix. All models include year and firm fixed effects. ***, **, * denotes significance at the 1 percent, 5 percent levels , and 10 percent levels, respectively. *t*-statistics are provided in parentheses.

ven by longer sentences rather than complex words. Based on the results found in the fourth column of panel B, the coefficient estimates suggest that a one standard deviation increase in readability, similarity, and diversity is associated with a 7, 6, and 5 percent higher increase in private information precision, respectively. This provides an explanation for why only textual similarity has a significant impact on consensus—it affects primarily private information generation, while more readable and textually diverse disclosure prompt similar improvements in both public and private information.[10]

Overall, our main results suggest that sell-side equity analysts, serving in the capacity of skilled information processors, infer meaningful signals from the qualitative elements of disclosure we consider, such that the information contained within the signals simultaneously confirm and update analysts' priors about the firm. As such, our analy-

10. All results are robust to including additional controls for institutional ownership, the number of management forecasts, and the standard deviation of returns.

TABLE 6
Precision of public and private information

Panel A: Precision of public information

	Expected sign	Dependent variable = *rH*				
		(1)	(2)	(3)	(4)	(5)
AbsCAR		0.107	−0.109	0.112	−0.157	−0.161
		(0.803)	(−0.608)	(0.845)	(−0.884)	(−0.903)
LMktvalue		0.058***	0.074**	0.058***	0.063**	0.064**
		(2.663)	(2.501)	(2.679)	(2.112)	(2.142)
LBM		0.052**	0.069**	0.051**	0.072**	0.071**
		(2.221)	(2.186)	(2.168)	(2.291)	(2.273)
Badnews		−0.003	−0.003	−0.003	−0.003	−0.003
		(−0.208)	(−0.173)	(−0.219)	(−0.192)	(−0.207)
Guide		0.041*	0.053	0.038*	0.052	0.052
		(1.779)	(1.505)	(1.649)	(1.477)	(1.480)
Readability	−	−0.013***			−0.014**	
		(−2.919)			(−2.249)	
Similarity	+		0.044		0.044	0.033
			(0.578)		(0.586)	(0.425)
Diversity	+			0.290***	0.351**	0.336**
				(2.689)	(2.389)	(2.269)
ComplexWords	−					−0.003
						(−0.652)
WordsSentence	−					−0.008**
						(−2.088)
Constant		0.426*	−0.008	0.191	0.233	0.223
		(1.860)	(−0.027)	(0.876)	(0.728)	(0.695)
Observations		1,751	995	1,750	994	994
R^2		0.218	0.182	0.218	0.194	0.194

Panel B: Precision of private information

	Expected sign	Dependent variable = *rS*				
		(1)	(2)	(3)	(4)	(5)
AbsCAR		0.281*	0.331*	0.293*	0.286	0.282
		(1.869)	(1.774)	(1.945)	(1.533)	(1.509)
LMktvalue		0.063**	0.104***	0.065***	0.092***	0.094***
		(2.560)	(3.359)	(2.632)	(2.960)	(3.013)
LBM		0.074***	0.101***	0.072***	0.104***	0.104***
		(2.771)	(3.060)	(2.681)	(3.169)	(3.142)
Badnews		−0.010	−0.022	−0.009	−0.022	−0.022
		(−0.692)	(−1.245)	(−0.678)	(−1.253)	(−1.277)
Guide		−0.047*	−0.076**	−0.049*	−0.077**	−0.077**
		(−1.786)	(−2.077)	(−1.876)	(−2.106)	(−2.100)
Readability	−	−0.016***			−0.016**	
		(−3.035)			(−2.347)	
Similarity	+		0.205**		0.208***	0.190**
			(2.567)		(2.617)	(2.341)

(The table is continued on the next page.)

TABLE 6 (continued)

Panel B: Precision of private information

	Expected sign	Dependent variable = *rS*				
		(1)	(2)	(3)	(4)	(5)
Diversity	+			0.173	0.288*	0.267*
				(1.412)	(1.862)	(1.712)
ComplexWords	−					−0.002
						(−0.455)
WordsSentence	−					−0.009**
						(−2.385)
Constant		0.184	−0.570*	−0.077	−0.297	−0.313
		(0.707)	(−1.807)	(−0.311)	(−0.884)	(−0.933)
Observations		1,776	1,012	1,775	1,011	1,011
R^2		0.231	0.266	0.227	0.274	0.275

Notes:

This table provides regression results for ranked change in public information precision (panel A) and private information precision (panel B). All variables are defined in the Appendix. All models include year and firm fixed effects. ***, **, * denotes significance at the 1 percent, 5 percent, and 10 percent levels, respectively. *t*-statistics are provided in parentheses.

sis demonstrates the value-relevance of both analysts as information intermediaries and latent qualitative disclosure by showing how several novel elements of qualitative disclosure impact analysts' information environment. Analysts appear to be superior processors of qualitative information who, upon receipt of public information signals, develop new and/or unique information and update their forecasts differentially. The superiority is ostensibly due to analysts' reliance, at least partially, on qualitative information and adjustments, which, in turn, allows them to generate private information in a parsimonious manner.

Cross-sectional analysis

We have established that several qualitative elements of disclosure are informative as they relate to changes in analysts' information environment. In this section, we examine whether these findings vary across firms. The ability of analysts to understand the qualitative features of disclosure may differ by, or analysts may be attracted to, firms with a rich disclosure environment, especially if there is an opportunity to develop idiosyncratic information. For example, firms with poor information environments may benefit more from providing additional information to investors. However, as our previous results show, the qualitative elements of disclosure provide an opportunity to develop new private information, which may exacerbate existing information asymmetry.

To examine these questions, we consider how the effect of firm size and analyst following impact our primary results. Both of these variables have been found to be related to firms' information environment (Bhushan 1989; O'Brien and Bhushan 1990; Lang and Lundholm 1996; Irvine 2001). These studies generally suggest that (a) large firms operate in a rich information environment and (b) analysts have greater incentives to generate idiosyncratic information about these firms, which leads to increases in analyst following.

TABLE 7
The effect of firm size and analyst following

Panel A: Firm size

	Dependent variable				
	rUncert (1)	*rCommUncert* (2)	*rCons* (3)	*rH* (4)	*rS* (5)
AbsCAR	−0.041	−0.079	−0.276	−0.186	0.283
	(−0.260)	(−0.506)	(−1.206)	(−1.045)	(1.513)
LMktvalue	−0.087***	−0.082***	−0.054	0.061*	0.096***
	(−3.170)	(−2.996)	(−1.345)	(1.946)	(2.915)
LBM	−0.096***	−0.092***	−0.055	0.077**	0.108***
	(−3.477)	(−3.339)	(−1.380)	(2.446)	(3.267)
Badnews	0.014	0.021	0.055***	−0.004	−0.021
	(0.993)	(1.484)	(2.612)	(−0.214)	(−1.222)
Guide	−0.012	0.001	0.095**	0.051	−0.075**
	(−0.398)	(0.024)	(2.115)	(1.458)	(−2.055)
Readability	0.023***	0.018**	0.008	−0.023***	−0.022***
	(3.297)	(2.556)	(0.803)	(−2.867)	(−2.609)
Similarity	−0.111	−0.188**	−0.271**	−0.010	0.218**
	(−1.314)	(−2.236)	(−2.211)	(−0.108)	(2.155)
Diversity	−0.496***	−0.516***	−0.118	0.380**	0.503**
	(−3.003)	(−3.121)	(−0.489)	(2.017)	(2.535)
Small × *Readability*	−0.023**	−0.016*	−0.001	0.022**	0.012
	(−2.524)	(−1.740)	(−0.084)	(2.045)	(1.073)
Small × *Similarity*	−0.093	0.055	0.095	0.186	−0.023
	(−0.717)	(0.422)	(0.502)	(1.258)	(−0.146)
Small × *Diversity*	0.192	0.311	0.432	0.003	−0.469*
	(0.810)	(1.311)	(1.247)	(0.012)	(−1.651)
Small	0.420*	0.123	−0.207	−0.550**	−0.036
	(1.929)	(0.563)	(−0.646)	(−2.197)	(−0.139)
Constant	0.940***	1.055***	1.261***	0.451	−0.299
	(3.106)	(3.483)	(2.841)	(1.304)	(−0.822)
Observations	1,011	1,011	994	994	1,011
R^2	0.303	0.346	0.154	0.199	0.278

Panel B: Analyst following

	Dependent variable				
	rUncert (1)	*rCommUncert* (2)	*rCons* (3)	*rH* (4)	*rS* (5)
AbsCAR	−0.066	−0.045	−0.150	−0.123	0.244
	(−0.420)	(−0.290)	(−0.638)	(−0.671)	(1.276)
LMktvalue	−0.077***	−0.073***	−0.059	0.060*	0.087***
	(−2.896)	(−2.777)	(−1.498)	(1.941)	(2.715)
LBM	−0.070**	−0.072**	−0.058	0.055	0.086**
	(−2.333)	(−2.399)	(−1.296)	(1.576)	(2.369)
Badnews	0.008	0.017	0.061***	0.002	−0.020
	(0.538)	(1.171)	(2.771)	(0.132)	(−1.128)

(The table is continued on the next page.)

TABLE 7 (continued)

Panel B: Analyst following

	Dependent variable				
	rUncert (1)	*rCommUncert* (2)	*rCons* (3)	*rH* (4)	*rS* (5)
Guide	0.005	0.017	0.091**	0.032	−0.085**
	(0.171)	(0.542)	(1.984)	(0.902)	(−2.283)
Readability	0.018***	0.015**	0.012	−0.019**	−0.016**
	(2.791)	(2.228)	(1.274)	(−2.481)	(−2.065)
Similarity	−0.092	−0.132*	−0.285**	−0.058	0.223**
	(−1.172)	(−1.703)	(−2.468)	(−0.639)	(2.360)
Diversity	−0.584***	−0.540***	0.097	0.533***	0.386**
	(−3.864)	(−3.602)	(0.433)	(3.057)	(2.113)
LowF × *Readability*	−0.008	−0.004	−0.003	0.007	−0.004
	(−1.060)	(−0.517)	(−0.232)	(0.797)	(−0.441)
LowF × *Similarity*	−0.046	−0.016	0.160	0.217	−0.075
	(−0.361)	(−0.125)	(0.837)	(1.453)	(−0.482)
LowF × *Diversity*	0.400**	0.357*	−0.201	−0.416*	−0.097
	(2.109)	(1.896)	(−0.718)	(−1.903)	(−0.423)
LowF	0.066	0.009	0.057	−0.148	0.109
	(0.361)	(0.051)	(0.210)	(−0.696)	(0.490)
Constant	0.960***	0.989***	1.148***	0.335	−0.270
	(3.269)	(3.396)	(2.636)	(0.985)	(−0.761)
Observations	947	947	931	931	947
R^2	0.323	0.371	0.163	0.212	0.294

Notes:

This table provides cross-sectional regression results for ranked changes of all information environment variables. Panel A examines firm size, whereas panel B examines analyst following. All variables are defined in the Appendix. All models include year and firm fixed effects. ***, **, * denotes significance at the 1 percent, 5 percent, and 10 percent levels, respectively. *t*-statistics are provided in parentheses.

This implies that the relations documented in our primary analysis between qualitative elements of disclosure and analysts' information environment are likely to be stronger in large and more heavily followed firms.

We construct two additional variables: (a) *Small*, equal to one if the firm's market value of equity is below the sample median, zero otherwise and (b) *LowF*, equal to one if the firm's analyst following is below the sample median, zero otherwise. We reestimate our original regressions to include interactions between these variables and our proxies for the qualitative characteristics of disclosure. The results are provided in Table 7.

Panel A shows that the effects of the qualitative elements of disclosure on analysts' information environment are particularly strong for large firms. For example, disclosure readability is highly statistically significant in the uncertainty and precision models but the interactions with small firms are in the opposite direction, counteracting the effects. More formally, *F*-tests (not tabulated) show that none of the coefficients on the qualitative disclosure variables for small firms are statistically different from zero. Panel B provides the results when interactions with low analyst following are included. As before, the effects of the qualitative elements of disclosure on analysts' information environment are particularly strong for heavily followed firms. However, *F*-tests (not tabulated) suggest that lexical diversity has a significant effect on uncertainty and public information precision for both sets of firms. Consistent with our prior evidence, the results also show that textual similarity affects primarily private information precision, while lexical diversity is strongly associated with both public and private information precision. Overall, the cross-sectional results in Table 7 suggest that our main results are driven by large and heavily followed firms, which is consistent with Bhushan (1989) and Irvine (2001) who suggest that analysts reap higher rewards from obtaining an information advantage about such firms.

5. Conclusion

We examine a routine and timely disclosure, earnings press releases, to determine the extent to which several novel qualitative elements of such disclosures are associated with changes in sell-side financial analysts' information environment. Using a comprehensive set of GARCH-based proxies (Sheng and Thevenot 2012), we examine how disclosure readability's components, across-document textual similarity, and within-document lexical diversity alter analysts' information environment. On one hand, if analysts are mere conduits for publicly available, *quantitative* information, there should be no relation between changes in the analysts' information environment around earnings press releases and the qualitative elements of disclosure. On the other, if analysts add value to their forecasts through their information-processing abilities and consequent *qualitative* adjustments, we expect the qualitative elements to impact analysts' information use. Our results are consistent with the latter conjecture.

We find that the qualitative elements of disclosure we consider provide analysts with embedded information cues, as they affect analysts' information environment in a predictable manner. First, consistent with concerns raised by Loughran and McDonald (2013), we find that less readable disclosure, in the form of longer sentences, is associated with increases in overall and common uncertainty and decreases in public and private information precision. Second, we further provide evidence that quarter-over-quarter textual similarity is associated with decreases in overall and common uncertainty and increases in private information precision. Third, lexical diversity is found to be associated with decreases in overall and common uncertainty and increases in public and private information precision. Finally, and consistent with large firms operating in rich information environments and analysts being attracted to firms with such environments, our results are stronger for large and heavily followed firms. Alternatively, analysts of such firms may be of higher ability to discern the more subtle, qualitative features of disclosure. We believe this may be a fruitful question to examine in future research.

Collectively, the results suggest that analysts use qualitative disclosure cues in earnings press releases to update their expectations and generate both public and private information. These findings should be of interest to analysts who may wish to explore the latent information embedded within the qualitative elements of disclosure, regulators who direct the form and content of disclosure, and academics who study the use (and possible misuse) of various forms of information and its presentation. Future research may examine whether qualitative disclosure leads to increases or decreases in information asymmetry or whether there is ambiguity in how analysts interpret such disclosure.

Appendix

Variable definitions

Uncert	Uncertainty in the 30 days following the earnings announcement (EA) less uncertainty in the 90 days prior to the EA, calculated using the GARCH approach of Sheng and Thevenot (2012). *rUncert* denotes ranked *Uncert*, where firms with negative (positive) *Uncert* are sorted evenly into groups zero to four (5–9). The resulting ranks are divided by 9 to create a ranked variable that ranges from zero to one
CommUncert	Common uncertainty in the 30 days following the EA less common uncertainty in the 90 days prior to the EA, calculated using the GARCH approach of Sheng and Thevenot (2012). *rCommUncert* denotes ranked *CommUncert*, where firms with negative (positive) *CommUncert* are sorted evenly into groups zero to four (5–9). The resulting ranks are divided by 9 to create a ranked variable that ranges from zero to one
Cons	Consensus in the 30 days following the EA less consensus in the 90 days prior to the EA, calculated using the GARCH approach of Sheng and Thevenot (2012). *rCons* denotes ranked *Cons*, where firms with negative (positive) *Cons* are sorted evenly into groups zero to four (5–9). The resulting ranks are divided by 9 to create a ranked variable that ranges from zero to one
H	Public information precision in the 30 days following the EA less public information precision in the 90 days prior to the EA, calculated using the GARCH approach of Sheng and Thevenot (2012). *rH* denotes ranked *H*, where firms with negative (positive) *H* are sorted evenly into groups zero to four (5–9). The resulting ranks are divided by 9 to create a ranked variable that ranges from zero to one
S	Private information precision in the 30 days following the EA less private information precision in the 90 days prior to the EA, calculated using the GARCH approach of Sheng and Thevenot (2012). *rS* denotes ranked *S*, where firms with negative (positive) *S* are sorted evenly into groups zero to four (5–9). The resulting ranks are divided by 9 to create a ranked variable that ranges from zero to one
AbsCAR	Absolute value of the five-day abnormal return centered on the EA
Mktvalue	Market value of equity at the beginning of the current quarter. If unavailable, *Mktvalue* is calculated at the fiscal year-end immediately prior. *LMktvalue* indicates the logarithm transformation of *Mktvalue*
BM	Book-to-market ratio at the beginning of the current quarter. If unavailable, *BM* is calculated at the fiscal year-end immediately prior. *LBM* indicates the logarithm transformation of *BM*
Badnews	One if the five-day abnormal return centered on the EA is negative, zero otherwise
Guide	One if managers provided guidance in the 90 days before or the 30 days after the earnings announcement, zero otherwise
Readability	The Fog index, expressed as the number of years of education required to comprehend the press release and computed as .4[(words/sentences) +100 (three or more syllable words/words)]
Diversity	The number of unique words (or types) divided by the total number of words (or tokens) in the earnings press release
Similarity	The degree of textual similarity between the current and preceding earnings press release, following Brown and Tucker (2010)
Small	Equal to one if the firm's market value of equity is below the sample median, zero otherwise
ComplexWords	Number of three or more syllable words divided by the total number of words
WordsSentence	Total number of words divided by the total number of sentences
LowF	Equal to one if the firm's analyst following is below the sample median, zero otherwise

References

Barron, O. E., O. Kim, S. C. Lim, and D. E. Stevens. 1998. Using analysts' forecasts to measure properties of analysts' information environment. *The Accounting Review* 73(4): 421–33.

Barron, O. E., D. Byard, and O. Kim. 2002. Changes in analysts' information around earnings announcements. *The Accounting Review* 4(4): 821–46.

Bhushan, R. 1989. Firm characteristics and analyst following. *Journal of Accounting & Economics* 11 (2-3): 255–74.

Bloomfield, R. 2002. The 'Incomplete Revelation Hypothesis' and financial reporting. *Accounting Horizons* 16(3): 233–43.

Bradshaw, M., M. Drake, J. Myers, and L. Myers. 2012. A re-examination of analysts' superiority over time-series forecasts of annual earnings. *Review of Accounting Studies* 17(4): 944–68.

Brown, S. V., and J. W. Tucker. 2010. Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research* 49(2): 309–46.

Brown, L., P. Griffin, R. Hagerman, and M. Zmijewski. 1987. Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings. *Journal of Accounting and Economics* 9(1): 61–87.

Brown, L., A. Call, M. Clement, and N. Sharp. 2014. Inside the "Black Box" of Sell-Side Financial Analysts. *Journal of Accounting Research*, forthcoming.

Davis, J. L., E. F. Fama, and K. R. French. 2000. Characteristics, covariances, and average returns: 1929 to 1997. *Journal of Finance* 1(1): 389–406.

Davis, A. K., J. M. Piger, and L. M. Sedor. 2012. Beyond the numbers: measuring the information content of earnings press release language. *Contemporary Accounting Research* 29(3): 845–68.

Francis, J., K. Schipper, and L. Vincent. 2002. Expanded disclosures and the increased usefulness of earnings announcements. *The Accounting Review* 77(3): 515–46.

Goel, S., J. Gangolly, S. R. Faerman, and O. Uzuner. 2010. Can linguistic predictors detect fraudulent financial filings? *Journal of Emerging Technologies in Accounting* 7: 25–46.

Humpherys, S. L., K. C. Moffitt, M. B. Burns, J. K. Burgoon, and W. F. Felix. 2011. Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems* 50(3): 585–94.

Irvine, P. J. A. 2000. Do analysts generate trade for their firms? Evidence from the Toronto stock exchange. *Journal of Accounting and Economics* 30(2): 209–26.

Kim, O., and R. E. Verrecchia. 1997. Pre-announcement and event-period private information. *Journal of Accounting and Economics* 24: 395–421.

Lang, M. H. 1991. Time-varying stock price response to earnings induced by uncertainty about the time-series process of earnings. *Journal of Accounting Research* 29(2): 229–57.

Lang, M. H., and R. J. Lundholm. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* 71(4): 467–92.

Lehavy, R., F. Li, and K. Merkley. 2011. The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review* 86(3): 1087–115.

Levitt, A. 1997. *Corporate Finance in the Information Age*. Remarks delivered to the Securities Regulation Institute. San Diego, California, January 23, 1997.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45(2-3): 221–47.

Loughran, T., and B. McDonald. 2014. Measuring readability in financial disclosures. *Journal of Finance* 69(4): 1643–71.

O'Brien, P. C., and R. Bhushan. 1990. Analyst following and institutional ownership. *Journal of Accounting Research* 28(3): 55–82.

Rees, L., and W. Thomas. 2010. The stock price effects of changes in dispersion of investor beliefs during earnings announcements. *Review of Accounting Studies* 15(1): 1–31.

Sheng, X., and M. Thevenot. 2012. A new measure of earnings forecast uncertainty. *Journal of Accounting and Economics* 53(1-2): 21–33.