

# How Does Readability Influence Investors' Judgments? Consistency of Benchmark Performance Matters

*Hun-Tong Tan*

*Nanyang Technological University*

*Elaine Ying Wang*

*University of Massachusetts Amherst*

*Bo Zhou*

*Shanghai University of Finance and Economics*

**ABSTRACT:** We conduct two experiments to investigate how readability (high versus low) and benchmark performance consistency (consistent versus inconsistent) influence investors' judgments. Using prior management guidance and year-ago quarter performance as two benchmarks against which to assess actual earnings performance, we manipulate whether the valence of guidance performance (positive or negative) and the valence of trend performance (positive or negative) are consistent with each other. We also manipulate the readability of trend performance in our main experiment. Our results show that when benchmark performance is inconsistent, higher as opposed to lower readability of positive (negative) trend performance leads to more (less) favorable investors' performance judgments. This effect of readability is smaller when benchmark performance is consistent. We also show that higher readability in the inconsistent benchmark performance condition improves investors' understanding of the firm's current-quarter performance, which in turn influences their judgments on the firm's future performance. In a supplementary experiment, we manipulate the readability of guidance performance in an inconsistent benchmark performance setting, and replicate the key finding that higher readability of positive guidance performance leads to more positive judgment on the firm's future performance.

**Keywords:** *readability; benchmark performance consistency; understanding; processing fluency; performance judgment.*

---

We appreciate helpful comments from Lisa Koonce (editor), two anonymous reviewers, John Harry Evans III (senior editor), Bob Libby, and workshop participants at Nanyang Technological University, Cornell University, and Peking University. We are grateful to Nanyang Technological University, University of Massachusetts Amherst, the Program for Innovative Research Team of Shanghai University of Finance and Economics, the National Natural Science Foundation of China (Project number 71402089), and the UOB Endowed Chair for financial support.

Editor's note: Accepted by Lisa Koonce.

*Submitted: October 2011*

*Accepted: June 2014*

*Published Online: July 2014*

## I. INTRODUCTION

Managers often compare their firms' current-period performance against different benchmarks such as year-ago quarter earnings, analysts' consensus forecast, or prior guidance in their earnings press releases (Graham, Harvey, and Rajgopal 2005). While they prefer that such comparisons to lead to favorable evaluations of their firms (Schrand and Walther 2000; Krische 2005), a firm's current-period performance relative to different benchmarks can often conflict, depending on which benchmark is used (Rees 2005). As a result, a firm's benchmark performance can vary in terms of consistency. For instance, in a sample of 52,123 firm-year observations analyzed in Rees (2005), 19,222 observations (36.9 percent) either beat analysts' consensus forecast, but report negative earnings change, or miss analysts' consensus forecast, but report positive earnings change. In the presence of such inconsistencies, managers have incentives to strategically vary the readability—the ease with which a text can be read and understood (Dale and Chall 1949)—of selected information in order to portray the firm in the most favorable light (Courtis 1998; Li 2008). Managers can improve the readability of content related to the benchmark with positive implications, and/or obfuscate content of the benchmark with negative implications. In turn, such actions can affect how investors react to managers' disclosures.

Investigating this issue is important for several reasons. First, anecdotal evidence reveals that firms with inconsistent benchmark performance make the positive benchmark performance more readable than the negative benchmark performance. For example, United Airlines' 2012 fourth-quarter earnings release reports an overall operating loss, but highlights the good news in arguably easier-to-read bullet points. Similarly, Eli Lilly's 2013 third-quarter earnings release lists positive facts about the company in easier-to-read bullet points, but discusses the decrease in net income and earnings per share (EPS) using what appears to be more difficult-to-read language; see Appendix A. This evidence is consistent with regulators' concern with managers selectively emphasizing information in public disclosures as favorably as possible (Pozen 2008; Koonce, Seybert, and Smith 2013). Second, given this evidence, it is possible that investors' welfare may be adversely affected to the extent that their ability to fully understand the implications of the negative performance information is reduced. If managers make only certain measures that reflect negatively on the firm less readable, will investors ignore or make little use of that negative information? Alternatively, if managers make only favorable performance measures more readable, will the positive information have a greater impact on investors' judgments? Extant research investigates the effects of readability on investors' reactions in settings where readability varies for the entire disclosure (You and Zhang 2009; Miller 2010; Rennekamp 2012), but does not directly shed light on these further issues. In particular, existing literature does not examine settings where readability varies only for some aspects of the disclosure, such as selected benchmark performance, and how readability effects are moderated when benchmark performance varies in consistency.

Comprehension theory indicates that coherence or consistency is a key message attribute that people attend to, and influences how people process the message (Kintsch and Van Dijk 1978). In the absence of inconsistencies, messages are relatively easier to understand (Albrecht and O'Brien 1993), irrespective of variations in the readability of some selected content. On the other hand, in the presence of inconsistencies, messages become complicated and difficult to understand. In such cases, high (low) readability of some selected content is more likely to improve (impair) individuals' understanding, which in turn influences their judgments (Burgoon 1975; Masson and Waldron 1994). In our setting with inconsistent benchmark performance, if the selected benchmark performance that has positive implications is made more (less) readable, then investors' judgments are correspondingly more (less) positive, and *vice versa* if the benchmark performance has negative implications.

We conduct experiments using M.B.A. students as proxies for investors to examine how the readability of selected benchmark performance measures and benchmark performance consistency jointly influence investors' judgments on the firm's future performance, hereafter, "performance

judgment.” We consider two performance benchmarks: (1) performance relative to previously issued management guidance, hereafter, “guidance performance,” and (2) performance compared to that in the same quarter one year ago, hereafter, “trend performance.” Research shows that managers generally compare current-quarter performance with that in the year-ago quarter, as well as previously issued management guidance, if available.<sup>1</sup> While managers manage earnings to avoid negative earnings trend performance (Degeorge, Patel, and Zeckhauser 1999), missing management’s own guidance is also perceived as a negative signal (Koch, Lefanowicz, and Shane 2012; Lee, Matsunaga, and Park 2012).

Our main experiment manipulates the readability of trend performance (high versus low), trend performance valence (positive versus negative), and benchmark performance consistency (inconsistent versus consistent). Benchmark performance consistency is determined by both trend performance valence and guidance performance valence. Specifically, in the positive (negative) trend performance condition, benchmark performance is consistent if guidance performance is also positive (negative), but inconsistent otherwise. We find that the readability of trend performance has a greater impact on participants’ performance judgments when benchmark performance is inconsistent than when it is consistent. When benchmark performance is inconsistent, higher readability of trend performance leads to higher performance judgment when trend performance is positive, but lower performance judgment when trend performance is negative. The readability of trend performance does not have an effect when benchmark performance is consistent.

To investigate the mechanism through which this readability effect occurs in the inconsistent benchmark performance condition, we examine two potential mediators. The first is processing fluency, an individual’s subjective feeling concerning the ease of processing information (Winkielman, Huber, Kavanagh, and Schwarz 2012). The second is understanding, an individual’s mental grasp of the meaning of the text content (Rumelhart 1984; Miele and Molden 2010). While improved understanding can help investors make more informed judgments (Securities and Exchange Commission [SEC] 1998; Bloomfield 2002), prior research also documents that processing fluency influences investors’ reliance on the disclosure (Shah and Oppenheimer 2007; Rennekamp 2012). Our results show that when benchmark performance is inconsistent, understanding rather than processing fluency mediates the effect of readability on performance judgment.

To test the generalizability on our findings from the main experiment, we conduct a supplementary experiment. We manipulate the readability of guidance performance, as opposed to trend performance in the main experiment, in an inconsistent benchmark performance setting where guidance performance is positive and trend performance is negative. We again find that higher readability of a positive benchmark performance, reflecting positive guidance performance, leads to higher performance judgment.

We extend the literature on readability by examining its effect on investor judgment in a context where variations in readability are selectively applied to parts of the disclosure, and where the disclosure contains either a consistent or inconsistent message. Prior studies show that investors’ reactions to bad news will be muted if the entire disclosure is difficult to read (You and Zhang 2009; Miller 2010; Rennekamp 2012).<sup>2</sup> We show that in the presence of inconsistent benchmark performance, investors’ reactions to unfavorable benchmark performance can be dampened by making only part of the disclosure containing bad news less readable. However, such strategic use of low readability in selected contents does not work when benchmark performance is consistent.

---

<sup>1</sup> Among listed firms, 40 percent issue either quarterly or annual guidance (Acito 2013; Call, Chen, Miao, and Tong 2014).

<sup>2</sup> In a recent study, Tan, Wang, and Zhou (2014) examine how the effect of tone is moderated by readability and investor sophistication. They vary tone in the earlier part of an earnings release, and readability in the later part of the earnings release. They find that the effect of tone on investors’ earnings judgments is magnified when readability is low versus when it is high, but that the directional effect of this interaction switches depending on investor sophistication.

We also contribute to the literature by identifying “understanding” as a mechanism through which readability can influence performance judgment. Prior studies (Bloomfield 2008; Li 2008) and regulators (SEC 1998) imply that high readability can improve investors’ understanding of management disclosures, but there has been no evidence of such benefits. Instead, Rennekamp (2012) finds that processing fluency rather than understanding explains investors’ greater reliance on more readable disclosures. One possible explanation for her finding is that the disclosure conveys a consistent message in her setting. We show that in the presence of inconsistent benchmark performance, high readability improves understanding, which in turn influences performance judgments by investors. More importantly, investors’ performance judgments are less affected by processing fluency in such cases. These results suggest that when messages are inconsistent, high readability enhances understanding and reduces investors’ reliance on heuristic cues, supporting regulators’ attempts to promote greater disclosure readability.

Our findings provide important insights on the effects of managers’ disclosure strategies on investors’ judgments. Prior studies demonstrate that managers can positively sway investors’ judgments by strategically comparing current results only against selected earnings benchmarks that make the firm’s performance appear more favorable (e.g., Schrand and Walther 2000). We show that managers can actually achieve the same objective by selectively making the information with unfavorable performance implications more difficult to read, even if they compare the firm’s performance against benchmarks that lead to both favorable and unfavorable inferences. Regulators, who are interested in knowing how investors are influenced by the way information is communicated in public disclosures and financial reports (SEC 2003; Pozen 2008), should also find our results important. Regulators should be concerned that even when managers appear transparent by discussing results against multiple benchmarks, they may also strategically obfuscate some important unfavorable information. Similarly, regulators should also be concerned when managers make the overall disclosure generally readable, but strategically use difficult-to-read language in selected parts of the disclosure. In promoting high readability, regulators can consider specifying additional guidelines warning against managers who strategically highlight favorable performance indicators and obfuscate unfavorable performance indicators.

The next section reviews related literature and develops our hypothesis. Section III describes our research design and experimental procedures. Sections IV, V, and VI report the study’s results. We conclude in Section VII.

## II. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

### Benchmark Performance Consistency

While there are many forms of inconsistencies in management disclosures, such as news relating to both positive and negative developments in revenue growth, we focus on the common situation of inconsistent messages concerning a firm’s current-period performance relative to different benchmarks. For example, Rees (2005) documents that over one-third of his sample observations report inconsistent benchmark performance. Furthermore, management’s discussion of a firm’s performance can highlight largely consistent evaluations relative to all benchmarks (e.g., earnings and revenue are higher relative to both prior management guidance and year-ago quarter performance), or include some inconsistent performance evaluations, such as earnings and revenue being higher than prior management guidance, but lower than year-ago quarter performance. Moreover, managers tend to strategically emphasize performance improvements relative to a chosen benchmark (Schrand and Walther 2000; Krische 2005). For example, managers may emphasize non-GAAP earnings performance to portray a more favorable firm performance when the GAAP earnings performance is unsatisfactory (Brown, Davis, and Matsumoto 2005). This reference to different benchmarks can

lead to mixed evaluative outcomes, such as better performance in terms of beating one benchmark, but worse performance in terms of missing another.

### **The Joint Effect of Readability and Benchmark Performance Consistency on Performance Judgments**

Prior studies find that readability of a disclosure affects investors' reactions. For example, [You and Zhang \(2009\)](#) find that investors under-react to 10-K filings with low readability, and [Miller \(2010\)](#) finds that small investors reduce their trading activities around the 10-K filing date when these 10-K filings have low readability. The cognitive difficulty in processing information with low readability has been proposed as an explanation in these archival studies. The implicit assumption is that low readability deters investors' understanding of financial reports ([Bloomfield 2008](#)), but no direct evidence exists to support this assumption.

We posit that low readability is more likely to deter investors' understanding of management disclosures in the presence of inconsistencies. Understanding captures individuals' mental grasp of the meaning of the text content ([Rumelhart 1984](#); [Miele and Molden 2010](#)). When the firm's performance valence varies depending on which benchmark is used such that benchmark performance is inconsistent, it is difficult for investors to make inferences and arrive at an integrated evaluation. In such cases, if management makes the negative benchmark performance measures more (less) readable, then investors will be more (less) likely to understand the negative implications, leading to more (less) negative investors' judgments concerning the firm's future performance. Similarly, if management makes the positive benchmark performance measures more (less) readable, then investors will be more (less) likely to understand the positive implications, resulting in more (less) positive investors' judgments concerning the firm's future performance. This suggests a two-way interaction between readability and the valence of selected benchmark performance measures in the presence of inconsistent benchmark performance.

In contrast, consider a case where a firm's benchmark performance is consistent because the current-period performance beats or misses both benchmarks. Here, the directional implication of the current-period performance is unambiguous and easily understood. Even varying the readability of selected benchmark performance measures is likely to have a limited role in influencing investors' understanding of the firm's performance. This is in line with the argument in [Kintsch and Van Dijk \(1978\)](#) that people cognitively accept a text's meaning if there is referential coherence because the arguments in the text are consistent. Since understanding does not substantially differ between disclosures with high and low readability, we expect that readability will have a limited impact on investors' judgments on the firm's future performance when the two benchmark performance measures are consistent. Considering both the consistent and inconsistent benchmark performance conditions, we predict a three-way interaction in H1:

- H1:** When a firm's performance relative to two different benchmarks is inconsistent, compared to lower readability, higher readability of the positive (negative) benchmark performance measures leads to more positive (negative) investors' judgments concerning the firm's future performance. This effect of readability is smaller when benchmark performance is consistent.

## **III. METHOD**

### **Materials and Manipulations**

We design our case based on an actual company listed on the New York Stock Exchange (NYSE). The materials first provide some background information and historical financial data for

the company, and then present the earnings release, which contains four paragraphs. The first and second paragraphs contain management's comments on current-quarter performance relative to previously issued management guidance, where we manipulate the *valence* of guidance performance.<sup>3</sup> The third paragraph presents the key financial highlights that compare current-quarter performance with year-ago quarter performance, where we manipulate the *valence* and *readability* of trend performance. The last paragraph contains management earnings guidance for the next quarter, which is held constant across conditions.<sup>4</sup> Appendix B illustrates our eight manipulated conditions within the four-paragraph earnings release.

We manipulate benchmark performance consistency by varying the signs of trend performance valence and guidance performance valence such that they are either consistent or inconsistent. In the *positive (negative)* guidance performance condition, current-quarter performance *beats (misses)* previously issued management guidance. In the *positive (negative)* trend performance condition, the majority (five out of seven) of the performance indicators in the current quarter show a *positive (negative)* trend compared to those in the year-ago quarter. Thus, when trend performance is positive, benchmark performance is *consistent (inconsistent)* if current-quarter performance *beats (misses)* previously issued management guidance. Similarly, when trend performance is negative, benchmark performance condition is *consistent (inconsistent)* if current-quarter performance *misses (beats)* previously issued management guidance.

The readability manipulation varies the extent to which participants can easily understand the implications of the trend performance. The low readability version is adapted from actual earnings releases, and we create the high readability version by rewriting the earnings release following plain English writing principles.<sup>5</sup> In our experiment, the *high* readability disclosure involves short sentences and organized structures with table/bullet presentations. In contrast, the *low* readability disclosure involves long sentences and mixed structures without table/bullet presentations. We summarize the manipulation of readability and trend performance valence in Appendix C.<sup>6</sup>

## Participants

Our participants are 131 M.B.A. students from a major U.S. university who received course credits for participating in the study. The participants have a mean work experience of 13 years. On average, they have taken 3.59 accounting and 2.34 finance courses. Eighty percent of the participants have stock investment experience and 86 percent (85 percent) of them have experience in reading earnings releases (annual reports).<sup>7</sup>

<sup>3</sup> There are minor variations in the readability of the guidance performance. Our results indicate that these variations do not differentially influence participants' understanding of the guidance performance.

<sup>4</sup> We manipulate the presentation order of guidance performance and trend performance in another experiment. Results show that presentation order has no effect on investors' judgments.

<sup>5</sup> We manipulate readability based on plain English writing principles, which include (1) short sentences; (2) definite, concrete, everyday words; (3) active voice; (4) tabular presentation or bullet lists for complex material whenever possible; (5) no legal jargon or highly technical business terms; and (6) no multiple negatives (SEC 1998). We use principles (1) and (4) in our manipulations, holding constant the other principles.

<sup>6</sup> In contrast to the manipulation of narrative versus list presentation in Sedor (2002), where causal order/links and certain key information vary across conditions, our manipulation of readability holds those factors constant between the low and high readability conditions and varies the ease with which participants can assimilate the information.

<sup>7</sup> Participants' background information in terms of courses taken, work experience, experience in investment, and experience in reading financial disclosures does not vary across manipulated conditions; smallest  $p = 0.12$ . Following Tan et al. (2014), who find that investor sophistication can influence the effects of linguistic features on investors' judgments under certain circumstances, we control for participants' background information, and our results remain unchanged.

Since nonprofessional investors are the primary beneficiaries of the plain English guidelines promoted by the SEC, we are mainly interested in the impact of readability for such investors. We consider our experimental task to involve moderate integrative complexity since it requires participants to integrate the implications of the firm's trend performance and guidance performance. Elliott, Hodge, Kennedy, and Pronk (2007) suggest that select M.B.A. students, who have completed the first year of an M.B.A. program and have elected to take a financial statement analysis course, are good proxies for nonprofessional investors for tasks with integrative complexity. Hence, select M.B.A. students are likely appropriate participants for our study. In their study, select M.B.A. students in Experiment 1 (Experiment 2) have taken 3.5 (3.6) accounting courses, 2.9 (3.5) finance courses, have 5.9 (4.4) years of work experience, and 93 percent (100 percent) of them have evaluated financial statements before. In our study, the M.B.A. students have taken 3.59 accounting courses, 2.34 finance courses, have 13 years of work experience, and 86 percent of them have read annual reports or earnings releases before, suggesting that our participants are comparable to the select M.B.A. students recruited in Elliott et al. (2007).

### **Procedure**

All participants are told to assume the role of a general investor. They first read the case materials, and then assess the firm's future earnings potential, the stock price appreciation potential, the likelihood of purchasing the stock, and the firm's price/earnings (P/E) ratio. Participants are instructed that they cannot go back to previous pages once they proceed to the next section. In the second section, we assess participants' understanding by asking them to indicate whether the firm's guidance performance indicators and trend performance indicators decrease, remain constant, or increase. Finally, participants answer debriefing questions, manipulation check questions on readability, performance valence, and benchmark performance consistency, and provide demographic information in the third section.

## **IV. RESULTS**

### **Manipulation Checks**

As a check on our readability manipulation, we ask participants to indicate the extent to which (1) the earnings release is difficult to read, (2) the earnings release is difficult to understand, and (3) the information in the earnings release is difficult to process, on an 11-point scale with endpoints 0 = "not at all difficult" and 10 = "extremely difficult." Participants' responses to these three questions are highly correlated (smallest Pearson correlation coefficient = 0.73,  $p < 0.01$ ; Cronbach's alpha = 0.93).<sup>8</sup> Thus, we average the three responses to get an overall readability manipulation check. The mean rating in the high readability condition (4.58) is significantly lower than that in the low readability condition (5.60),  $t = 2.91$ ,  $p < 0.01$ , suggesting that our readability manipulation is successful.

As a check on our manipulation of trend performance valence (positive versus negative), we ask participants to indicate the extent to which the firm's current-quarter earnings performance is favorable compared to year-ago quarter in the key financial highlights section, on an 11-point scale with endpoints 0 = "extremely unfavorable" and 10 = "extremely favorable." The mean rating of 5.55 in the positive trend performance condition is significantly higher than the mean rating of 3.78 in the negative trend performance condition,  $t = 5.46$ ,  $p <$

---

<sup>8</sup> All p-values are two-tailed unless otherwise specified.

0.01, suggesting that our manipulation of trend performance valence is successful. As a check on our manipulation of positive versus negative guidance performance valence, we ask participants to indicate the extent to which the firm's current-quarter earnings performance is favorable compared to prior management guidance, on an 11-point scale with endpoints 0 = "extremely unfavorable" and 10 = "extremely favorable." The overall mean rating of 5.95 in the beat guidance setting is significantly higher than the mean rating of 3.93 in the miss guidance setting,  $t = 6.44$ ,  $p < 0.01$ , suggesting that our manipulation of guidance performance valence is successful.

Finally, as a check on our manipulation of benchmark performance consistency, we ask participants to indicate the extent to which the benchmark performance/message is consistent, on an 11-point scale with endpoints  $-5 =$  "extremely inconsistent" and  $5 =$  "extremely consistent." Participants' responses to these two questions are significantly positively correlated (Pearson correlation coefficient = 0.51,  $p < 0.01$ , Cronbach's alpha = 0.67). Thus, we average the two responses to get an overall consistency manipulation check.<sup>9</sup> The mean ratings in the consistent and inconsistent benchmark performance conditions are significantly different from each other (0.04 versus  $-0.67$ ,  $t = 2.30$ ,  $p = 0.02$ ). For all our manipulation check questions, no other main or interaction effect is significant (smallest  $p = 0.26$ ).

### Dependent Variable

We asked two questions to assess participants' judgments on the firm's future performance: "[t]o what extent do you agree that the firm's earnings performance will be strong/stock price will go up in the near future?" Both responses are measured on 11-point scales with endpoints  $-5 =$  "strongly disagree" and  $5 =$  "strongly agree." Participants' earnings judgments and stock price judgments are significantly positively correlated (Pearson correlation coefficient = 0.88,  $p < 0.01$ , Cronbach's alpha = 0.94). Thus, we average the two responses to get an overall performance judgment score.<sup>10</sup>

### Test of Hypothesis

H1 predicts that when trend performance is positive and benchmark performance is inconsistent, higher as opposed to lower readability of this positive trend performance leads to higher performance judgment. Similarly, when trend performance is negative and benchmark performance is inconsistent, higher as opposed to lower readability of this negative trend performance leads to lower performance judgment. This effect of readability is smaller with consistent benchmark performance in which performance is positive or negative for both benchmarks. Given this framing, H1 suggests a two-way interaction between trend performance valence and readability when benchmark performance is inconsistent, as shown in Figure 1, Panel A, and no such interaction when benchmark performance is consistent, as shown in Figure 1, Panel B. H1 also suggests a three-way interaction involving readability, trend performance valence, and benchmark performance consistency.

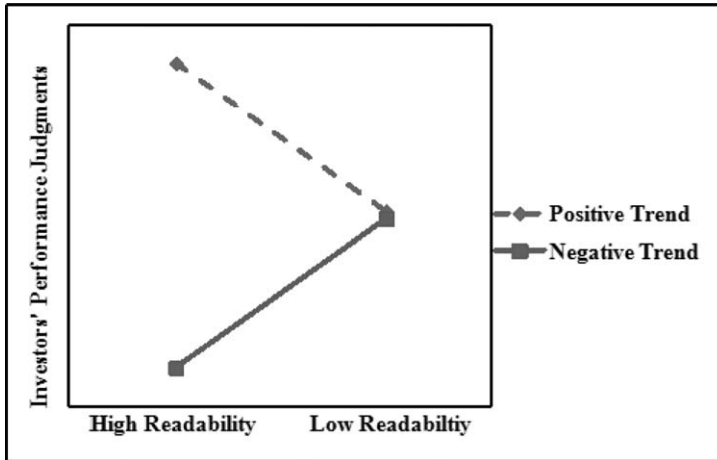
<sup>9</sup> Our construct of benchmark performance consistency is different from the construct of plausibility of management disclosures. [Barton and Mercer \(2005\)](#) show that explanations in management disclosure lead to higher or lower analysts' earnings judgments depending on whether the explanation is plausible. It is possible that consistent benchmark performance is considered to be more plausible than inconsistent benchmark performance. To examine this possibility, we ask participants to indicate the plausibility of management disclosures on an 11-point scale. Our results show that benchmark performance consistency does not influence the perceived plausibility of management disclosure (5.75/5.33 in the consistent/inconsistent benchmark performance condition,  $t = 1.19$ ,  $p = 0.23$ ).

<sup>10</sup> We obtain similar results if earnings performance judgments and stock price judgments are separately analyzed.

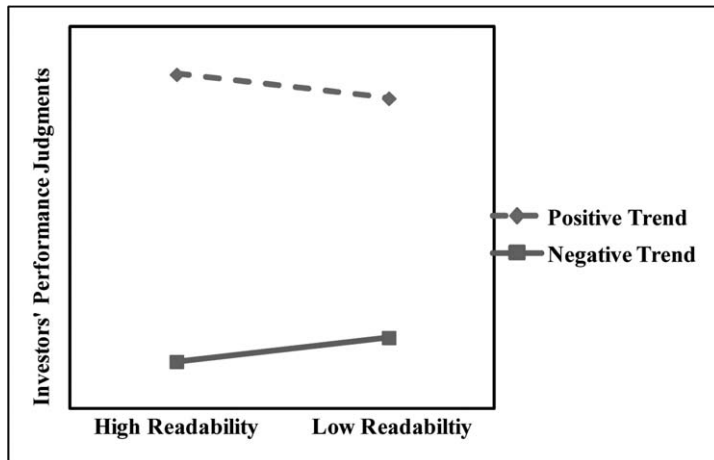


**FIGURE 1**  
**Predicted Effects of Readability and Trend Performance Valence**

**Panel A: Inconsistent Benchmark Performance**



**Panel B: Consistent Benchmark Performance**



We conduct an analysis of variance (ANOVA) for participants' performance judgments, with readability, trend performance valence, and benchmark performance consistency as the independent variables. The results are shown in Table 1, with Panel A showing the descriptive statistics and Panel B presenting the three-way ANOVA results. Consistent with H1, Table 1, Panel B shows a significant three-way interaction effect ( $F = 5.05, p = 0.03$ ), supporting our prediction that the effect of readability on performance judgment is larger when benchmark performance is inconsistent, with the directional effect varying with trend performance valence. We also find a significant main effect of trend performance valence ( $F = 10.23, p < 0.01$ ), and a significant interaction effect between

**TABLE 1**  
**Investors' Performance Judgments**

**Panel A: Descriptive Statistics—Means (Standard Deviations) [Sample Size]**

Read-ability	Consistent Benchmark Performance		Inconsistent Benchmark Performance	
	Positive Trend Performance	Negative Trend Performance	Positive Trend Performance	Negative Trend Performance
High	0.47 (1.98)	-0.75 (1.74)	1.59 (2.08)	-1.17 (2.21)
Low	0.41 (2.23)	-0.78 (2.37)	-0.36 (2.05)	0.17 (1.87)

**Panel B: Three-Way ANOVA Test**

Source	Type III Sum of Squares	df	Mean Square	F-stat	p-value
Readability	1.03	1	1.03	0.24	0.63
Trend Performance Valence	43.95	1	43.95	10.23	<0.01
Benchmark Performance Consistency	1.59	1	1.59	0.37	0.54
Trend Performance Valence × Benchmark Performance Consistency	0.07	1	0.07	0.02	0.90
Trend Performance Valence × Readability	22.38	1	22.38	5.21	0.02
Benchmark Performance Consistency × Readability	0.58	1	0.58	0.14	0.71
Trend Performance Valence × Consistency × Readability	21.69	1	21.69	5.05	0.03
Error	528.64	123	4.30		

Participants are asked to indicate “[t]o what extent do you agree that the firm’s earnings performance will be strong (stock price will go up) in the near future” on an 11-point scale, with endpoints -5 = “Strongly Disagree” and 5 = “Strongly Agree.” Participants’ earnings judgments and stock price judgments are highly correlated (Pearson correlation coefficient = 0.88,  $p = 0.00$ ; Cronbach’s alpha = 0.94). We average the two responses to get an overall performance judgement score. Panel A shows the descriptive statistics. Panel B presents the results of the three-way ANOVA.

TABLE 2

## Investors' Performance Judgments

## Separate Analyses for the Inconsistent and Consistent Benchmark Performance Conditions

## Panel A: The Inconsistent Benchmark Performance Condition

Source	df	Type III Sum of Squares	Mean Square	F-stat	p-value
Readability	1	1.54	1.54	0.36	0.55
Trend Performance Valence	1	19.83	19.83	4.69	0.04
Trend Performance Valence $\times$ Readability	1	43.01	43.01	10.17	<0.01
Error	60	253.68	4.23		

## Panel B: The Consistent Benchmark Performance Condition

Source	df	Type III Sum of Squares	Mean Square	F-stat	p-value
Readability	1	0.03	0.03	0.01	0.93
Trend Performance Valence	1	24.30	24.30	5.57	0.02
Trend Performance Valence $\times$ Readability	1	0.00	0.00	0.00	0.98
Error	63	274.96	4.34		

Participants are asked to indicate "[t]o what extent do you agree that the firm's earnings performance will be strong (stock price will go up) in the near future" on an 11-point scale, with endpoints  $-5 =$  "Strongly Disagree" and  $5 =$  "Strongly Agree." Participants' earnings judgments and stock price judgments are highly correlated both in the inconsistent benchmark performance condition (Pearson correlation coefficient = 0.87,  $p = 0.00$ ; Cronbach's alpha = 0.93) and in the consistent benchmark performance condition (Pearson correlation coefficient = 0.90,  $p = 0.00$ ; Cronbach's alpha = 0.94). We average the two responses to get an overall performance judgment score in both conditions. Panel A presents the two-way ANOVA results for the inconsistent benchmark performance condition, and Panel B presents the two-way ANOVA results for the consistent benchmark performance condition.

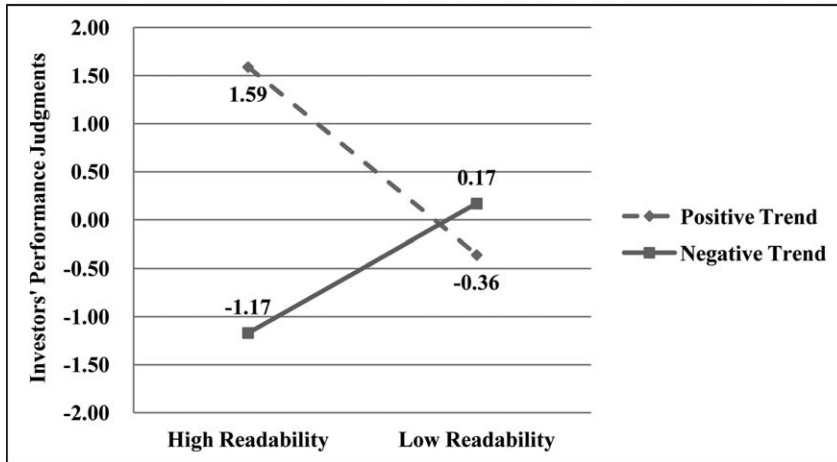
readability and trend performance valence ( $F = 5.21$ ,  $p = 0.02$ ). No other effect is significant (smallest  $p = 0.54$ ).<sup>11</sup>

We separately analyze investors' performance judgments when benchmark performance is inconsistent versus consistent with results in Table 2 and Figure 2. Table 2, Panel A shows the performance judgment results when benchmark performance is inconsistent. As predicted, we find a significant interaction effect between readability and trend performance valence,  $F = 10.17$ ,  $p < 0.01$ . Compared to participants in the low readability condition, those in the high readability condition make higher performance judgments when trend performance is positive (1.59 versus  $-0.36$ ,  $t = 2.76$ ,  $p < 0.01$ , one-tailed), but lower performance judgments when trend performance is negative ( $-1.17$  versus  $0.17$ ,  $t = 1.79$ ,  $p = 0.04$ , one-tailed). These results support our prediction that higher readability of positive (negative) trend performance leads to higher (lower) performance judgment in the inconsistent benchmark performance condition. In addition, we find that performance judgments are

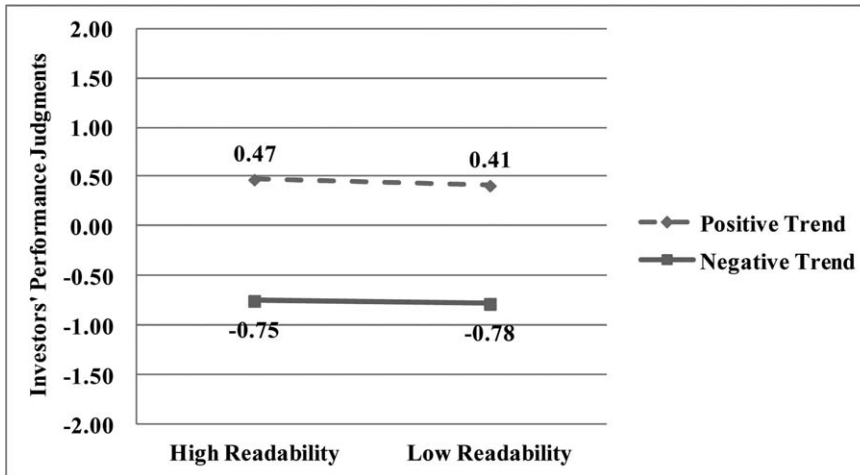
<sup>11</sup> The main effect of benchmark performance consistency is not significant ( $F = 0.37$ ,  $p = 0.54$ ), suggesting that it does not influence performance judgment. Specifically, when readability is high and trend performance is positive, there is no significant difference in participants' performance judgments between the consistent condition (0.47) and the inconsistent condition (1.59),  $t = 1.57$ ,  $p = 0.13$ . Similarly, when readability is high and trend performance is negative, participants' performance judgments in the consistent condition ( $-0.75$ ) and that in the inconsistent condition ( $-1.17$ ) are not significantly different from each other, as well,  $t = 0.60$ ,  $p = 0.55$ .

**FIGURE 2**  
**Effects of Readability and Trend Performance Valence**

**Panel A: Inconsistent Benchmark Performance**



**Panel B: Consistent Benchmark Performance**



significantly higher in the positive trend performance condition than those in the negative trend performance condition when readability is high ( $t = 3.58, p < 0.01$ ), but insignificantly so when readability is low ( $t = 0.77, p = 0.45$ ). These results suggest that high readability helps investors distinguish between the positive and negative trend performance in the presence of inconsistent benchmark performance.

Table 2, Panel B presents the performance judgment results when benchmark performance is consistent. We find that participants' performance judgments are higher in the positive trend performance condition than those in the negative trend performance condition (0.44 versus  $-0.77, F = 5.57, p = 0.02$ ). Neither the main effect of readability ( $F = 0.01, p = 0.93$ ) nor the interaction effect

between readability and trend performance valence ( $F = 0.00$ ,  $p = 0.98$ ) is significant, suggesting no effect of readability on participants' performance judgments when benchmark performance is consistent. Overall, these results support H1.<sup>12</sup>

## V. MEDIATION ANALYSIS

Our analysis above shows that readability has an impact on participants' performance judgments only in the presence of inconsistent benchmark performance. Here, we explore the mechanism through which this readability effect occurs. We are interested in two potential mediators: (1) processing fluency, and (2) understanding. While processing fluency is an individual's subjective feeling of ease in processing information (Winkielman et al. 2012), understanding is a comprehension measure that captures an individual's mental grasp of the meaning of the text content (Rumelhart 1984; Miele and Molden 2010). Readability, as a linguistic characteristic of the disclosure, can affect both. For example, Rennekamp (2012) shows that processing fluency mediates the effect of readability on investor reaction, but her setting involves one where the disclosure does not contain inconsistencies. In contrast, our theory suggests that understanding can play a magnified role in the presence of inconsistencies. Hence, we now examine the mediating role of both processing fluency and understanding when benchmark performance is inconsistent.

### Processing Fluency

Following prior studies on processing fluency (Winkielman et al. 2012), we measure processing fluency by asking participants to indicate the extent to which the information in the earnings release is difficult to process, on an 11-point scale with endpoints 0 = "not at all difficult" and 10 = "extremely difficult." Our results show that higher readability is associated with lower ratings on this difficulty measure, which reflects higher fluency, when benchmark performance is inconsistent (4.84 versus 6.03,  $t = 2.32$ ,  $p = 0.02$ ). However, the processing fluency measure is not associated with participants' performance judgments (Pearson correlation coefficient = 0.07,  $p = 0.58$ ), suggesting that processing fluency cannot explain the readability effect when the disclosure contains inconsistent benchmark performance.<sup>13</sup>

### Understanding

When benchmark performance is inconsistent, the valence of guidance performance and that of trend performance always have opposite performance implications. As a result, to assess the impact of understanding on participants' performance judgments, we use a "net understanding" measure that reflects participants' comprehension of trend performance relative to that of guidance performance.

---

<sup>12</sup> We repeat our main analysis using participants' responses to the question on how likely they will purchase the firm's stock in the near future, and find similar results. Further, participants' performance judgments fully mediate the joint effect of readability, trend performance valence, and benchmark performance consistency on participants' likelihood of purchasing the firm's stock. However, we do not find any effect on P/E ratio (smallest  $p = 0.12$ ), probably because our manipulated variables have similar effects on expected earnings performance and stock price performance. For example, if investors expect a firm to have both higher EPS and higher stock price in the future, then their expected P/E ratio may not change.

<sup>13</sup> A common issue with studies that investigate the effects of readability on processing fluency is that the measure of processing fluency is often indistinguishable from a manipulation check for readability (Miele and Molden 2010; Rennekamp 2012). Similarly, we use "difficult-to-process" as a measure of processing fluency, although this measure can also be interpreted as a readability manipulation check and was one of three readability manipulation check questions in our study. We also use "difficult-to-read" and "difficult-to-understand" as measures of processing fluency, and neither of them is associated with performance judgments ( $p > 0.73$ ).

To measure “understanding,” psychology studies often ask participants to complete multiple-choice questions that test the correctness of their inferences based on the message they read (Melby-Lervag and Lervag 2014). To choose a correct answer from different options that provide alternative interpretations of the message, participants must understand not only the explicitly stated information, but also the overall implication of the text (Rawson and Dunlosky 2002; Miele and Molden 2010). Accordingly, we measure participants’ understanding of the firm’s performance by asking them to evaluate changes (increase, remain constant, or decrease) in four trend performance indicators (net sales, unit sales volume, earnings per share, and earnings per share from continuing operations) and four guidance performance indicators (net sales, sales growth, earnings per share, and earnings per share from continuing operations).<sup>14</sup> A participant’s response is coded as 1 if his or her answer is correct, and 0 otherwise. We then add the four coded understanding measures on trend performance to form “trend understanding,” and the four coded understanding measures on guidance performance to form “guidance understanding.” We subtract guidance understanding from trend understanding to give “net understanding,” a measure that captures the extent to which participants have comprehended information in the trend performance relative to that in the guidance performance.

Our readability manipulation mainly occurs in the trend performance section, and we expect that high readability improves trend understanding, which we find (2.90 versus 2.03,  $t = 2.51$ ,  $p = 0.01$ , one-tailed). While there are minor wording changes in describing guidance performance between the high and low readability conditions, guidance understanding does not vary with readability (2.00 versus 1.88,  $t = 0.31$ ,  $p = 0.86$ ). Net understanding is not correlated with our proxy for processing fluency (Pearson correlation coefficient = 0.00,  $p = 0.99$ ), suggesting that these two measures capture different constructs.

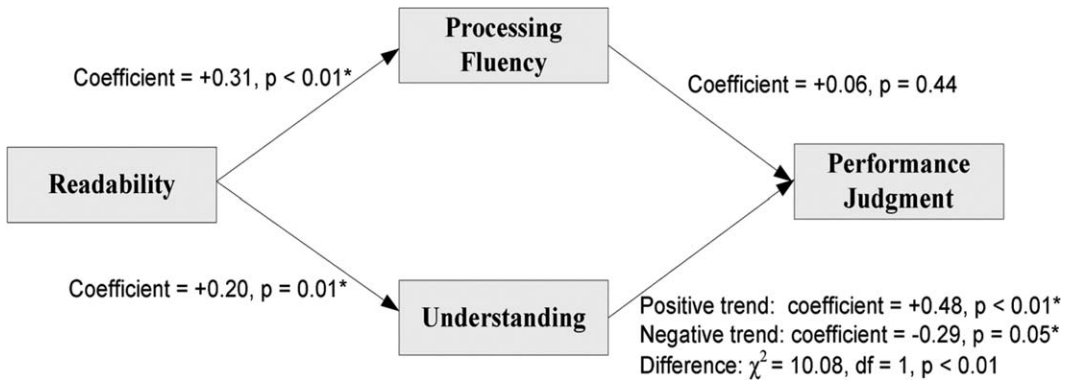
### Structural Equation Modeling Results

We conduct a structural equation modeling (SEM) analysis with readability as the independent variable, processing fluency and understanding as two mediators, and performance judgment as the dependent variable. As shown in Figure 3, our model describes the relationships in the data well as the  $\chi^2$  statistic is insignificant ( $\chi^2 = 10.54$ ,  $df = 12$ ,  $p = 0.57$ ). Specifically, we find that higher readability leads to higher net understanding (coefficient = 0.20,  $p = 0.01$ , one-tailed). Moreover, higher net understanding results in higher performance judgment when trend performance is positive (coefficient = 0.48,  $p < 0.01$ , one-tailed), but lower performance judgment when trend performance is negative (coefficient =  $-0.29$ ,  $p = 0.05$ , one-tailed). The interaction between net understanding and valence of trend performance on performance judgment is significant (difference  $\chi^2 = 10.08$ ,  $df = 1$ ,  $p < 0.01$ ). On the other hand, although readability is positively associated with processing fluency (coefficient = 0.31,  $p < 0.01$ , one-tailed), processing fluency is not associated with performance judgment (coefficient = 0.06,  $p = 0.44$ ).<sup>15</sup> These results suggest that net

<sup>14</sup> We focus on participants’ understanding of these four trend performance indicators because they provide inferences on the firm’s performance that contradict those implied by the four guidance performance indicators in the inconsistent benchmark performance conditions. Hence, correct responses to these four trend performance indicators would better reflect participants’ actual understanding of trend performance. In contrast, the other three trend performance indicators provide inferences about the firm’s performance that are directionally consistent with those provided by the guidance performance indicators. Correct responses to these three trend performance indicators may, therefore, merely reflect the influence of the guidance performance indicators. We also use all seven trend performance indicators to measure trend understanding and find similar results.

<sup>15</sup> We also use an alternative measure with trend understanding as numerator, and the sum of trend understanding and guidance understanding as denominator. This proportion of trend understanding measure is highly correlated with the net understanding measure (Pearson correlation coefficient = 0.94,  $p = 0.00$ ). Using this proportion measure, we find similar results. Our results also remain unchanged if we use the average of “difficult-to-read,” “difficult-to-understand,” and “difficult-to-process” to form the processing fluency measure.

**FIGURE 3**  
**The Mediating Role of Understanding versus Processing Fluency**



\* one-tailed p-values. This figure shows structural-equation modeling results for the mediating role of understanding and processing fluency in the main experiment when benchmark performance is inconsistent. We present the standardized coefficients and corresponding p-values next to each link.

Overall model fit indices are as follows: (a)  $\chi^2 = 10.54$ ,  $df = 12$ , and  $p = 0.57$ , and (b) Root Mean Square Error of Approximation (RMSEA) value = 0.00, below the cut-off point of 0.08 for a good fit (MacCallum, Browne, and Sugawara 1996).

understanding, rather than processing fluency, mediates the readability effect when benchmark performance is inconsistent.

## VI. SUPPLEMENTARY EXPERIMENT

In our main experiment, the readability manipulation mainly occurs in the description of the firm's trend performance. To test whether our results generalize to a different performance benchmark, guidance performance, we conduct another  $1 \times 2$  between-subjects experiment where we vary the readability of guidance performance (see Appendix D). In contrast to the main experiment where there are minor readability differences in sections other than trend performance, readability differences only occur in the guidance performance section in the supplementary experiment. In addition, while the manipulation of readability is placed in the lower half of the earnings release in our main experiment, the readability manipulation is placed at the earlier part of the earnings release in the supplementary experiment.

We choose an inconsistent benchmark performance setting where current-quarter performance generally beats prior management guidance, but is lower than that in the year-ago quarter. We continue to use seven indicators, and describe trend performance in a difficult-to-read manner. Our participants are 48 M.B.A. participants with the same profile as those in the main experiment, and they are evenly distributed between the two conditions. Our theory predicts that in the presence of inconsistent benchmark performance, higher readability of the positive guidance performance will lead to higher performance judgment, which we find (1.20 versus -0.13,  $t = 2.33$ ,  $p = 0.01$ , one-tailed). To test the mediating role of understanding versus processing fluency, we compute a similar net understanding measure that reflects the difference between guidance understanding and trend understanding. We use the same processing fluency

measure as that in the main experiment.<sup>16</sup> Untabulated results from the SEM analysis show that higher readability of positive guidance performance leads to higher net understanding (coefficient = 0.35,  $p < 0.01$ , one-tailed), which then results in higher performance judgment (coefficient = 0.35,  $p < 0.01$ , one-tailed). As in the main experiment, processing fluency is not associated with performance judgment (coefficient = 0.13,  $p = 0.33$ ). The  $\chi^2$  statistics suggest that our model has a good model fit ( $\chi^2 = 3.30$ ,  $df = 2$ ,  $p = 0.19$ ).

## VII. CONCLUSION

We conduct two experiments using M.B.A. students to investigate how readability and benchmark performance consistency jointly influence investors' judgments. In our main experiment, we manipulate the readability and valence of trend performance, and whether the sign of trend performance valence and that of guidance performance valence are consistent with each other. We find that readability influences investors' judgments to a greater extent when the two performance valences are inconsistent than when they are consistent. In the presence of inconsistent benchmark performance, high readability of trend performance helps investors better understand the firm's performance than low readability, which in turn leads to higher performance judgments when trend performance is positive, but lower performance judgments when the trend performance is negative. In a supplementary experiment, we manipulate the readability of positive guidance performance in an inconsistent benchmark performance setting in which trend performance is negative, and again find that higher readability of positive guidance performance leads to more favorable performance judgments by investors.

Prior research examines the effect on investor judgment when there are variations in the readability of an entire disclosure that conveys either consistently good or bad news (Rennekamp 2012). We extend this work by examining the effect on investor judgment in a context where variations in readability are selectively applied to parts of the disclosure, and where the disclosure either contains a consistent or inconsistent message. More specifically, we identify message consistency in the form of benchmark performance consistency as a factor that moderates the effects of readability on investors' judgments. This is important because we show that investors' welfare may be impaired when managers strategically make certain unfavorable performance measures difficult to read in the presence of inconsistent benchmark performance measures. On the other hand, when management disclosures convey consistent messages, such strategic use of low readability with selected benchmark performance measures may not work. Our findings suggest that in assessing the impact of readability of selected disclosure content, researchers and practitioners should consider message consistency because these two attributes jointly affect investors' judgments.

Our study also contributes to the literature by identifying understanding as a mediating mechanism through which readability affects investors' judgments when benchmark performance is inconsistent. Extant literature in accounting suggests that readability operates through the psychological mechanism of processing fluency (Rennekamp 2012), but examines settings where the disclosure does not contain inconsistencies. Further, we provide direct evidence on a benefit of high readability in the form of improved investors' understanding. This benefit is of interest to regulators because it provides insight concerning a previously unverified assumption made by the SEC (1998).

Our study has several limitations. First, we provide participants with short management disclosures, when the disclosures that investors encounter in practice can be significantly longer. While we believe that the effect of readability is magnified when the disclosure becomes more

<sup>16</sup> In the supplementary experiment, guidance performance has five positive indicators and two negative indicators, while trend performance has five negative indicators and two positive indicators. Our results remain unchanged whether we use all (seven) or majority (five) indicators to form the understanding measure, or if we use the average of "difficult-to-read," "difficult-to-understand," and "difficult-to-process" to form the processing fluency measure.



voluminous, it is possible that investors lose focus and become indifferent to readability in such a situation. Future research can investigate this issue. Second, we examine only one particular context, involving benchmark performance inconsistency in terms of contradictions in performance based on prior management guidance and year-ago quarter performance. Other possible contexts include contradictions in inferences based on different financial statement line items, such as revenue versus net earnings or current earnings versus future prospects.

## REFERENCES

- Acito, A. A. 2013. *Does Quarterly Earnings Guidance Increase or Reduce Earnings Management?* Working paper, Michigan State University.
- Albrecht, J. E., and E. J. O'Brien. 1993. Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19 (5): 1061–1070.
- Barton, J., and M. Mercer. 2005. To blame or not to blame: Analysts' reactions to external explanations for poor financial performance. *Journal of Accounting and Economics* 39 (3): 509–533.
- Bloomfield, R. 2002. The “incomplete revelation hypothesis” and financial reporting. *Accounting Horizons* 16 (3): 233–243.
- Bloomfield, R. 2008. Discussion of: Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (August): 248–252.
- Brown, R. M., A. K. Davis, and D. A. Matsumoto. 2005. Emphasis on pro forma versus GAAP earnings in quarterly press releases: Determinants, SEC intervention, and market reactions. *The Accounting Review* 89 (4): 1011–1138.
- Burgoon, J. K. 1975. Conflicting information, attitude, and message variables as predictors of learning and persuasion. *Human Communication Research* 1 (2): 133–144.
- Call, A. C., S. Chen, B. Miao, and Y. H. Tong. 2014. Short-term earnings guidance and accrual-based earnings management. *Review of Accounting Studies* 19 (2): 955–987.
- Courtis, J. K. 1998. Annual report readability variability: Tests of the obfuscation hypothesis. *Accounting, Auditing and Accountability Journal* 11 (4): 459–472.
- Dale, E., and J. S. Chall. 1949. The concept of readability. *Elementary English* 26 (1): 19–26.
- DeGeorge, F., J. Patel, and R. Zeckhauser. 1999. Earnings management to exceed thresholds. *Journal of Business* 72 (1): 1–33.
- Elliott, W., F. Hodge, J. Kennedy, and M. Pronk. 2007. Are M.B.A. students a good proxy for nonprofessional investors? *The Accounting Review* 82 (1): 139–168.
- Graham, J., C. Harvey, and S. Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40 (1/3): 3–73.
- Kintsch, W., and T. A. Van Dijk. 1978. Toward a model of text comprehension and production. *Psychological Review* 85 (5): 363–394.
- Koch, A. S., C. E. Lefanowicz, and P. B. Shane. 2012. *Earnings Guidance and Earnings Management Constraints*. Working paper, University of Virginia and The College of William & Mary.
- Koonce, L., N. Seybert, and J. Smith. 2013. *Management Speaks, Investors Listen: Are Investors Too Focused on Managerial Disclosures?* Working paper, The University of Texas at Austin, University of Maryland, and University of Saskatchewan.
- Krische, S. D. 2005. Investors' evaluations of strategic prior period benchmark disclosures in earnings announcements. *The Accounting Review* 80 (1): 243–268.
- Lee, S., S. R. Matsunaga, and C. W. Park. 2012. Management forecast accuracy and CEO turnover. *The Accounting Review* 87 (6): 2095–2122.
- Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2/3): 221–247.

- MacCallum, R. C., M. W. Browne, and H. M. Sugawara. 1996. Power analysis and determination of sample size for covariance structure modeling. *Psychological Method* 1 (2): 130–149.
- Masson, M. E. J., and M. A. Waldron. 1994. Comprehension of legal contracts by non-experts: Effectiveness of plain language redrafting. *Applied Cognitive Psychology* 8 (1): 67–85.
- Melby-Lervag, M., and A. Lervag. 2014. Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychology Bulletin* 140 (2): 409–433.
- Miele, D. B., and D. C. Molden. 2010. Naïve theories of intelligence and the role of processing fluency in perceived comprehension. *Journal of Experimental Psychology: General* 139 (3): 535–557.
- Miller, B. 2010. The effects of reporting complexity on small and large investor trading. *The Accounting Review* 85 (6): 2107–2143.
- Pozen, R. 2008. *Final Report of the Advisory Committee on Improvements to Financial Reporting to the United States Securities and Exchange Commission*. (August 1). Washington, DC: Securities and Exchange Commission Advisory Committee.
- Rawson, K. A., and J. Dunlosky. 2002. Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28 (1): 69–80.
- Rees, L. 2005. Abnormal returns from predicting earnings thresholds. *Review of Accounting Studies* 10 (4): 465–496.
- Rennekamp, K. 2012. Processing fluency and investors' reactions to disclosure readability. *Journal of Accounting Research* 50 (5): 1319–1354.
- Rumelhart, D. E. 1984. *Understanding Reading Comprehension: Cognition, Language, and the Structure of Prose*, edited by Flood, J. Newark, DE: International Reading Association.
- Schrand, C. M., and B. R. Walther. 2000. Strategic benchmarks in earnings announcement: The selective disclosure of prior-period earnings components. *The Accounting Review* 75 (2): 151–177.
- Securities and Exchange Commission (SEC). 1998. *A Plain English Handbook: How to Create Clear SEC Disclosure*. Available at: <http://www.sec.gov/pdf/handbook.pdf>
- Securities and Exchange Commission (SEC). 2003. *Interpretation: Commission Guidance Regarding Management's Discussion and Analysis of Financial Condition and Results of Operations*. Available at: <http://www.sec.gov/rules/interp/33-8350.htm>
- Sedor, L. M. 2002. An explanation for unintentional optimism in analysts' earnings forecasts. *The Accounting Review* 77 (4): 731–753.
- Shah, A. K., and D. M. Oppenheimer. 2007. Easy does it: The role of fluency in cue weighting. *Judgment and Decision Making* 2 (6): 371–379.
- Tan, H. T., E. Y. Wang, and B. Zhou. 2014. When the use of positive language backfires: The joint effect of tone, readability, and investor sophistication on earnings judgments. *Journal of Accounting Research* 52 (1): 273–302.
- Winkielman, P., D. E. Huber, L. Kavanagh, and N. Schwarz. 2012. Fluency of consistency: When thoughts fit nicely and flow smoothly. In *Cognitive Consistency: A Fundamental Principle in Social Cognition*, edited by Gawronski, B., and F. Strack, 89–111. New York, NY: Guilford Press.
- You, H., and X. Zhang. 2009. Financial reporting complexity and investor under-reaction to 10-K information. *Review of Accounting Studies* 14 (4): 559–586.

## APPENDIX A

### Examples of Strategic Use of Readability in Selected Content

#### Example No. 1: Excerpt from United Airlines' Earnings Release United Announces Full-Year and Fourth-Quarter 2012 Results

CHICAGO, January 24, 2013—United Continental Holdings, Inc. (NYSE: UAL) today reported full-year 2012 net income of \$589 million, or \$1.59 per diluted share, excluding \$1.3

billion of special charges. Including special charges, UAL reported a full-year 2012 net loss of \$723 million, or \$2.18 per share. UAL reported a fourth-quarter 2012 net loss of \$190 million, or \$0.58 per share, excluding \$430 million of special charges. Including special charges, UAL reported a fourth-quarter 2012 net loss of \$620 million, or \$1.87 per share.

- UAL full-year 2012 consolidated passenger revenue increased 0.2 percent year-over-year. Consolidated passenger revenue per available seat mile (PRASM) increased 1.7 percent in 2012 compared to 2011.
- Superstorm Sandy reduced fourth-quarter revenue by approximately \$140 million and profit by approximately \$85 million.
- Full-year 2012 consolidated unit costs (CASM), holding fuel rate and profit sharing constant and excluding special charges and third-party business expense, increased 2.5 percent year-over-year on a consolidated capacity reduction of 1.5 percent. Full-year 2012 consolidated CASM increased 6.7 percent year-over-year.
- UAL ended 2012 with \$7.0 billion in unrestricted liquidity.
- Co-workers earned \$119 million in profit sharing for full-year 2012, which will be distributed on February 14, 2013.

#### **Fourth-Quarter Revenue and Capacity**

For the fourth quarter of 2012, total revenue was \$8.7 billion, a decrease of 2.5 percent year-over-year. Fourth-quarter consolidated passenger revenue decreased 3.6 percent to \$7.5 billion, compared to the same period in 2011. Consolidated revenue passenger miles (RPMs) decreased 3.2 percent on a consolidated capacity (available seat miles) decrease of 4.2 percent year-over-year for the fourth quarter, resulting in a fourth-quarter consolidated load factor of 82.3 percent. Fourth-quarter 2012 consolidated PRASM increased 0.6 percent compared to the same period in 2011. Consolidated yield for the fourth quarter of 2012 decreased 0.4 percent year-over-year.

#### **Example No. 2: Excerpt from Eli Lilly's Earnings Release Lilly Reports Third-Quarter 2013 Results**

- Worldwide revenue increased 6 percent, driven by solid growth for Cymbalta, insulin, Animal Health, Alimta, Cialis, and Trajenta.
- Higher revenue and ongoing cost containment drove strong operating income growth.
- Earnings per share totaled \$1.11 for the third quarter of 2013.
- 2013 earnings per share guidance narrowed to the range of \$4.33–\$4.38 (reported), or \$4.10–\$4.15 (non-GAAP).
- Company reaffirms commitment to return cash to shareholders through its dividend and share repurchase program.

Eli Lilly and Company (NYSE: LLY) today announced financial results for the third quarter of 2013.

Certain financial information for 2013 and 2012 is presented on both a reported and a non-GAAP basis. Some numbers in this press release may not add due to rounding. Reported results were prepared in accordance with generally accepted accounting principles (GAAP) and include all revenue and expenses recognized during the period. Non-GAAP measures exclude the items described in the reconciliation tables later in the release. The non-GAAP measures are presented in order to provide additional insights into the underlying trends in the company's business. The company's 2013 financial guidance is also being provided on both a reported and a non-GAAP basis.

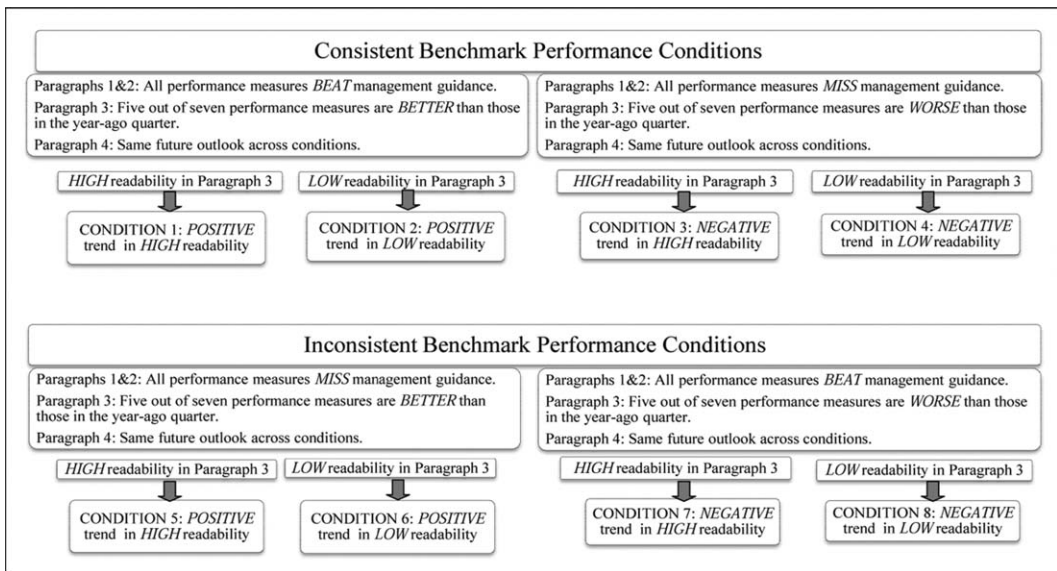
**Key Events Over the Last Three Months**

- Dulaglutide was submitted for regulatory review in both the U.S. and Europe as a potential treatment for type 2 diabetes.
- The U.S. rolling submission was completed for ramucirumab as a single-agent treatment for patients with advanced gastric cancer who have had disease progression after initial chemotherapy. A submission for ramucirumab for the same indication was also made in Europe.
- Top-line results were announced from two global Phase III studies of ramucirumab.

**Third-Quarter Reported Results**

In the third quarter of 2013, net income and earnings per share decreased to \$1.203 billion and \$1.11, respectively, compared with third-quarter 2012 net income of \$1.327 billion and earnings per share of \$1.18. The decreases in net income and earnings per share were driven by the early payment of the exenatide revenue-sharing obligation in the third quarter of 2012, partially offset by higher operating income and a lower effective tax rate in the third quarter of 2013. Earnings per share also benefited from a lower number of shares outstanding in the third quarter of 2013 compared to the third quarter of 2012.

**APPENDIX B**  
**Eight Manipulated Conditions in the Main Experiment**



APPENDIX C

Panel A: Manipulation of Readability and Trend Performance Valence in the Main Experiment

Low Readability

*Positive Trend Performance*  
 Compared with year-ago quarter, net sales grew 6 percent to \$19.8 billion for the January–March quarter, mainly due to foreign exchange impacts of 7 percent as the U.S. dollar weakened. Unit sales volume increased 3 percent largely due to general economic conditions, market recovery, prior-year acquisitions, and share gains. Gross margin expanded 2.9 percent to 52.6 percent behind price increases, lower commodity costs, and manufacturing cost savings. Operating margin was 22.5 percent, a decrease of 1.6 percent versus the prior-year period, driven by higher SG&A expenses as a percentage of net sales, partially offset by higher gross margin. Earnings per share were \$1.08, an increase of 1 percent primarily due to higher net sales, partially offset by the loss on the sale of Acxon in Japan, which decreased earnings from discontinued operations by 2 percent to \$0.24 billion. Earnings from continuing operations were \$3.02 billion, up 3 percent behind foreign exchange impacts as the U.S. dollar weakened, higher net sales, and higher base-period acquisition cost.

High Readability

See “Quarterly Results Summary” (Positive) Below

*Negative Trend Performance*

Compared with year-ago quarter, net sales declined 6 percent to \$19.8 billion for the January–March quarter, mainly due to foreign exchange impacts of 7 percent as the U.S. dollar strengthened. Unit sales volume decreased 3 percent largely due to general economic conditions, market contractions, prior-year divestitures, and share losses. Gross margin expanded 2.9 percent to 52.6 percent behind price increases, lower commodity costs, and manufacturing cost savings. Operating margin was 22.5 percent, a decrease of 1.6 percent versus the prior-year period, driven by higher SG&A expenses as a percentage of net sales, partially offset by higher gross margin. Earnings per share were \$1.08, a decrease of 1 percent primarily due to lower net sales, partially offset by the gain on the sale of Acxon in Japan, which increased earnings from discontinued operations by 2 percent to \$0.24 billion. Earnings from continuing operations were \$3.02 billion, down 3 percent behind foreign exchange impacts as the U.S. dollar strengthened, lower net sales, and higher base-period divestiture gains.

See “Quarterly Results Summary” (Negative) Below

Adapted from Tan et al. (2014).

(continued on next page)

## APPENDIX C (continued)

## Panel B: Readability: Positive Quarterly Results Summary (Positive Version—the word “positive” is not shown to the participants)

	2012 Q1	Compared with 2011 Q1	Sources/Reasons
Net sales	\$19.8 billion	↑ Grew 6%	Foreign exchange impacts of 7% as U.S. dollar weakened.
Unit sales volume	NA	↑ Increased 3%	General economic conditions, market recovery, prior-year acquisitions and share gains.
Gross margin	52.6%	↑ Expanded 2.9%	Price increases, lower commodity costs, and manufacturing cost savings.
Operating margin	22.5%	↓ Decreased 1.6%	Higher SG&A expenses as a percentage of net sales, partially offset by higher gross margin.
Earnings per share	\$1.08	↑ Increased 1%	Higher net sales, partially offset by the loss on sale of Acxon in Japan.
Earnings from discontinued operations	\$0.24 billion	↓ Decreased 2%	Loss on sale of Acxon in Japan.
Earnings from continuing operations	\$3.02 billion	↑ Up 3%	Foreign exchange impacts as U.S. dollar weakened, higher net sales, and higher base-period acquisition cost.

## Panel C: Readability: Negative Quarterly Results Summary (Negative Version—the word “negative” is not shown to the participants)

	2012 Q1	Compared with 2011 Q1	Sources/Reasons
Net sales	\$19.8 billion	↓ Declined 6%	Foreign exchange impacts of 7% as U.S. dollar strengthened.
Unit sales volume	NA	↓ Decreased 3%	General economic conditions, market contractions, prior-year divestitures and share losses.
Gross margin	52.6%	↑ Expanded 2.9%	Price increases, lower commodity costs, and manufacturing cost savings.
Operating margin	22.5%	↓ Decreased 1.6%	Higher SG&A expenses as a percentage of net sales, partially offset by higher gross margin.
Earnings per share	\$1.08	↓ Decreased 1%	Lower net sales, partially offset by the gain on sale of Acxon in Japan.
Earnings from discontinued operations	\$0.24 billion	↑ Increased 2%	Gain on sale of Acxon in Japan.
Earnings from continuing operations	\$3.02 billion	↓ Down 3%	Foreign exchange impacts as U.S. dollar strengthened, lower net sales, and higher base-period divestiture gains.

**APPENDIX D**  
**Manipulation of Readability in the Supplementary Experiment**

	2013 Q1	Guidance	Change
<i>High Readability</i>			
Net sales	\$19.8 billion	\$19.3 billion	↑ \$0.5 billion higher
Sales growth	1%	-3% to 0%	↑ higher
Sales volume growth	3%	4%	↓ lower
Cash position	\$5.4 billion	\$5.3 billion	↑ \$0.1 billion higher
Earnings per share from discontinued operations	\$0.07	\$0.09	↓ \$0.02 lower
Earnings per share from continuing operations	\$1.01	\$0.94	↑ \$0.07 higher
Earnings per share	\$1.08	\$1.03	↑ \$0.05 higher
<i>Low Readability</i>			
<p>Net sales were \$19.8 billion, \$0.5 billion higher than company's guidance of \$19.3 billion. Sales growth was 1 percent, higher than a guidance range of flat to -3 percent. Sales volume growth was 3 percent, lower than the guidance of 4 percent increase. Cash position was \$5.4 billion, \$0.1 billion higher than the guidance of \$5.3 billion. Earnings per share from discontinued operations were lower than guidance of \$0.09 at \$0.07. Higher than company's guidance of \$0.94 for earnings per share from continuing operations and \$1.03 for earnings per share, earnings per share from continuing operations were \$1.01 and earnings per share were \$1.08.</p>			

Copyright of Accounting Review is the property of American Accounting Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.