Do Financial Statement Users Judge Relevance Based on Properties of Reliability?

Kathryn Kadous
Emory University

Lisa Koonce
The University of Texas at Austin

Jane M. Thayer
The University of Georgia

ABSTRACT: Relevance and reliability (now referred to as "representational faithfulness") are qualities of financial information that both the Financial Accounting Standards Board and the International Accounting Standards Board use in setting standards for financial reporting. Despite their importance, very little research has addressed how financial statement users apply these constructs. Via experiments set within the fair value context, we show that users do not view them as independent constructs. Instead, variations in properties that are associated with the *reliability* of a measurement influence users' assessments of the *relevance* of fair value. The relationship between assessed relevance and assessed reliability is unidirectional, in that factors underlying reliability influence judgments of relevance, but factors underlying relevance do not influence judgments of reliability. Our findings are important because inappropriate assessments of relevance can influence firm valuation. The results are particularly meaningful in the context of fair value because such measurements can vary widely in reliability.

Keywords: relevance; reliability; fair value; valuation.

JEL Classifications: M41.

I. INTRODUCTION

R elevance and reliability (now referred to as "representational faithfulness") are important qualities of financial reporting that both the Financial Accounting Standards Board (FASB) and the International Accounting Standards Board (IASB) use in setting standards (FASB 2010; IASB 2010).[1] Relevance addresses the pertinence of an economic construct (e.g., fair value, historical cost) to a user's decision. Reliability addresses how well that economic construct, or phenomenon, is depicted or measured (e.g., fair value based on a market transaction versus a model). Standard setters note that although relevance and reliability are both necessary for information to be useful in decision making, information can be relevant without being reliable and can be reliable without being relevant (FASB 2010, ¶QC 17).[2] In this paper, we assert and experimentally test the idea that financial statement users conflate these two constructs when judging the relevance of fair value measurement; that is, we test the idea that users do not view these two constructs as independent.

Investigating how financial statement users think about the constructs of relevance and reliability in the context of fair value measurement is important for several reasons. First, we lack empirical evidence regarding how users assess these qualities. Even the limited research and debates to date suggest that the answer is unclear. Some researchers indicate that users view relevance and reliability as inversely related (e.g., Schipper 2003; Dye and Sridhar 2004), while others assert a positive relationship (McCaslin and Stanga 1983; Duncan and Moores 1988). Second, one potential reason why many financial statement users oppose fair value accounting may originate in their conflating the concepts of relevance and reliability. Indeed, some of the criticisms raised against fair value accounting during the recent financial crisis appear to be due to individuals judging the relevance of fair value based on the reliability of its measurements (American Bankers Association 2009). By providing scholarly evidence on this issue, we provide to standard setters information that will help them make important decisions about fair value measurements in accounting.

In this paper, we report the results of several experiments in which we investigate the hypothesis that financial statement users' assessments of the relevance of an economic construct are influenced by factors (i.e., potential for bias and error) underlying the reliability of its measurement. We examine this idea in the context of fair value measurement of an asset, as relevance and reliability issues are of particular practical concern there.[3] Finally, we also investigate the economic importance of our ideas by determining whether the relationships noted above influence users' judgments about a firm's valuation.

Our hypotheses are grounded in psychology theories about attribute substitution (Kahneman and Frederick 2002). Attribute substitution is a heuristic used by individuals asked to evaluate a difficult-to-assess attribute. The theory predicts that when asked to evaluate a less accessible

---

[1] The FASB and IASB recently replaced the term "reliability" with "representational faithfulness" to address the possibility that auditors would associate reliability with vouchability in addition to the factors it was intended to incorporate (Schipper 2005). This modification to the conceptual framework does not affect our study because we examine features of reliability (i.e., error and bias) that are unrelated to the change. We use the more familiar term, reliability, herein.

[2] The FASB distinguishes relevance and reliability from a third construct, decision usefulness. They note that "[i]nformation must be both relevant and faithfully represented if it is to be useful. Neither a faithful representation of an irrelevant phenomenon, nor an unfaithful representation of a relevant phenomenon, helps users make good decisions" (FASB 2010, ¶QC 17). Their comments imply that an unreliable measure of a highly relevant phenomenon does not diminish the relevance of the phenomenon to the decision, but reduces the decision usefulness of the information.

[3] Although our paper is set in the context of fair value, it does not address the issue of whether fair value is more or less relevant than historical cost (see Koonce et al. [2011] for a study of that issue). Rather, our study focuses on how the judged relevance of fair value is influenced by variations in the reliability of its measurement.

attribute, individuals often (without awareness) substitute an evaluation of a more accessible attribute of the same entity. The accessibility of an attribute relates to an individual's ability to conceptualize the value of an attribute, with increases in accessibility due, in part, to the frequency with which the attribute is evaluated as part of perception and comprehension (Kahneman and Frederick 2002). We provide evidence that reliability is a basic property of fair value measurement that users understand well. On the other hand, relevance is decision specific and requires a more difficult, case-by-case analysis. Thus, we maintain that in the fair value context, reliability is a more accessible attribute than is relevance. We, therefore, predict that users will substitute the more accessible assessment of measurement reliability when asked to assess the relevance of a fair value.

The results of our experiments support our hypothesis that variations in factors associated with the reliability of an asset's fair value measurement affect financial statement users' assessments of the relevance of a fair value. Specifically, we find that a measurement's source and method of derivation (i.e., a market transaction versus a model) strongly influence users' judgments about relevance. Because attribute substitution is unidirectional, users substitute evaluations of more accessible attributes for evaluations of less accessible attributes. However, the opposite does not occur. Accordingly, we also show that manipulations of factors underlying relevance do not affect users' assessments of reliability. Finally, results support the idea that because of this attribute substitution, factors underlying reliability may be over-weighted in users' valuation judgments; that is, these factors influence valuation not only through their effect on assessed reliability, but also through their influence on assessed relevance. While we cannot establish a normative benchmark for valuation in our experiment, these results suggest that properties underlying reliability have too much emphasis in valuation.

Our study has several important implications. First, our results shed light on one reason that efforts of the FASB and the IASB to extend the use of fair values in financial reports are being met with resistance. We show that acceptance of the concept of fair value measurement may be conditioned primarily on how well the items are measured rather than on separate assessments of the relevance of the economic construct of fair value and the reliability of its measurement. This idea finds anecdotal support when, during the recent financial crisis, observers argued that thin markets and unreliable measurements created fair value accounting outcomes that were not relevant to decision makers (American Bankers Association 2009). We believe our study is important to this debate; it indicates that at least part of the reason users argue against the use of fair value may be that they do not consider reliability and relevance to be independent constructs. The primacy of reliability in users' minds may indicate that standard setters should focus on standards that enhance the reliability of fair value measurements in order to make the biggest gains in acceptance of fair value. Additionally, standard setters may wish to focus on ways to better communicate the importance of relevance of particular constructs separate from their efforts to improve the reliability of the measurements.

Second, users' conflation of the qualities of reliability and relevance may hamper additional attempts to improve the actual reliability of fair value measures or the communication of reliability. Although the FASB has previously issued guidance about the communication of fair value measurement reliability (e.g., SFAS 157), we argue that further efforts to improve those disclosures may be thwarted if users are unable to consider the relevance of the fair value construct *independent of* the reliability of its measurement. By documenting that users do not view these two qualities separately, we hope that our study provides a foundation for greater clarity on what can be improved (i.e., measurement reliability or its disclosure) and what is relatively fixed (i.e., relevance).

Third, our results suggest that financial statement users' conflation of relevance and reliability may influence their valuation judgments. That is, strengths or deficits in reliability due to potential error appear to be over-weighted in valuations. The natural next question is

whether users' association of relevance with reliability can be moderated. Our theory and results are consistent with the ideas of attribution substitution theory, which implies that decision makers are *unaware* of the substitution. However, we did not explicitly test this assumption. Psychology researchers have long struggled with determining how individuals unconsciously evaluate their environment, judge other individuals, and undertake actions, all while consciously thinking about something else (Wilson 2002). A better understanding of whether financial statement users are aware of the attribute substitution is important, as the success of attempts to moderate its effects is likely contingent on their awareness. Stated differently, if decision makers are unaware of the attribute substitution, then training in assessing relevance may not be sufficient to correct their behavior (e.g., see Tversky and Kahneman 1974).

In the following section, we develop our theoretical predictions. Sections III and IV describe our two primary experiments, respectively. Section V provides robustness tests and Section VI summarizes and concludes the paper.

II. HYPOTHESES DEVELOPMENT

The Independence of Relevance and Reliability

Relevance and reliability are the two primary qualitative characteristics of accounting information that determine its usefulness in decision making (FASB 1980, 2010). Relevance applies to a construct, a type of information. Reliability, on the other hand, is a property of a specific measurement of that construct. Both relevance and reliability influence the reported measure's decision usefulness (FASB 2010, ¶ QC17).

In its conceptual framework, the FASB notes that information must be both relevant and reliable to be useful, and that neither a reliable depiction of an irrelevant phenomenon nor an unreliable depiction of a relevant phenomenon is decision useful (FASB 2010, ¶ QC17). The FASB's comments imply that all four cells of a 2 (relevance: lower, higher) × 2 (reliability: lower, higher) matrix could be populated, as illustrated in Panel A of Figure 1. The FASB further suggests to preparers that when considering a particular decision, they should follow a sequence of steps, looking first for the most relevant information for that decision and then ascertaining the reliability of a measure of that information. If that information is lacking in reliability (or is just not available), then the decision maker should select the next most relevant measure, then again check its reliability. This process is repeated as necessary (FASB 2010, ¶ QC18; IASB 2010, ¶ QC18). Although many assert that the FASB's conceptual framework is *suggestive* of the independence of relevance and reliability (Maines and Wahlen 2006; Li et al. 2010), the framework does not *explicitly* state that the concepts are independent, and so the FASB's view on this issue is unknown.

Although the FASB's conceptual framework is unclear on the idea that relevance and reliability are independent qualities—that is, that a particular level of one does not imply a particular level in another—viewpoints outside of accounting maintain that the two constructs should be viewed as independent. In marketing, information usefulness is deemed a function of its diagnosticity (i.e., relevance) and, separately, its reliability (Dick et al. 1990). In law, information is useful if it is admissible (i.e., has relevance) and has probative value (i.e., is reliable) (Mueller and Kirkpatrick 1995). In research settings, a construct is useful if it is relevant and its measurement is reliable (Herzog 1997).

The following example, outside of the financial reporting context, helps to illustrate the view that relevance and reliability are distinct and independent constructs. Consider an individual who is spending the day in Chicago. For that individual, the air temperature in Honolulu is not highly relevant to his decision about whether to wear an overcoat when going outside. Nonetheless, the

FIGURE 1
Relevance and Reliability as Independent Constructs

Panel A: Relevance and Reliability for a Decision about What Value to Place on an Asset

	Lower Reliability	**Higher Reliability**
Lower Relevance	Model estimate without comparables, Stand-alone value of asset in use	Comparable sales value, Stand-alone value of asset in use
Higher Relevance	Model estimate without comparables, Stand-alone value of asset held for sale	Comparable sales value, Stand-alone value of asset held for sale

Panel B: Relevance and Reliability for a Decision about Whether to Wear an Overcoat in Chicago

	Lower Reliability	**Higher Reliability**
Lower Relevance	Broken thermometer, Today's Honolulu temperature	Accurate thermometer, Today's Honolulu temperature
Higher Relevance	Broken thermometer, Today's Chicago temperature	Accurate thermometer, Today's Chicago temperature

temperature in Honolulu can be measured by a very unreliable process or instrument, or by a very reliable one. The air temperature in Chicago, on the other hand, is highly relevant to the individual for the decision about whether to wear an overcoat. Likewise, the temperature in Chicago can be measured by a very unreliable process or instrument, or by a very reliable one. Indeed, the construct of "Chicago air temperature" is relevant to the person in Chicago; the reliability of the current measurement of the Chicago temperature is a separate, independent factor that the decision maker must consider.

Do Users View Them as Independent?

While we argue that the constructs of reliability and relevance should be viewed as independent, we also question whether financial statement users judging the relevance of fair value treat them as such. As explained more fully below, prior research indicates that financial statement users have fairly accurate ideas about how to assess reliability, and so we expect that they will be able to do so readily. On the other hand, they are less likely to have a ready framework for assessing

relevance, particularly in the fair value context (Koonce et al. 2011). As a result, we posit that they will rely on attribute substitution to judge fair value relevance (Kahneman and Frederick 2002). That is, they are likely to substitute an easier question (i.e., how reliable is this fair value measure?) when asked the relatively difficult question "how relevant is fair value?" We explain our argument in more detail below.

Reliability is a basic property of information, and individuals are well acquainted with its determinants. To support this assertion, we note that numerous studies in accounting and psychology provide evidence that individuals understand that factors underlying reliability, such as the objectivity of the measurement source and the error-proneness of that source, affect the assessed decision usefulness of a reported measure. For example, Hirst (1994) showed that auditors discount the inferential value of information obtained from sources that are low in competence or objectivity relative to that of information obtained from sources higher in these characteristics. Similarly, market participants discount forecasts made by management with low prior accuracy in forecasting earnings relative to those with high prior accuracy (Williams 1996; Hirst et al. 1999). Moreover, market participants' reactions to management forecasts are, in part, a function of the credibility of management (Jennings 1987). Investors glean manager credibility from factors such as their competence, trustworthiness, and situational incentives (Mercer 2004). Additional studies indicate that users of other forms of financial reporting can identify when a source is more or less credible and will utilize the information accordingly. For example, users' assessments of financial analysts' credibility depend, in part, on analysts' prior accuracy, and users' willingness to purchase future reports from these analysts is associated with their assessments of analysts' credibility (Kadous et al. 2009). We thus propose that reliability is relatively easy to assess and, therefore, is likely to be an accessible attribute for most users.

Unlike reliability, the relevance of an item depends on the specific decision to be made and its context (FASB 1980, 2010, ¶ BC3.11). Because each situation is unique, users are less likely to have a ready schema for assessing relevance. That is, rather than relying on well-established relationships (e.g., more competent sources tend to produce more reliable measurements), users have to carefully evaluate their decision task and decision context to assess relevance. There are limited circumstances in which relevance is easy to assess (e.g., judging the relevance of the air temperature in Chicago to the decision to wear a coat). However, based on prior research by Koonce et al. (2011), we propose that in the fair value context, relevance is a less accessible attribute than reliability (also see Joyce et al. [1982] and Jurney [2008] for evidence supporting this idea in financial reporting, more generally).[4]

Because assessing the relevance of fair value will be difficult for financial statement users, we believe that attribute substitution will apply (Kahneman and Frederick 2002). In general, a person who is asked a question about an accessible (i.e., easily retrievable from their knowledge) attribute such as reliability can answer the question directly and easily. In contrast, a person who is asked a difficult question will follow a different process. The individual, being unable to access information about the target attribute, will instead answer an easier question about a more accessible attribute of the same entity or object (Kahneman 2002, 469).

In our fair value setting, we argue that the more accessible attribute is reliability and the less accessible attribute is relevance. Thus, we propose that financial statement users asked to assess the reliability of a fair value measurement of a financial statement item will readily respond to that question, as reliability is an accessible attribute. In contrast, financial statement users will have difficulty responding to a question asking them to assess the relevance of an item's fair value

---

[4] While these studies demonstrate users' relative difficulty in assessing relevance versus reliability, they do not address why this occurs nor do they test our hypothesis that variations in properties associated with reliability influence relevance judgments.

because relevance is less accessible. Following the logic of attribute substitution, users asked to assess relevance will instead answer an easier question about the reliability of the measurement of the financial statement item in question. In other words, they will assess the relevance of a reported economic construct by assessing how well it represents, or is similar to, a reliable measurement.[5] This reasoning implies that financial statement users will not only use factors underlying the reliability of a measurement in assessing its reliability, but they will also rely on those same variables in assessing relevance. This logic leads to our first hypothesis:

> **H1:** Factors underlying the *reliability* of a fair value measurement influence financial statement users' assessments of the *relevance* of fair value information.

Although our theory predicts that assessments of reliability and relevance will be positively associated, it is important to note that we also expect that the relationship is unidirectional. Manipulations of factors underlying the reliability of a measurement will influence assessments of the reported construct's relevance (and, of course, of measurement reliability). However, we do not expect the opposite relationship to hold—that is, manipulations of factors underlying relevance will not affect users' assessments of reliability. The latter is unlikely to occur because reliability is a relatively accessible attribute and so users will be able to readily answer a question about reliability.

The decision usefulness of information is a function of both its relevance and its reliability (FASB 1980, 2010). Thus, we expect users will rely on their own assessments of the relevance and reliability of information in determining how useful that information is for subsequent decisions, such as valuation. If variations in properties underlying reliability affect assessments of reliability, as is appropriate, but also affect assessments of relevance (H1), and if users incorporate assessments of both relevance and reliability into their valuation judgments, then variations in the properties underlying reliability will influence valuation through two distinct paths—one through users' assessments of the relevance of the item and another through their assessments of the reliability of the item's measurement. We summarize this prediction in the following hypothesis:

> **H2:** Factors underlying the *reliability* of a measurement influence financial statement users' valuation judgments via two effects—one through assessed reliability and a second through assessed relevance.

Two caveats apply. First, Kahneman and Frederick (2002, 53) indicate that attribute substitution usually occurs without decision maker awareness. Indeed, those who substitute one question for another simply fail to notice they are answering a different question (Kahneman 2002, 469). Although we rely on the idea of attribute substitution for our predictions, we do not test the assumption that individuals are unaware of the substitution. Second, although it is difficult to establish a normative benchmark for weights that users should apply to various factors when making valuation judgments, our ideas suggest that users place too much weight on the properties underlying reliability when they make valuation judgments. That is, reliability may have too much emphasis in valuation. This idea is consistent with prior research that indicates that highly accessible values tend to be over-weighted in judgments (Strack and Mussweiler 1997; Epley and Gilovich 2002; Chapman and Johnson 2002; Epley and Gilovich 2006).

---

[5] Judgments made in this fashion often are described as following the representativeness heuristic (Tversky and Kahneman 1974).

III. EXPERIMENT 1

Participants, Design, and Variables

We tested our hypotheses in an experiment with 129 M.B.A. students from two *BusinessWeek* top-25 schools.[6] We asked participants to assume the role of a potential investor in a manufacturing firm. All were told that the company had purchased a plot of land for $41 million two years ago. The current value of the land, generated by external consultants who relied on a model, was estimated at $53 million. As explained more fully below, we manipulated two variables relating to properties underlying the reliability and relevance of the land's current value in a 2×2 full-factorial, between-participants design. We randomly assigned participants to conditions.

We manipulated the competence of the measurement source (i.e., external consultants) at two levels. Those participants in the high competence (i.e., high reliability) conditions were provided with the following information regarding the consultants:

> Based on your experience and knowledge, you believe the consultants are unbiased, and based on the accuracy of their work you have noted on other valuations, you believe they are highly competent.

We provided participants in the low competence (i.e., low reliability[7]) conditions with the following statement:

> Based on your experience and knowledge, you believe the consultants are unbiased, but based on errors and reasoning flaws you have noted on other valuations, you believe they are not very competent.

Our second manipulation relates to a property we expect financial statement users to associate with relevance—the length of time until management plans to sell the asset (FASB 2006, SFAS 157, ¶20). We manipulated the time to sale at two levels. Approximately half of the participants were told that management intends to sell the plot of land very soon ("sell soon" conditions), which corresponds to higher relevance for the valuation decision we assigned participants. The remaining participants were told that management does not intend to sell the plot of land for at least 20 more years ("no sell" conditions), which corresponds to lower relevance for this decision.

There is debate as to whether time to sale *should* affect relevance judgments in a valuation context (Hague and Willis 1999). While some individuals believe that the relevance of fair values depends on whether something is to be sold soon or sold later, others consider the distinction as unimportant. Our objective is not to resolve this debate, but rather to leverage the finding—as shown in recent research (Koonce et al. 2011)—that, on average, the time to sale is important to investors' relevance judgments in our design. This manipulation allows us to keep constant the measurement basis (i.e., fair value). Other possible ways to manipulate relevance are less desirable

---

[6] Participants came from three sources: 24 were full-time M.B.A. students from one institution (average of 4.7 years of work experience, 4.9 finance classes, 3.1 accounting classes, 71 percent had invested in individual stocks previously and 96 percent plan to do so), 55 were full-time M.B.A. students from a second institution (average of 5.5 years of work experience, 2.5 finance classes, 2.1 accounting classes, 65 percent had invested in stocks previously and 91 percent plan to), and 50 were Executive M.B.A. students from the first institution. Time constraints prevented us from collecting individual demographics from the Executive students; however, the class as a whole averaged 12 years of work experience. Source of participants did not have significant effects on any of our dependent measures.

[7] This manipulation represents higher and lower levels of reliability rather than absolute levels of high and low. For ease of exposition, however, we refer to levels of this variable, as well as other analogous variables in the paper, as "high" and "low" rather than "higher" and "lower."

FIGURE 2
Example of Materials Used in Experiment 1
High Competence, Sell Soon Condition

Several years ago, WRSwann Co., a manufacturing company, purchased a plot of land for $41,000,000.

You learn that management intends to sell the plot of land very soon.

The current value of the land is estimated at $53,000,000. This current value number was generated by external consultants.

	Consultants' Estimate of
Original Value	Current Value
--	
$41,000,000	$53,000,000

The consultants generated this number with a model that they created. The model uses the market values for several comparable plots of land. However, no individual plot is completely comparable to the one owned by WRSwann, so the model incorporates additional factors that could affect its value (i.e., location, size, desirability, population growth). The weights that are placed on these additional factors can vary depending on the various assumptions made by the consultants.

Based on your experience and knowledge, you believe the consultants are unbiased, and, based on the accuracy of their work you have noted on other valuations, you believe they are highly competent.

as they involve changing the underlying economic information that is communicated (e.g., comparing fair value to historical cost). Figure 2 provides an example of the materials.

After presenting participants with the information described above, we asked for their responses to five questions. Our key dependent variables include assessments of the relevance of "the current value information," the reliability of "the number assigned to the land's current value," and how the "situation" affects the valuation of the firm. We worded these relevance and reliability questions carefully to ensure that participants were asked the relevance question at the construct level (i.e., relevance of current value information) and the reliability question at the operational level (i.e., reliability of its measurement). Participants assessed the relevance of the current value information to their evaluation of the firm on a 101-point scale, with 0 labeled "not at all relevant" and 100 labeled "extremely relevant." The responses to this question serve as the main dependent measure for our first hypothesis. Participants assessed the reliability of the number assigned to the land's value on a 101-point response scale ranging from 0, "very unreliable," to 100, "very reliable."

Participants provided assessments of valuation in two stages. They first assessed whether the situation described in the case materials increases, decreases, or neither increases nor decreases the value they would place on the firm. Those who indicated a change in the value also indicated the magnitude of the change on a 101-point scale, ranging from "slightly increases (decreases) how much I value the company" (0) to "greatly increases (decreases) how much I value the company" (100). We created our valuation dependent variable by coding the magnitude of change as a positive amount for increases in value and a negative amount for decreases in value. This variable is zero for participants who indicated neither an increase nor a decrease in valuation due to the reported

information. Two additional questions ask for participants' assessments of the amount of bias in the current value measurement and of management's forthcomingness.[8]

Results

Preliminary Tests

We first assess whether our manipulation of competence influenced participants' judgments of the reliability of the current value number. Participants in the high source competence conditions assessed the reliability of the current value number significantly higher (mean of 67.14) than did those in the low source competence conditions (mean of 45.39) ($F_{1,125} = 42.05$, $p < 0.01$). As expected, neither time to sale nor the interaction of time to sale and source competence significantly influenced reliability assessments (both p-values > 0.30). Also as expected, participants' ratings of bias and management forthcomingness did not significantly differ across levels of source competence (both p-values > 0.10).[9]

Tests of Hypotheses

Descriptive (inferential) statistics for participants' assessments of relevance and their valuation judgments are presented in Panel A (Panel B) of Table 1. Turning first to the relevance dependent measure, the analysis of variance (ANOVA) model reveals a significant main effect of the source competence variable ($F_{1,125} = 10.92$; $p < 0.01$), but no other significant effects. Cell means indicate that participants receiving the measurement from a high competence source viewed the current value information as significantly more relevant (mean of 65.66) than did participants receiving the measurement from a low competence source (mean of 51.44). Thus, financial statement users' determination of the relevance of current value information is based on their determination of the reliability of its measurement, supporting H1.[10] This finding is consistent with our theory, as it implies that evaluations of relevance are based on the more accessible reliability attribute. Further support for participants' reliance on attribute substitution is indicated by a positive and highly significant correlation of their assessments of reliability and relevance ($r = 0.48$, $p < 0.01$) (Kahneman and Frederick 2005).

As noted above, there were no significant effects in the model for participants' relevance ratings other than that of source competence. The manipulation of time to sale did not have the expected effect on ratings of relevance. Thus, we have evidence that source competence influences users' assessments of both reliability and relevance and that time to sale does not influence

---

[8] An alternative approach for Experiment 1 would have been to ask participants about the relevance of fair value and then provide them with the reliability of a specific fair value measurement. Such an approach would have mimicked the process recommended by standard setters in their conceptual frameworks (FASB 2010; ¶ QC18; IASB 2010 ¶ QC18). We did not follow this approach because financial statement users typically receive measures that already have characteristics of relevance and reliability. That is, they typically receive information about both constructs simultaneously. Only preparers and standard setters are likely to be in a situation to consider them in a sequential fashion. Our materials thus mimic this important contextual feature for users.

[9] An ANOVA model for bias revealed a marginally significant main effect of time to sale (p = 0.085) and a significant interaction of competence and time to sale (p < 0.01). While participants rated bias as approximately equal in the two higher competence conditions, they rated bias as higher for lower competence sources when management had no plans to sell and lower for lower competence sources when management planned to sell the asset soon.

[10] We also manipulated bias (high versus low) in Experiment 1. Bias is an underlying property of reliability and, thus, allows us another opportunity to test our hypothesis about attribute substitution. Holding competence at a high level, our results of this additional manipulation reveal that participants in the low bias condition rated the relevance of the current value information higher than did those in the high bias condition (65.66 for low bias versus 60.29 for high bias), consistent with our predictions. However, this effect is only marginally significant (p = 0.10, one-tailed).

TABLE 1

Relevance Assessments and Valuation Judgments, Experiment 1

Panel A: Mean Judgments (Standard Deviation)

	Assessed Relevance Dependent Variable Source Competence			Valuation Effect Dependent Variable Source Competence		
Time to Sale	**Low**	**High**	**Row Means**	**Low**	**High**	**Row Means**
No Sell	47.47	66.21	57.29	9.21	34.75	22.83
	(24.40)	(21.78)	(24.74)	(33.29)	(34.64)	(36.09)
	n = 30	n = 33	n = 63	n = 28	n = 32	n = 60
Sell Soon	54.94	65.09	59.86	22.35	33.63	27.60
	(22.92)	(29.61)	(26.66)	(36.25)	(32.56)	(34.75)
	n = 34	n = 32	n = 66	n = 31	n = 27	n = 58
Column Means	51.44	65.66		16.12	34.24	
	(23.73)	(25.73)		(35.20)	(33.42)	
	n = 64	n = 65		n = 59	n = 59	

Panel B: Analysis of Variance

	Assessed Relevance				Valuation Effect			
Source	**df**	**MS**	**Statistic**	**Two-Tailed p-value**	**df**	**MS**	**Statistic**	**Two-Tailed p-value**
Time to Sale	1	325	F = 0.53	0.47	1	1060	F = 0.90	0.34
Source Competence	1	6718	F = 10.92	< 0.01	1	9944	F = 8.46	0.01
Time to Sale × Source Competence	1	594	F = 0.97	0.33	1	1942	F = 1.27	0.26
Residual	125	615			114	1176		

Participants assessed the relevance of the current value information to their overall evaluation of the firm. The competence of the external consultants determining the measurement was varied at two levels, high and low. Management's plans to sell the asset were varied at two levels, sell soon versus no sell. Participants provided relevance assessments on a 101-point scale, with 0 labeled "not at all relevant" and 100 labeled "extremely relevant." Participants also assessed whether the current value information increases, decreases, or neither increases nor decreases the value they place on the firm. Participants who indicated a change in value also indicated the magnitude of the change on a 101-point scale, ranging from "slightly increases (decreases) how much I value the company" (0) to "greatly increases (decreases) how much I value the company" (100). Valuation effect is calculated by coding the magnitude of change as a positive amount for participants who indicated increases in value, a negative amount for those who indicated decreases in value, and 0 for participants who indicated no change in value.

reliability. However, because our manipulation of time to sale did not also affect assessed relevance, we cannot be sure that we have effectively manipulated factors underlying relevance. In sum, Experiment 1 allows us to test our H1 (i.e., manipulations of reliability influence relevance judgments). However, it does not allow us to test the idea that the reliability-relevance relationship is unidirectional (i.e., that manipulation of properties underlying relevance do not result in differences in assessed reliability). We are able to explore the latter issue in Experiment 2.

Table 1 presents participants' valuation judgments. The ANOVA results show a significant main effect of source competence on valuation effects ($F_{1,114} = 8.46$, $p < 0.01$). Consistent with the relevance assessments, participants receiving the measurement from a high competence source

indicated more positive valuation effects (mean of 34.24) than did those receiving the measurement from a low competence source (mean of 16.12). Time to sale and the interaction of source competence and time to sale are not significant in the ANOVA.

H2 examines the paths by which properties underlying reliability influence valuation judgments. This hypothesis is important for two reasons. First, although we find support for H1, it is possible that the relationship between manipulated source competence and assessed relevance is not meaningful in the sense of influencing valuation. Second, H2 verifies that users incorporate *both* assessed reliability and assessed relevance in their valuations (i.e., that decision usefulness is a function of both assessments). We test our theory underlying both H1 and H2 using a structural-equations model, as presented in Figure 3. Note that our full hypothesized model would also include a link from manipulated factors influencing relevance (time to sale) to assessed relevance. However, because the manipulation of time to sale did not influence participants' assessments of relevance, we omit this variable from the model.

FIGURE 3
Structural-Equations Model, Experiment 1
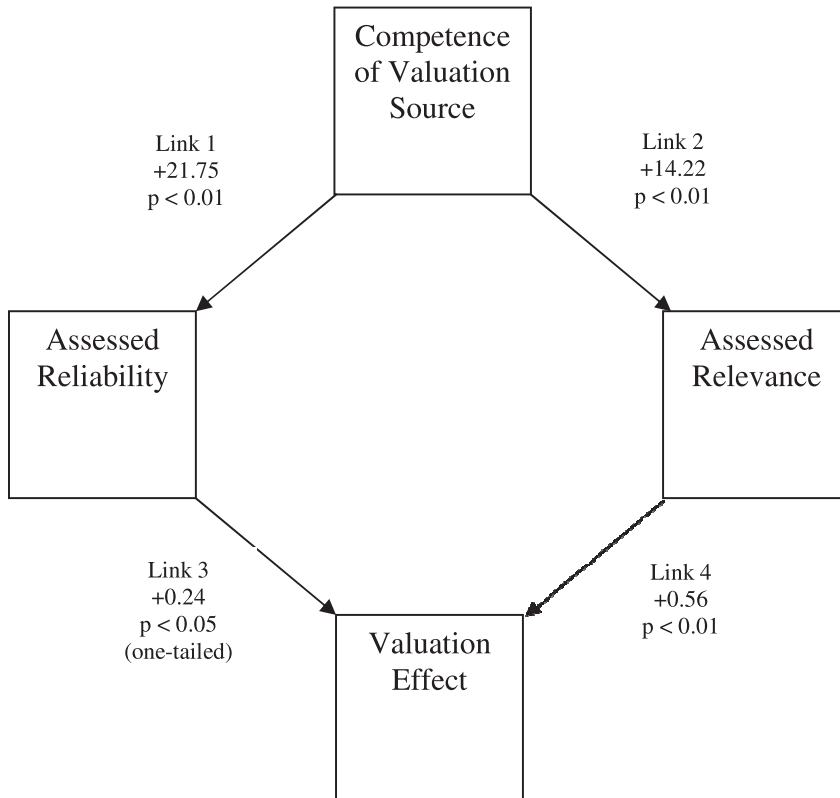
This figure summarizes results from Experiment 1. It illustrates that manipulated source competence, a property related to reliability, significantly influences users' assessments of both the reliability of the measurement and the relevance of current value information. It further demonstrates that users take their assessments of both reliability and relevance into account when determining how much weight to put on the reported information in valuing the firm.

Overall, the model fits the data well. The traditional $\chi 2$ test shows a good fit ($\chi 2 = 0.99$, p $= 0.32$), as do other standard fit measures.[11] Turning to the path coefficients, the relationship between our manipulated source competence variable and assessed reliability (Link 1) is positive and significant ($p < 0.01$), consistent with a successful manipulation of factors underlying reliability. The relationship between competence of the valuation source and assessed relevance (Link 2) is positive and significant ($p < 0.01$), consistent with H1. Finally, both assessed reliability (Link 3) and assessed relevance (Link 4) are, in turn, positively associated with the valuation effect ($p < 0.05$ and $p < 0.01$, respectively). Both of the indirect effects of source competence on valuation (through assessed reliability and through assessed relevance) are significant ($p < 0.01$) by the product of Z-Scores test (Holbert and Stephenson 2003), indicating that both paths are meaningful. These results support H2.

These Experiment 1 results support our theory that variations in source competence cause users to assess differences in reliability, as one would expect. Moreover, our results also support our theory by showing that reliability has an additional effect on relevance judgments. Both assessed reliability and assessed relevance, in turn, affect valuation judgments (i.e., their joint influence is decision useful). These results strongly suggest that properties underlying reliability have too much emphasis in valuation.

IV. EXPERIMENT 2

Overview

Experiment 1 provides evidence supporting our theory that variations in a property underlying reliability (i.e., source competence) influence the assessed relevance of current value information. Additionally, Experiment 1 supports the idea that source competence, as a property of reliability, influences users' valuation judgments by two distinct paths—through both assessed reliability and assessed relevance.

In our second experiment, we test whether our theory holds for a second property related to the reliability of a measurement—whether the measurement arises from a market transaction of a comparable item or from a model. We also attempt a stronger manipulation of factors underlying relevance in Experiment 2, to allow a test of the directionality of the relationship between relevance and reliability. The use of different manipulations of the underlying properties of reliability and relevance is also helpful in demonstrating the generalizability of our theory.

Participants, Design, and Variables

Ninety-nine M.B.A. students from a *BusinessWeek* top-25 school volunteered to participate in Experiment 2 in return for a small fixed payment. The participants had an average of 5.0 years of work experience. Seventy-two percent of participants had previously made investments in common stocks, and 97 percent either had invested or planned to invest in common stocks in the future. Participants had taken an average of 4.7 finance and 3.0 accounting classes. We paid participants $10 for their participation in this and an unrelated study.[12]

---

[11] The Comparative Fit Index, a measure of the proportion of improvement of the fit of our model to the null model, is 1.00, which is above the generally accepted minimum value of 0.95 (Byrne 2001). The Incremental Fit Index (1.00) is above the recommended minimum of 0.95 (Byrne 2001), and the Root Mean Square Error of Approximation (0.00) is below the recommended maximum of 0.01.

[12] Participants completed the unrelated study before participating in Experiment 2. We counterbalanced conditions of the other study with those of our experiments, alleviating the potential for carryover effects. Statistical tests reveal no systematic carryover effects.

As in Experiment 1, we asked participants to assume the role of a potential investor in a fictitious company. All were told that the company had acquired a plot of land for $41 million two years ago that was recently estimated to have a current value of $53 million. Similar to the first experiment, Experiment 2 makes use of a 2×2 full-factorial, between-participants design. We manipulated underlying properties of both the reliability of the current value measurement and the relevance of current value information. Specifically, we manipulated the valuation method used in deriving the land's current value as based on the market price of a comparable asset (e.g., Level 2 of SFAS 157 valuation hierarchy) or estimated from a model (e.g., Level 3). We did not use SFAS 157 terminology in our materials. The two levels correspond to high and low reliability and are referred to as "comparable" and "model," respectively.[13]

We manipulated the intended use of the asset as an underlying property of relevance.[14] At one level ("sell," corresponding to high relevance), the land is owned by a real estate development firm that holds the land for development and sale. At the other level ("use," corresponding to low relevance), the land is owned by a cosmetics manufacturing firm that holds it for purposes of building its own factory. Intended use incorporates time to sale (i.e., presumably the real estate developer will sell the land first, making its fair value potentially more timely) and separability of the asset from other operations (i.e., the separate value of land to be developed and sold is likely more relevant for the assigned valuation task than is the fair value of land that is part of productive capacity, because the latter would be valued "in use" rather than separately) (SFAS 157). Figure 4 illustrates the manipulations.

After receiving the information described above, participants made four assessments. The first three response measures are identical to the first three measures captured in Experiment 1 (i.e., relevance of the current value information [i.e., the construct], reliability of the number assigned to the land's current value [i.e., the measurement], and valuation effect of the information). Participants also assessed management trustworthiness using a 101-point scale ranging from "not at all trustworthy" (0) to "extremely trustworthy" (100).

Results

Preliminary Tests

As anticipated, reliability was assessed higher by those in the comparable conditions (mean of 64.06) as compared with those in the model conditions (mean of 42.60) ($F_{1,95} = 31.76$, p < 0.01). The manipulation of intended use (i.e., relevance) did not influence reliability judgments (not tabulated), a point we return to in later tests for this second experiment. That is, assessed reliability was not statistically different in the sell and use conditions (means 54.14 and 52.29, respectively, $F_{1,95} = 0.19$, p = 0.67). Further, the interaction of intended use and valuation method was not significant ($F_{1,95} = 0.21$, p = 0.65).[15]

---

[13] Following the argument of Song et al. (2010), this manipulation is a joint manipulation of bias and potential for error. Specifically, those authors contend that investors' downward valuation adjustment for reported Level 3 fair values is likely due to two factors: (1) a downward adjustment to the cash flow effect of the item due to potential bias on the part of management, and (2) an upward adjustment to the discount rate applied to the reported amount due to uncertainty or noise in the fair value estimate.

[14] The issue of whether intended use *should* normatively affect relevance judgments is subject to debate. For the purposes of this study, however, it is not necessary to establish whether individuals should or should not normatively react to intended use, but only that it will influence assessments of relevance and is unrelated to reliability.

[15] If users expected the real estate company to more reliably estimate the value of the land than the cosmetics company, then our intended use manipulation could be viewed as a manipulation of both relevance and reliability. However, the lack of any significant effects of intended use on judged reliability rules out this possibility.

FIGURE 4
Example of Materials Used in Experiment 2

Panel A: Sell, Model Condition

> In Year 1, WRSwann Co. purchased a plot of land for $41,000,000. WRSwann Co. is a real estate development firm that routinely purchases, develops, and sells land. The land is ideally suited for development, given its proximity to dense, high-income neighborhoods and transportation.
>
> In Year 3, WRSwann Co. places a current value on the land in the amount of $53,000,000.
>
> This current value number was generated internally by company management. Management based the current value number on an internally developed model (i.e., marked-to-model). There are no comparable plots of land upon which to determine the current value; hence, management used a model to determine the value.
>
> | Year 1 Value | Year 3 Value |
> |---|---|
> | $41,000,000 | $53,000,000 |

Panel B: Use, Comparable Condition

> In Year 1, WRSwann Co. purchased a plot of land for $41,000,000. WRSwann Co. is a cosmetics manufacturing company that plans to build its factory on the land. The land is ideally suited for the factory, given its proximity to an appropriate workforce and an already established distribution center.
>
> In Year 3, WRSwann Co. places a current value on the land in the amount of $53,000,000.
>
> This current value number was generated by an external valuation firm. This external valuation firm based the current value number on a comparable plot of land (i.e., adjacent to the one purchased by WRSwann, same size, and same location desirability). This comparable plot of land was recently sold.
>
> | Year 1 Value | Year 3 Value |
> |---|---|
> | $41,000,000 | $53,000,000 |

An ANOVA model for participants' ratings of management's trustworthiness (not tabulated) revealed a significant main effect of valuation method ($F_{1,95} = 13.83$, $p < 0.01$). Participants in the comparable conditions rated management trustworthiness significantly higher than did those in the model conditions (58.78 for comparable versus 49.52 for model). Neither intended use ($F_{1,95} = 0.01$, $p = 0.92$) nor the interaction of intended use and valuation method ($F_{1,95} = 1.22$, $p = 0.27$) is significant in the model. Collectively, these results indicate that our manipulation of the valuation method successfully influenced participants' views of the reliability of the asset's reported value.

Tests of Hypotheses

Descriptive (inferential) statistics for participants' assessments of relevance are presented in Panel A (Panel B) of Table 2. The ANOVA reveals a significant main effect of valuation method ($F_{1,95} = 8.44$; $p < 0.01$) such that participants in the comparable conditions judged the relevance of current value information (mean of 68.49) significantly higher than did those in the model conditions (mean of 53.94). Thus, underlying factors (e.g., error and bias) of the reliability of a current value measurement influence financial statement users' assessments of the relevance of current value information, supporting H1. Consistent with the findings from Experiment 1, this supports our theory that users base their evaluations of relevance on the more accessible reliability attribute. Support for participants' reliance on attribute substitution is also indicated by a positive and highly significant correlation of assessed reliability and relevance ($r = 0.48$, $p < 0.01$).

The ANOVA for relevance also reveals a significant main effect of intended use ($F_{1,95} = 4.97$; $p = 0.03$) and an insignificant interaction of valuation method and intended use ($F_{1,95} = 0.19$; $p =$

<div align="center">

TABLE 2

Relevance Assessments and Valuation Judgments, Experiment 2

</div>

Panel A: Mean Judgments (Standard Deviation)

Intended Use	Assessed Relevance Dependent Variable Valuation Method			Valuation Effect Dependent Variable Valuation Method		
	Model	**Comparable**	**Row Means**	**Model**	**Comparable**	**Row Means**
Use	49.48	61.75	55.49	15.00	29.96	22.17
	(24.13)	(25.49)	(25.32)	(37.54)	(29.22)	(34.30)
	n = 25	n = 24	n = 49	n = 25	n = 23	n = 48
Sell	58.40	74.96	66.68	23.24	49.40	36.32
	(24.84)	(24.29)	(25.71)	(41.46)	(32.24)	(39.06)
	n = 25	n = 25	n = 50	n = 25	n = 25	n = 50
Column Means	53.94	68.49		19.12	40.08	
	(24.65)	(25.51)		(39.37)	(32.04)	
	n = 50	n = 49		n = 50	n = 48	

Panel B: Analysis of Variance

Source	Assessed Relevance				Valuation Effect			
	df	**MS**	**Statistic**	**Two-Tailed p-value**	**df**	**MS**	**Statistic**	**Two-Tailed p-value**
Intended Use	1	3029	F = 4.97	0.03	1	4688	F = 3.71	0.06
Valuation Method	1	5141	F = 8.44	0.01	1	10341	F = 8.18	0.01
Intended Use × Valuation Method	1	113	F = 0.19	0.67	1	768	F = 0.61	0.44
Residual	95	609			95	1264		

Participants assessed the relevance of the current value information to their overall evaluation of the firm. We manipulated whether the current value was derived from a market transaction for a similar asset (e.g., level 2 of SFAS 157 valuation hierarchy) or from a model (e.g., level 3). We also manipulated the intended use of the asset as either "use" (the land is held for purposes of building the owner's factory) or "sell" (the land is held for development and sale). The dependent measures are described in the notes to Table 1.

0.67). The significant effect of intended use arises from participants in the sell conditions assessing higher relevance for current value information (mean of 66.68) than participants in the use conditions (mean of 55.49). This result indicates that we successfully manipulated participants' views of relevance independent of reliability.

The fact that our manipulation of intended use significantly influenced assessed relevance without affecting assessed reliability provides evidence in support of the unidirectional relationship between relevance and reliability. While factors underlying reliability affect both assessed reliability and assessed relevance, factors underlying relevance influence only assessed relevance. To further test the direct effect of these underlying properties on users' assessments of reliability and relevance as well as the indirect effect of these properties on users' valuation decisions, we employ structural equations analysis.

As shown in Table 2, an ANOVA model for participants' valuation judgments reveals significant main effects of intended use and valuation method. Valuation judgments were more positive when the current value was based on a comparable market transaction (mean of 40.08) versus a model (mean of 19.12) ($F_{1,95} = 8.18$, $p < 0.01$). They also were more positive when the company intended to develop and sell the asset (mean of 36.32) than when they were intending to use the asset in operations (mean of 22.17) ($F_{1,95} = 3.71$, $p = 0.06$). There was no evidence of an interaction of these two variables ($F_{1,95} = 0.61$, $p = 0.44$).

For H2, recall that we predict that factors underlying the reliability of a current value measurement will influence valuation via two paths—through users' assessments of the reliability of the measurement and through their assessments of relevance. Thus, we expect that the valuation method (i.e., comparable versus model) will influence both assessed reliability and assessed relevance and that each will separately influence valuation. We test H2 using a structural-equations model, as presented in Figure 5.

The model is similar to that tested in Experiment 1, except that we include in the model the intended use variable, which significantly influenced participants' relevance judgments. Before exploring the links in the model, we first test the overall fit using various indices. The traditional $\chi2$ test shows a good fit ($\chi2 = 1.92$, $p = 0.59$), as do other standard fit measures.[16] Thus, our model describes well the relations in the data.

Figure 5 presents path coefficients and statistical significance. The relationship between valuation method, with model coded as 0 and comparable coded as 1, and assessed reliability (Link 1) is positive and significant ($p < 0.01$), consistent with a successful manipulation of factors underlying reliability. Consistent with H1, the relationship between valuation method and assessed relevance (Link 2) is also positive and significant ($p < 0.01$). Further, both assessed reliability (Link 3, $p < 0.01$) and assessed relevance (Link 4, $p < 0.01$) are significantly associated with valuation, indicating that both influence decision usefulness. Finally, both indirect effects of valuation method on valuation—one through assessed reliability and the other through assessed relevance—are significant ($p < 0.01$, product of Z-scores test), indicating there are two distinct and meaningful paths by which factors underlying reliability affect valuation. These findings support H2.

Further analysis of our results shows that while properties underlying reliability influence relevance judgments, the reverse is not true. That is, properties associated with relevance do not influence reliability judgments. To see this, turn to our manipulation of intended use of the asset (with "use" coded as 0 and "sell" coded as 1). There, we see a significant positive relationship with assessed relevance (Link 6, $p = 0.02$), indicating a successful manipulation of factors underlying

---

[16] The Comparative Fit Index is 1.00, which is above the generally accepted minimum value of 0.95 (Byrne 2001); the Incremental Fit Index (1.01) is above the recommended minimum of 0.95 (Byrne 2001); and the Root Mean Square Error of Approximation is 0.00, which is below the recommended maximum of 0.01.

American Accounting Association

FIGURE 5
Structural-Equations Model, Experiment 2
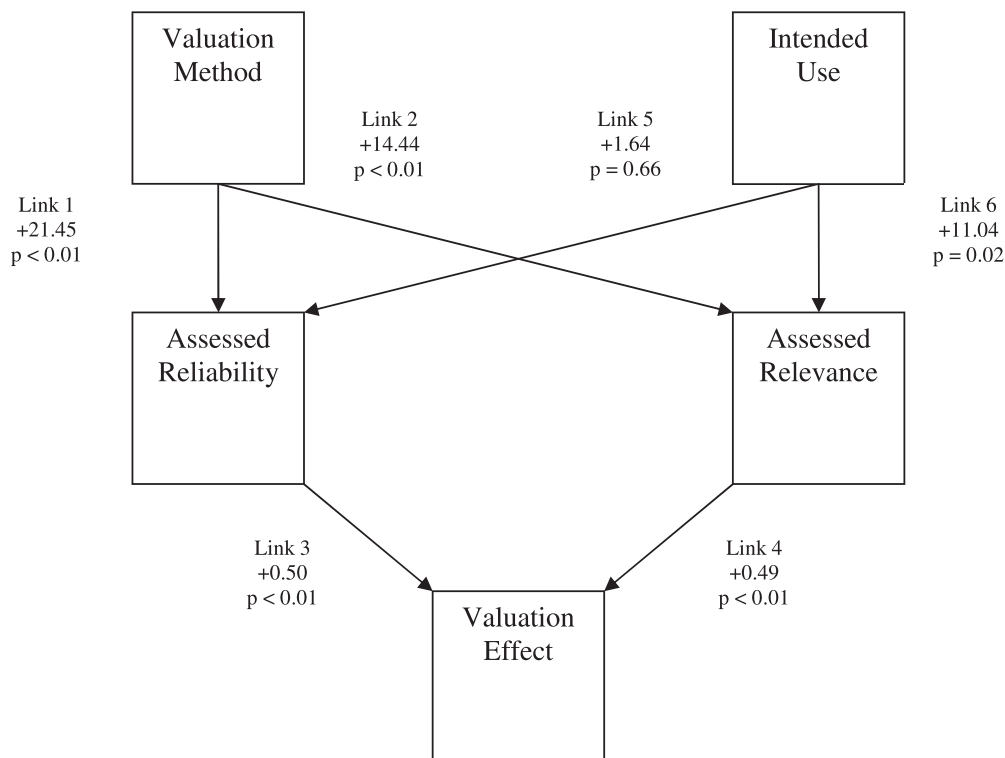
This figure summarizes results from Experiment 2. It illustrates that manipulated valuation method (i.e., comparable transaction versus model), a property related to reliability, significantly influences users' assessments of both the reliability of the measurement and the relevance of current value information. It demonstrates the directionality of this "confusion" between reliability and relevance by showing that the assets' intended use (use as a factory site versus held for sale), a determinant of relevance, significantly influences assessed relevance but does not influence assessed reliability. It further demonstrates that users take their assessments of both reliability and relevance into account when determining how much weight to put on the reported information in valuing the firm.

relevance. Most pertinent to our purposes, the link between the intended use variable and assessed reliability is insignificant (Link 5, $p = 0.66$), indicating that the reliability-relevance relationship is strictly unidirectional, as we posit.

An alternative specification of this model includes a link from assessed reliability to assessed relevance instead of the link from valuation method to assessed relevance. This model is computationally equivalent to the model we test for Experiments 1 and 2, and so fit measures are identical. Both our model and the alternative model are consistent with our theory, as they imply that evaluations of relevance rely on the more accessible reliability attribute.

Ruling Out an Alternative Substitute Attribute

Our theory predicts that users have difficulty assessing relevance, and so will substitute the reliability attribute for relevance. We believe reliability to be the most likely candidate for attribute

substitution. Reliability is a basic property of measurement that prior research indicates individuals assess relatively accurately, and thus it is highly accessible. Results of our experiments support this theory. However, an alternative possibility is that users substitute decision usefulness for relevance. That is, they may use the label "relevance" to capture the idea of decision usefulness, which includes the effects of both relevance and reliability (FASB 1980, 2010). We conduct additional analyses to rule out this possibility.

First, recall that the path analysis in Experiment 2 shows that manipulations of properties underlying relevance and reliability affect participants' valuation judgments.[17] If participants used the relevance label to capture decision usefulness, then we would observe the assessed relevance measure capturing *all* effects of factors underlying both relevance and reliability on our valuation dependent measure. However, our analysis does not support this result. Instead, as seen in Figure 5 (Link 3), assessed reliability has a significant effect on valuation above and beyond the impact of assessed relevance on valuation. As a further test of this same idea, we estimated a model identical to that in Figure 5, except that we omitted Link 3. This model also tests the idea that judged relevance is used as a proxy for decision usefulness. The alternative model is a poor fit to the data, and a nested model test shows that our model fits the data significantly better ($\chi_1^2 = 8.482$, p < 0.01). These results indicate that our participants did not consider relevance as a substitute for decision usefulness.

V. ROBUSTNESS TESTS

We conducted two additional experiments to test the robustness of the findings from our main experiments. First, to determine whether these results hold even if participants are provided with the FASB's definitions of relevance and reliability, we undertook a third experiment with the same design as Experiment 2 with a separate group of 94 participants from a similar population (M.B.A. students with mean work experience of 8.4 years, average of 3.5 finance and 3.1 accounting classes, 81 percent had previously invested in common stocks, and 89 percent planned to do so). Our results are similar to those reported for Experiments 1 and 2. That is, the manipulation of valuation method (i.e., market price of a comparable asset versus model estimate) influenced participants' assessments of relevance ($F_{1,90} = 7.38$; p < 0.01) despite the prominent display of FASB's definitions in the materials. Participants assessed relevance as higher in the comparable conditions (i.e., high reliability; mean of 75.60) than in the model conditions (i.e., low reliability; mean of 62.95).

Second, to ensure that participants interpreted the relevance question as referring to the general construct of current value information (versus a specific measurement), we conducted a fourth experiment based on Experiment 2 with a separate group of 47 M.B.A. student participants (mean work experience of 8.4 years, average of 3.8 finance and 3.6 accounting classes, 72 percent had previously invested in common stocks and 91 percent planned to do so). We used the model and comparable conditions for the high relevance condition (i.e., sell) from Experiment 2, with one change. We asked about the relevance of "current value information" rather than "*the* current value information." Again, we replicated our Experiment 2 findings. Assessments of relevance were significantly higher in the comparable condition than in the model condition (68.5 versus 50.1, $F_{1,45} = 6.68$; p $= 0.01$).

VI. CONCLUSIONS

In this study, we experimentally investigate the hypothesis that factors underlying the reliability of a fair value measurement influence financial statement users' assessments of the

---

[17] We did not conduct this test for Experiment 1 as it is less meaningful there, because the manipulation of time to sale (i.e., relevance) showed an insignificant main effect on our dependent measures.

relevance of fair value. Such an assessment is inconsistent with the qualities of relevance and reliability being independent constructs. We base our tests on attribute substitution theory, which predicts that this relationship arises because financial statement users can access more readily the reliability attribute than the relevance attribute. Thus, when asked to assess relevance, financial statement users substitute an easier question—one about reliability—in forming a response. Our results support our hypothesis, revealing that factors underlying the reliability of a fair value measurement—namely, the competence of the source of the measurement (Experiment 1) and whether the measurement is based on a market transaction for a comparable item or a model calculation (Experiment 2)—strongly influence users' assessments of the relevance of fair value information. Attribute substitution further predicts that this relationship will be unidirectional. That is, when asked to assess reliability, financial statement users will do so rather than substitute the relevance attribute. Our results also support this assertion.

We also demonstrate that factors underlying the reliability of a measurement affect users' valuation judgments through two paths: (1) their influence on assessed reliability, and (2) their influence on assessed relevance. Both of these paths are highly significant in our experiments. While these results are highly suggestive of an over-weighting of factors underlying reliability in valuation, without a normative benchmark we cannot be certain that this is occurring. Future research can draw on our insights to further explore this issue.

Future research also could examine boundary conditions on attribute substitution for relevance and reliability. That is, does what we observed in a fair value context hold in other financial reporting situations? In many financial reporting contexts, users receive a measurement possessing a given level of reliability and relevance, and relevance is typically more difficult to assess than reliability. Thus, we expect the results we observed to obtain generally in these contexts. However, we caution readers that our results may not generalize to other financial reporting contexts in which relevance is made particularly salient, and is therefore more accessible, or in contexts in which reliability is particularly difficult to assess. In such cases, attribution substitution either would not occur or would take a different form than we observe.

Future research could further explore the extent to which financial statement users are aware of the attribution substitution. Recall that attribute substitution theory indicates that individuals are generally unaware of the substitution they are making. Although we did not design our tests to investigate the role of awareness, we nevertheless report results that are consistent with theory, with no obvious alternative explanation for the observed effects. Future research could address the awareness issue by employing, for example, verbal protocol analysis to determine if users are aware of their use of factors underlying reliability when assessing relevance. If this research shows that individuals are aware of the substitution, additional research also could investigate whether participants purposefully use reliability information when judging relevance. We believe such intent is unlikely, since the substitution effect persists when we gave participants definitions of the terms. The fact that this substitution occurs is important regardless of awareness or intentionality, given that it is at odds with a view of relevance and reliability as independent constructs.

This study has important implications for standard setters, firm management, and financial statement users. First and foremost, our findings suggest that financial statement users are likely to make errors in judgments. Because these errors are apparently driven by the difficulties of assessing relevance and the resulting use of a heuristic, training may not be effective in correcting the problem. For example, even individuals with extensive statistical training ignore base rates and rely on representativeness when making intuitive judgments (Tversky and Kahneman 1974, 1130).

The phenomena that we document are also likely to interfere with standard setters' efforts to expand the use of fair value in financial reports. Observers focusing on unreliable fair value measurements have argued that fair values are irrelevant, as a result. We note that this is akin to asserting that the outdoor temperature is not relevant to one's decision whether to wear an overcoat

simply because the thermometer at hand is unreliable. A better understanding of the underlying issue—users' inability to assess relevance independently from reliability—indicates that standard setters may wish to focus on addressing how to communicate the relevance of the particular constructs separate from their efforts to improve the reliability of their measurements.

REFERENCES

American Bankers Association. 2009. What people are saying about fair value. Available at: http://www. aba.com/Press+Room/PR_Accounting_FVAQuotes.htm#Steve

Byrne, B. M. 2001. *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Chapman, G. B., and E. J. Johnson. 2002. Incorporating the irrelevant: Anchors in judgments of belief and value. In *Heuristics and Biases*, edited by T. Gilovich, D. Griffin, and D. Kahneman, 120–138. New York, NY: Cambridge University Press.

Dick, A., D. Chakravarti, and G. Biehal. 1990. Memory-based inferences during consumer choice. *Journal of Consumer Research* (17): 82–93.

Duncan, K., and K. Moores. 1988. Usefulness of CCA information for investor decision making: A laboratory experiment. *Accounting and Business Research* 18 (70): 121–132.

Dye, R. A., and S. S. Sridhar. 2004. Reliability-relevance trade-offs and the efficiency of aggregation. *Journal of Accounting Research* 42 (1): 51–88.

Epley, N., and T. Gilovich. 2002. Putting adjustment back in the anchoring and adjustment heuristic. In *Heuristics and Biases*, edited by T. Gilovich, D. Griffin, and D. Kahneman, 139–149. New York, NY: Cambridge University Press.

Epley, N., and T. Gilovich. 2006. The anchoring-and-adjustment heuristic: Why the adjustment is insufficient. *Psychological Science* 17 (4): 311–318.

Financial Accounting Standards Board (FASB). 1980. *Qualitative Characteristics of Accounting Information*. Statement of Financial Accounting Concepts No. 2. Norwalk, CT: FASB.

Financial Accounting Standards Board (FASB). 2006. *Fair Value Measurements*. Statement of Financial Accounting Standards No. 157. Norwalk, CT: FASB.

Financial Accounting Standards Board (FASB). 2010. *Conceptual Framework for Financial Reporting*. Chapter 3, "Qualitative Characteristics of Useful Financial Information." Statement of Financial Accounting Concepts No. 8. Norwalk, CT: FASB.

Hague, I., and D. Willis. 1999. Old price or new? *CA Magazine* (February): 47–49.

Herzog, T. 1997. *Research Methods and Data Analysis in the Social Sciences*. Upper Saddle River, NJ: Prentice Hall.

Hirst, E. 1994. Auditors' sensitivity to source reliability. *Journal of Accounting Research* 32 (Spring): 113–126.

Hirst, E., L. Koonce, and J. Miller. 1999. The joint effect of management's prior forecast accuracy and the form of its financial forecasts of investor judgment. *Journal of Accounting Research* 37 (Supplement): 101–124.

Holbert, R. L., and M. T. Stephenson. 2003. The importance of indirect effects in media effects research: Testing for mediation in structural equation modeling. *Journal of Broadcasting and Electronic Media* 47 (4): 556–572.

International Accounting Standards Board (IASB). 2010. *The Conceptual Framework for Financial Reporting*. London, U.K.: IASC Foundation.

Jennings, R. 1987. Unsystematic security price movements, management earnings forecasts, and revisions in consensus analyst earnings forecasts. *Journal of Accounting Research* 25 (1): 90–110.

Joyce, E. J., R. Libby, and S. Sunder. 1982. Using the FASB's qualitative characteristics in accounting policy choices. *Journal of Accounting Research* 20 (2): 654–675.

Jurney, S. 2008. *How Do Nonprofessional Investors Understand and Use Relevance and Reliability of Financial Information?* Dissertation: The University of Oklahoma.

Kadous, K., M. Mercer, and J. Thayer. 2009. Is there safety in numbers? The effects of forecast accuracy and forecast boldness on financial analysts' credibility with investors. *Contemporary Accounting Research* 26 (3): 933–968.

Kahneman D. 2002. Nobel prize lecture: Maps of bounded rationality: A perspective on intuitive judgment and choice. In *Nobel Prizes 2002: Nobel Prizes, Presentations, Biographies, and Lectures*, edited by T. Frangsmyr, 416–499. Stockholm, Sweden: Almqvist and Wiksell Int.

Kahneman, D., and S. Frederick. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, and D. Kahneman, In, *Heuristics and Biases: The Psychology of Intuitive Judgment*, pp. 49–81. New York: Cambridge University Press.

Kahneman, D., and S. Frederick. 2005. A model of heuristic judgment. In *The Cambridge Handbook of Thinking and Reasoning*, edited by K. Holyoak, and R. Morrison, 267–294. New York, NY: Cambridge University.

Koonce, L., K. K. Nelson, and C. Shakespeare. 2011. How do investors assess the relevance of fair value for financial instruments? *The Accounting Review* 86 (6): 2075–2098.

Li, V., T. Shevlin, and D. Shores. 2010. *Revisiting Earnings and Accrual Quality*. Working paper, University of Washington.

Maines, L. A., and J. M. Wahlen. 2006. The nature of accounting information reliability: Inferences from archival and experimental research. *Accounting Horizons* 20 (4): 399–425.

McCaslin, T. E., and K. G. Stanga. 1983. Related qualities of useful accounting information. *Accounting and Business Research* 14: 35–41.

Mercer, M. 2004. How do investors assess the credibility of management disclosures? *Accounting Horizons* 18 (3): 185–196.

Mueller, C., and L. Kirkpatrick. 1995. *Federal Rules of Evidence*. Boston, MA: Little, Brown and Company.

Schipper, K. 2003. Principles-based accounting standards. *Accounting Horizons* 17 (1): 61–72.

Schipper, K. 2005. *Lecture on Recent Activities at the FASB*. American Accounting Association Annual Conference, San Francisco, CA: California.

Song, C. J., W. B. Thomas, and H. Yi. 2010. Value relevance of FAS 157 fair value hierarchy information and the impact of corporate governance mechanisms. *The Accounting Review* 85 (4): 1375–1410.

Strack, F., and T. Mussweiler. 1997. Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology* 73: 437–446.

Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185: 1124–1131.

Williams, P. 1996. The relation between a prior earnings forecast by management and analyst response to a current management forecast. *The Accounting Review* 71 (1): 103–113.

Wilson, T. 2002. *Strangers to Ourselves*. Cambridge, MA: The Belknap Press of Harvard University Press.