

# Documentation Requirements and Quantified versus Qualitative Audit Risk Assessments

M. David Piercey

**SUMMARY:** The “not documented, not done” requirements of Public Company Accounting Oversight Board (PCAOB) Auditing Standard No. 3 substantially increased auditors’ obligations to document their risk assessments. This study examines a potentially unintended consequence of such a requirement on auditors who have pressure to reach lenient, client-preferred risk assessments. Because documentation requirements potentially expose auditors’ lenient judgments to more *ex post* scrutiny (e.g., regulator inspection, litigation), one would ordinarily *not* expect that adding them would cause auditors with client pressures to become *more* lenient. However, I expect that adding documentation requirements leads auditors who assess risk in *qualitative* (rather than quantified) terms to engage more in a specific word-smithing strategy that is shown by prior research to help rationalize reaching more lenient audit conclusions. Thus, even though documentation potentially exposes more lenient judgments to scrutiny, I show that auditors assessing risk in qualitative terms respond to this pressure by rationalizing their lenient assessments even more vigilantly. This leads to more lenient judgments, ironically, as a result of adding the documentation requirement. Adding documentation requirements does not have this effect on quantified risk assessments. Prior research also suggests that auditors typically assess risk in words. Thus, under common conditions, the PCAOB’s documentation requirements may have unintended effects, with adverse implications for audit effectiveness contrary to their regulatory intent.

**Keywords:** quantified versus qualitative audit risk assessments; elastic re-definition; documentation requirements.

**Data Availability:** Contact the author.

---

*M. David Piercey is an Assistant Professor at the University of Massachusetts Amherst.*

---

This manuscript is based on part of my dissertation at the University of Illinois. I thank my dissertation committee, David Budescu, Mark Peecher (chair), Rachel Schwartz, and Ira Solomon, for their helpful comments. I also thank Ken Trotman (editor), two anonymous reviewers, Efrim Boritz, Bill Dilla, Anne Farrell, Mike Gibbins, Frank Hodge, Patrick Hopkins, Karim Jamal, Jane Kennedy, Tom Kida, Ron King, Natalia Kotchetova, Susan Krische, Joan Luft, Bill Messier, Kimberly Moreno, Mark Nelson, Derek Oler, Jamie Pratt, Jim Smith, Hun-Tong Tan, Chris Wolfe, participants in workshops at the University of Alberta, Georgia State University, Indiana University, University of Massachusetts, Michigan State University, University of Washington, Washington University, University of Waterloo, and in the AAA Annual Meeting. I gratefully acknowledge the financial assistance of the Deloitte & Touche Doctoral Fellowship. I also thank the AAA, which awarded an earlier version of this manuscript the Accounting, Behavior, and Organizations Outstanding Dissertation Award.

Editor’s note: Accepted by Ken Trotman.

*Submitted: May 2011*

*Accepted: August 2011*

*Published Online: November 2011*

## INTRODUCTION

In recent years, the requirements on auditors to document their auditing judgments have expanded substantially. These increased requirements are one of many reforms prompted by widespread concerns that auditors are too influenced by pressures to placate their clients. At the end of 2004, the U.S. Public Company Accounting Oversight Board (PCAOB) implemented Auditing Standard No. 3—*Audit Documentation* (PCAOB 2004). As with any auditing standard, an implicit intention of the new “not documented, not done” standard is to improve auditor judgment, including auditor objectivity and professional skepticism. The standard requires that auditors now document all of their judgments in such detail that a reviewer can reconstruct all parts of their judgments (PCAOB 2004). Since this standard was released alongside new regulatory inspections of public auditing firms, the standard also provides regulatory inspectors, plaintiffs, and others with more material for scrutinizing auditors’ judgments *ex post* (e.g., PCAOB 2008).

This study examines the effects of adding such documentation requirements on auditors with a preference to reach lenient, client-preferred audit conclusions. The PCAOB (2008) reports that, based on its first four years of inspections, serious concerns remain as to the objectivity and independence of auditors from client pressures to reach lenient auditing conclusions.

Ordinarily, one would *not* expect that adding documentation requirements would make auditors with client pressures *more* lenient. After all, adding documentation requirements simply adds more potential for *ex post* scrutiny of lenient auditing judgments. However, this study finds evidence that documentation requirements can have this surprising effect. Professional standards give auditors the option of quantifying audit risks (i.e., putting them into numbers) or expressing them using qualitative words (e.g., International Federation of Accountants [IFAC] 2009; PCAOB 2007; American Institute of Certified Public Accountants [AICPA] 2008). I find no significant differences between auditors’ documented and undocumented judgments when they assess risk in numbers. However, when auditors assess risk in words, their *documented* qualitative judgments are, ironically, more lenient than their *undocumented* qualitative judgments.

This finding is surprising, since documentation requirements simply open qualitative risk assessments to more potential scrutiny. However, building upon recent research, I predict and find that adding documentation requirements leads auditors who assess risk in words to engage in a particular word-smithing strategy. Piercey (2009) shows that auditors can rationalize a lenient audit risk assessment by retaining the option to act as though the qualitative low-risk phrase really means something else—and could mean many things—in the event of adverse outcomes. This behavior manifests itself in a specific empirical pattern that is detectable with a specific empirical test. In this study, I show that documentation requirements increase this behavior. Thus, even though documentation requirements potentially expose auditors’ lenient qualitative risk assessments to *ex post* scrutiny, this can, ironically, lead auditors to “defensively bolster” (Tetlock 1999) lenient auditing judgments more vigilantly, leading to lower assessed risk levels.<sup>1</sup> Kachelmeier and King (2002) call for experiments that test potential unintended consequences of accounting standards, as well as provide a better understanding of why those unintended consequences occur.

Supplemental findings in this study suggest that these effects are due to the way that documentation requirements affect how humans naturally assess risk, rather than undesirable behaviors that *auditors and only auditors* exhibit. Thus, regulators should understand how their regulations can interact with the fundamental psychology of how humans assess risk, in potentially

<sup>1</sup> Piercey (2009) demonstrates this word-smithing behavior and shows that auditors can use it to rationalize more lenient auditing judgments. However, that study does not consider the effects of documentation requirements either on increasing this behavior or on lowering assessed audit risk levels. Nor does that study consider the potential for PCAOB inspection. They are the focus of the current study.

unintended ways (e.g., Kadous et al. 2003). Such an understanding can help them better predict when and why their standards may create unintended consequences. This study contributes to theory with a better understanding of when and why auditors with client pressures reach more lenient audit conclusions. Auditors generally prefer to assess risk in words rather than numbers (PCAOB 2007; Simon 2002; Olson and Budescu 1997; Martinov and Roebuck 1998; Laswad and Mak 1997; McFadgen 1994). Thus, documentation requirements can cause audit risk assessments to become more lenient under certain conditions. When auditors assess audit risks more leniently, they perform fewer tests, rely more readily on internal controls, and collect less substantive audit evidence to reach an unqualified opinion of a client's financial statements (AICPA 2008; IFAC 2009). This suggests a potentially unintended consequence of the PCAOB's (2004) auditing standard, and should help inform regulators and others about the relative costs and benefits of the "not documented, not done" auditing standard. This study can also help auditors with client pressures better understand what factors can influence their judgments in ways that they may not anticipate.

In the remainder of this paper, I develop theory and hypotheses, describe the experimental method and results, and discuss conclusions and implications for practice, theory, and research.

## THEORY AND HYPOTHESES

### Client Pressures and Motivated Reasoning

During an audit, auditors continuously form and revise their beliefs about the probability of material misstatements in a client's financial statement accounts, and of material weaknesses in its internal controls (Bell et al. 2005). Professional standards refer to these probabilities as audit risks (AICPA 2008). Auditors assess such risks both formally (e.g., during initial planning of the audit) and informally throughout the audit (e.g., assessments and reassessments of the risk of material misstatement that appear in work papers).

Auditors have pressures from a variety of sources, some providing them with incentives to be skeptical and conservative (e.g., regulatory and litigation pressures), some providing them with incentives to be lenient (e.g., client pressures) (Nelson 2006).<sup>2</sup> While it is unlikely that auditors experience constant, over-arching pressures to be lenient, there is likely to be variance in client pressures, over time, across audit clients, and across auditors. Researchers argue that client pressures remain an ongoing concern in practice, even since passage of the Sarbanes-Oxley Act (SOX) and similar regulations (e.g., see Moore et al. 2006; Nelson 2006; Bazerman et al. 2006). Similarly, the PCAOB (2008) reports that independence from client pressures remains a serious ongoing concern, based upon its inspections of audit firms. Thus, while many recent regulations are intended to reduce independence risks, Peecher et al. (2010, 1,764) argue that pressures "to reach client-preferred auditing judgments . . . continue to put audit effectiveness at risk, even post-SOX." In this study, I consider a setting in which auditors have client pressures.

Prior accounting research has shown that auditors' assessments of their clients' financial statements are influenced by their clients' preferences, despite professional requirements to maintain independence and objectivity (e.g., AICPA 1988). When client pressures motivate auditors to reach client-preferred audit conclusions, auditors can use biased information search, interpretation of audit evidence, and word-smithing to justify reaching more lenient auditing

<sup>2</sup> As used in prior studies, the term "client pressures" refers to pressures to reach lenient, client-preferred conclusions, whether auditors receive those pressures directly from the client or through a supervisor (e.g., Peecher et al. 2010).

judgments (e.g., [Brown et al. 1999](#); [Kadous et al. 2003](#); [Peecher 1996](#); [Peecher et al. 2010](#); [Piercey 2009](#); [Wilks 2002](#)).

However, individuals engaged in such motivated reasoning generally do not just overtly bias their judgments in ways that would be blatant to others or even themselves. Rather, they reach a preferred conclusion as if building a justifiable case in their minds for defending their judgments as objective to an outside observer ([Kunda 1990](#)). The more auditors can mentally maintain this “illusion of objectivity,” the more likely they are to reach client-preferred audit risk assessments ([Pyszczynski and Greenberg 1987](#), 302).

### Motivated Reasoning and Quantified versus Qualitative Audit Risk Assessments

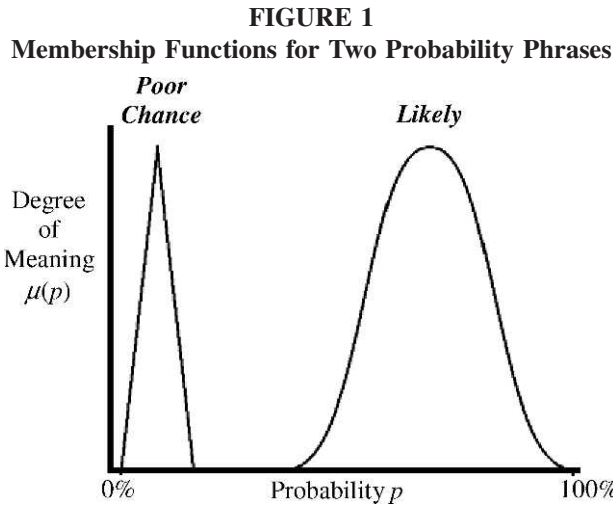
Professional auditing standards explicitly give auditors the choice of assessing audit risks either quantitatively or qualitatively, entirely at the option of the auditor ([AICPA 2008](#), 258; [IFAC 2009](#), 19). In other cases, standards directly encourage a preference for qualitative assessments (e.g., [PCAOB 2007](#), 419; [Simon 2002](#)). For example, qualitative risk assessments express the likelihood of material misstatement as a phrase (e.g., “a reasonable possibility”) ([PCAOB 2007](#)), while quantified risk assessments can express it in percentages ([AICPA 2008](#); [IFAC 2009](#)).

Unlike a percentage, a qualitative audit risk assessment (e.g., “somewhat possible”) implies a variety of different likelihoods of material misstatement, and some more than others. Psychology theory models this characteristic of qualitative probability expressions using *membership functions* ([Budescu et al. 2003](#); [Wallsten et al. 1986](#)). As Figure 1 shows, membership functions provide a rich representation of the meaning of a qualitative risk assessment by showing on the vertical axis the extent to which the phrase implies (or does not imply) various probabilities along the horizontal axis.

Also, unlike a percentage, a qualitative risk assessment has several semantic characteristics. Two membership function characteristics capture the probability “level” conveyed by a phrase: the *Peak* of the membership function is the probability that a phrase describes best, while the *Center* (i.e., the first moment of the membership function, analogous to the mean of a distribution) is the “average” probability implied by the phrase. The second moment (analogous to the variance of a distribution) is the *Vagueness* of a phrase, with narrower membership functions indicating a more precise specification of risk, and wider membership functions indicating a more vague risk assessment (Figure 1).

The third moment of a membership function is its *Skewness*. Two phrases may convey the same probability level, but have entirely different emphases. For example, the phrases “there is a chance” and “there is only a chance” convey similar probability levels. However, the former emphasizes that an event may occur, while the latter emphasizes that it may not ([Teigen and Brun 2000, 1995](#); [Moxey and Sanford 2000](#); [Teigen 1988](#)). [Budescu et al. \(2003\)](#) show that the skewness of a phrase’s membership function measures this relative emphasis, even after controlling for the probability level conveyed by the phrase. For example, a phrase can be skewed with its “tail” assigned to lower probabilities below its center, and higher membership values assigned to probabilities above its center. This skewness indicates that the risk assessment is emphasizing the higher probabilities within its connotation relatively more, while de-emphasizing its lower probabilities (i.e., emphasizing that an event may occur and de-emphasizing that it may not). This relative emphasis or de-emphasis of higher or lower probabilities within phrases influences subsequent judgment and decision-making, even when those phrases convey the same probability level (e.g., [Teigen and Brun 1999](#)).

Despite auditors’ widespread use of qualitative risk assessments, and the multiple ways that qualitative risk assessments differ from quantified assessments ([Budescu and Wallsten 1995](#); [Budescu et al. 2003](#)), research has only recently begun to use membership functions to better

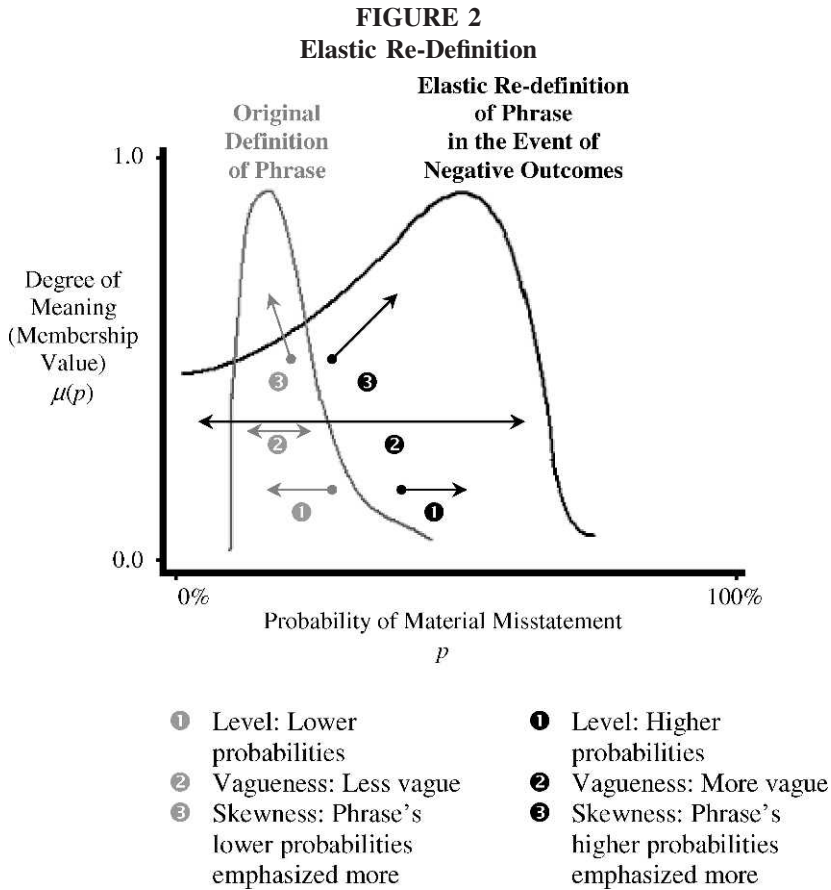


Source: Wallsten et al. (1986). Figure 1 used with permission from the American Psychological Association.

understand qualitative judgments in accounting contexts. Piercey (2009) uses this theory to study auditors' pressures to reach client-preferred auditing conclusions. He shows that auditors trying to justify a lenient qualitative audit risk assessment can retain the option, in case of negative outcomes, of acting as though the phrase has really meant something more skeptical all along and, for that matter, means many things. This optional re-definition of the assessment (if it is ever needed) emphasizes this alternative meaning more, but retains some de-emphasized connection to the original, lenient sense of the audit risk assessment. This de-emphasized connection can help maintain a mental illusion of justifiability, helping an auditor rationalize, for now, more lenient audit judgments.

For example, an auditor with client pressures to be lenient could assess the probability of material misstatement with a relatively low-probability phrase (e.g., "somewhat possible"). An auditor may feel that (s)he could justify this lenient judgment to an outside observer (Kunda 1990) if his or her judgment were ever subjected to *ex post* scrutiny. One can easily rethink how the phrase "somewhat possible" could really mean something different (like a higher probability warning that material misstatement is indeed *somewhat possible*) and, in fact, reconsider how it could even have a variety of meanings. This rethought definition would tend to emphasize how the phrase could actually mean something quite different than the original, lenient connotation. On the other hand, this re-definition would also likely retain some connection to the original, lenient risk assessment (rather than break completely from the original definition), so as to maintain some illusion of objectivity. Yet, this connection will be de-emphasized in the phrase's alternative meaning in case of adverse outcomes.

Figure 2 illustrates this specific type of motivated reasoning and the way it manifests itself empirically. The membership function on the left represents an original, lenient characterization of risk (e.g., "somewhat possible"), conveying lower probabilities of material misstatement. The membership function on the right represents how an auditor could mentally recast that same phrase as though it really conveys higher probabilities of material misstatement along the right side of the horizontal axis. The membership function on the right is also more vague (i.e., wider), as if the phrase now means *more things*. Finally, the membership function on the right is also now skewed to



Source: [Piercey \(2009\)](#). Figure 2 used with permission from Elsevier.

emphasize the new meaning, but retains some common membership in its skewed “tail” with the original meaning of the phrase, albeit now de-emphasized. [Piercey \(2009, 333\)](#) calls this specific three-part justification strategy “elastic re-definition” because it shows how a phrase can be re-defined, but in a way that reaches back elastically (cf. [Hsee 1996](#)) to the original meaning of the phrase, helping maintain an illusion of objectivity.<sup>3</sup>

<sup>3</sup> Other contexts provide vivid examples of elastic re-definition. The phrase “person of interest” entered into common usage after the 1996 Olympics bombing in Atlanta, and the characterization of a subsequently cleared security guard who discovered the bomb as a “suspect” (which drew a large public controversy and lawsuits of several media outlets). [Shaw \(2006\)](#) notes that since then, “person of interest” has evolved into a mere proxy for “suspect” in reports that quite unambiguously use the phrase to imply suspicion. However, the phrase seems to maintain an alternative re-definition, if needed, to act as if the phrase only implies some vague connection to the crime (with suspicion de-emphasized) in case the subject of the assessment is subsequently cleared of wrongdoing. The sense that the phrase can be elastically re-defined may provide the authors of such reports with more comfort, for now, to imply suspicion. In fact, elastic re-definition is likely a fairly automatic behavior that people develop across a wide variety of contexts in their daily lives. As a result, this is likely a behavior that auditors and nonauditors alike exhibit as human beings, rather than an undesirable behavior unique to auditors (cf. [Sedor 2002](#); [Libby et al. 2002](#); [Peecher and Solomon 2001](#)).

The possibility of mentally re-defining a qualitative risk assessment in this way potentially allows auditors, for now, to justify reaching more lenient assessments of the client. In contrast, quantified risk assessments (e.g., “30 percent”) are precisely specified and much more difficult for an auditor to plausibly re-define as meaning something else while maintaining an illusion of objectivity (e.g., it is harder to justify to oneself how an originally assessed 30 percent could really be taken to mean 70 percent, even in the event of adverse outcomes).

### Documentation Requirements

“Not documented, not done” audit risk assessments require auditors to document virtually every single judgment in such detail as to facilitate a regulator’s inspection of their judgments (PCAOB 2004). In this study, I consider auditors who simultaneously face client pressures to be lenient and a risk of *ex post* scrutiny. Rather than these two incentives merely canceling out, auditors must continue to deal with client pressures while simultaneously coping with a risk of inspection.

In this study, I posit that documentation requirements make elastic re-definition more likely to occur as a rationalization strategy. Specifically, requiring auditors to document every single material audit judgment in such detail as to facilitate potential review of their judgments prompts auditors to “choose their words carefully,” or word-smith their qualitative audit risk assessments as they form them. Consciously or unconsciously (Lerner and Tetlock 1999), this would include contemplating the elasticity of a lenient phrase and how it could mean something less lenient, mean more things, etc., in case of adverse outcomes, as depicted in Figure 2. On the other hand, when auditors do not have such documentation requirements, their assessments are not directly subject to review. There is less need to contemplate how one’s words could be elastically re-defined as meaning something more skeptical and vague in the event of a regulatory inspection.

Ordinarily, one would not expect that adding such documentation requirements to auditors’ client pressures would cause their judgments to become more *lenient*. However, the more that individuals engage in motivated reasoning, and sense that they could justify their judgments to an outside observer, the greater the illusion of objectivity that they feel, and the more likely they are to reach preferred conclusions biased by pressures to reach a preferred conclusion (Kunda 1990). For example, Kadous et al. (2003) demonstrate that asking auditors to provide seemingly objective evaluations of their clients’ financial reporting quality actually amplifies the effects of their preferences for client-preferred judgments. Peecher et al. (2010) demonstrate that when supervising auditors craft and provide outwardly objective guidance to their subordinates, this elevates the supervisors’ sense of justifiability, and they form auditing judgments even more influenced by their client pressures to be lenient.

As Lerner and Tetlock (1994) and Tetlock (1999) note, individuals who face conflicting accountability pressures engage in “defensive bolstering” of the positions preferred by the more immediate source of pressure. Note that client pressures to be lenient are generally more immediate and certain to auditors than an increase in the deferred chance of future regulatory inspection. Elastic re-definition is a specific word-smithing strategy that can enable auditors to accommodate more immediate client pressures to be lenient, while maintaining an illusion of justifiability in case of a potential future scrutiny of their judgments. If documentation requirements increase elastic re-definition, then they ultimately prompt auditors’ more vigilant rationalization of their lenient audit risk assessments as justifiable and defensible. Thus, even though documentation requirements potentially expose auditors to *ex post* scrutiny, this may have a truly ironic effect when auditors are expressing their auditing judgments in qualitative words: they may engage in more defensive bolstering of their lenient auditing judgments. Auditors facing the same client pressures—but without documentation requirements—would not engage in this type of word-smithing to

preemptively rationalize more lenient audit risk assessments. Therefore, they would not engage in as much defensive bolstering and motivated reasoning (see Boiney et al. 1997).

Put differently, documentation requirements may prompt auditors with preferences for client-preferred conclusions to “gear up” for an imagined challenge, and defensively word-smith the justifiability of their preferred positions in case their documented judgments should ever face *ex post* scrutiny. As a result, when auditors assess risk in words, their documented judgments may be more lenient than their undocumented judgments, even though documentation opens their lenient judgments up to potential *ex post* scrutiny. Because elastic re-definition only applies to risks assessed in words, I do *not* predict this effect of documentation requirements on audit risks assessed in numbers. This discussion suggests the following three hypotheses in situations where there are pressures to be lenient:

- H1:** Effect of documentation requirements on elastic re-definition of *qualitative* audit risk assessments: *documented* qualitative audit risk assessments will exhibit elastic re-definition (Figure 2) in the event of negative outcomes, while *undocumented* qualitative risk assessments will *not* exhibit elastic re-definition (Figure 2) in the event of negative outcomes.
- H2:** Effect of documentation requirements on *qualitative* audit risk assessment levels: *documented* qualitative assessments will be *more lenient* (i.e., lower) than *undocumented* qualitative assessments, even though documentation potentially exposes lenient judgments to greater *ex post* scrutiny.
- H3:** Effect of documentation requirements on *quantified* audit risk assessment levels: *documented* quantified assessments will *not* be more lenient than *undocumented* quantified assessments.<sup>4</sup>

### Documentation Requirements, Audit Risk Assessment, and Auditor Attributes

H2, in particular, suggests that documentation requirements can create quite unintended consequences when auditors assess audit risks in qualitative terms. To avoid such unintended consequences, standard setters should better understand why auditors persuade themselves to support client-preferred judgments, and how their mandates interact with those factors. Prior research would suggest that the effects predicted in H1–H3 are due to documentation requirements interacting with features of human risk assessment in the judgment setting, rather than undesirable behavior unique to auditors and only auditors. People assess the likelihoods of uncertainties in qualitative terms every day and in many different contexts. In particular, people show a robust fluency in the “language of uncertainty” that they apply across a variety of tasks and contexts (Teigen 1988; Budescu and Wallsten 1987, 1995; Moxey and Sanford 2000). As Piercey (2009) states, the characteristics of qualitative (versus quantified) risk assessment and elastic re-definition

<sup>4</sup> Whether documentation requirements would have no effect on quantified risk assessments or cause them to become more *strict* would depend on whether the documentation leads auditors to act as if they anticipate *a priori* justifying themselves to a more objective, impartial outsider (Kunda 1990), or leads them to vividly anticipate a biased, strict outsider *a priori*. This is simply an empirical question, and so H3 makes no directional prediction for quantified judgments. In contrast, for qualitative judgments, H1 and H2 do make directional predictions. Regardless of whether audit documentation suggests a chance of either impartial or strict *ex post* scrutiny of an auditor’s judgments, either is sufficient to predict auditors assessing risk *in words* to defensively bolster their preferred *lenient* conclusion (via elastic re-definition). As Kunda (1990) notes, the chance of even just an impartial *ex post* review of one’s judgment is enough to prompt defensive bolstering of a preferred conclusion (let alone the chance of an unfair review). The ability to elastically re-define words makes such defensive bolstering much more likely for risks assessed qualitatively.



are “behaviors [that] individuals likely develop as they reach and justify preferred conclusions in a variety of contexts of daily life” (e.g., see footnote 3 of this paper, above). Additionally, motivated reasoning theory would suggest that the effects posited in H1–H3 are very likely to be unconscious effects that people can adapt rather automatically to different circumstances, settings, and contexts (cf. [Kunda 1990](#); [Lerner and Tetlock 1999](#)). Thus, it is likely that the unintended consequences of auditors reaching more lenient judgments are created by the regulatory mandate interacting with the way that humans naturally assess uncertainties in words versus numbers. It is less likely that these effects are due to auditors somehow behaving worse than others would in the same situation (e.g., [Sedor 2002](#), 744–745; [Libby et al. 2002](#); [Peecher and Solomon 2001](#)).

However, there is likely to be individual variance in how a person uses words to assess risks. Some individuals are more persuasive than others ([Friestad and Wright 1999](#)). Some people may tend to be more tactical (i.e., strategic, calculated, planned, deliberate) in their choice of words, while others tend to be more frank (i.e., blunt, direct, outspoken) ([Carter and Russell 2001](#)). [Rich et al. \(1997\)](#) posit that the persuasive attributes of individual auditors influence the characteristics of their auditing judgments. The robust fluency that individuals exhibit in the language of uncertainty across contexts suggests that auditors will tend to apply the approach they take to qualitative risk assessment in other contexts of daily life to their role as an auditor in auditing contexts. As a result, I expect individuals who are more persuasive (less persuasive) and more tactical (more frank) in general contexts to form more (less) vague qualitative auditing judgments.

**H4:** Individuals who are more persuasive (less persuasive) and more tactical (more frank) in general contexts will form qualitative audit risk assessments that are more (less) vague.

If supported, this hypothesis may provide insights to auditors and audit firms as to which auditors are more or less likely to have adequately considered a risk when performing qualitative risk assessment. More vague audit risk assessments create only an illusion of having adequately assessed a risk ([Budescu and Wallsten 1995](#)). Support for this hypothesis would provide audit firms with initial empirical evidence of a potential cost to more persuasive and verbally calculated auditors ([Rich et al. 1997](#)), especially on high-risk audits with client pressures. It also suggests another potential cost to qualitative (as opposed to quantified) audit risk assessments, which auditing standards currently see as an inconsequential choice (e.g., [PCAOB 2007](#); [AICPA 2008](#); [Stone and Dilla 1994](#); [Laswad and Mak 1997](#); [Simon 2002](#)).

## METHOD

### Participants

As discussed above, prior theory suggests that the hypothesized effects would be driven by the way humans naturally assess risks in words versus numbers, rather than behavior unique to auditors (e.g., [Budescu and Wallsten 1987, 1995](#)). Although many behavioral accounting theorists call for the use of student participants in such situations (e.g., [Libby et al. 2002](#), 802–803; [Peecher and Solomon 2001](#), 199–201; [Sedor 2002](#), 744–745), I use both auditors and students to test this assertion. Such a test can inform the PCAOB whether unintended consequences of their standards are due to undesirable behaviors unique to auditors, or to ways in which their standards interact with the task. This can better help standard setters predict when and why their standards may create unintended consequences. This also increases the power of the hypothesis tests while measuring and controlling for any differences attributable to participant type ([Neter et al. 1996](#)).

One hundred thirty-eight auditors from two large public accounting firms and 76 senior accounting students from a highly ranked accounting program participated in the audit risk-assessment task. There was only one significant difference between the auditors and students in any

**TABLE 1**  
**Sample Demographics**

|   | <u>Auditors</u> | <u>Students</u> |
|---|-----------------|-----------------|
| Gender  |                 |                 |
| Females   | 64              | 26              |
| Males   | 73              | 49              |
| Nonrespondents                                      | 1               | 1               |
| Total   | 138             | 76              |
| Experience: Months of Professional Audit Experience |                 |                 |
| Mean  | 42.9            | NA              |
| Standard Deviation                                  | 39.4            | NA              |
| Experience: Rank                                    |                 |                 |
| Staff Auditors                                      | 37              | NA              |
| Seniors and In-Charges                              | 87              | NA              |
| Managers and Senior Managers                        | 10              | NA              |
| Partners  | 4               | NA              |
| Total   | 138             | 76              |

aspect of their assessments. Auditors' risk assessments were, on average, 7.6 percentage points higher than those of accounting students, across all experimental conditions, replicating the general finding that auditors are more conservative than nonauditors (Smith and Kida 1991). However, this overall main effect does not vary significantly by experimental condition (i.e., regardless of whether the assessments were quantified, qualitative, documented, or undocumented) and, therefore, does not affect the tests of hypotheses (Peecher and Solomon 2001; Neter et al. 1996). Table 1 shows the demographics of the sample. The auditors included all ranks, averaged 43 months of experience in public accounting, and ranged up to 23 years of experience. Experience had no significant effects on auditors' responses. Participants were randomly assigned to experimental conditions.

### Overview of the Experiment and Experimental Design

In this experiment, participants read a case about a hypothetical client and a subjective accounting estimate in its financial statements, and assessed the risk of material misstatement in that financial statement account. All participants received a directional preference for client-preferred judgments from their engagement partner, as well as a reminder that the PCAOB was currently inspecting other audits at the firm (but not this audit). The first experimental factor, *documentation requirements*, manipulated between subjects whether their risk assessment would need to be documented in the working papers. The second experimental factor, *response mode*, manipulated how participants would express their risk assessments at three levels: quantified, qualitative, and qualitative-ERD (i.e., qualitative-elastic re-definition).

In the quantified conditions, participants expressed their risk assessments as a 0-to-100 percent probability of material misstatement (see Appendix A). In the qualitative and qualitative-ERD conditions, participants assessed the risk of material misstatement by selecting a phrase (e.g., "somewhat possible") to describe the probability of material misstatement (see Appendix B).

After participants in the qualitative conditions selected a phrase as their audit risk assessment (e.g., "somewhat possible"), the case then asked them to define that previously selected phrase in

terms of its membership function (i.e., Figure 1). In contrast, after participants in the qualitative-ERD conditions selected their phrase, the case first introduced a PCAOB inspector, who had now selected the participants' audit for inspection and had strong preferences against overly lenient auditing judgments, and *then* asked them to define that previously selected phrase in terms of its membership function.

The sole purposes for this PCAOB inspector are (1) to demonstrate elastic re-definition, and (2) to show that it occurs for *documented* qualitative risk assessments, but not for *undocumented* qualitative risk assessments (as H1 predicts in this study).<sup>5</sup> As Figure 3 suggests, the qualitative versus qualitative-ERD comparison allows me to demonstrate empirically how participants elastically re-define their mental representation of a previously specified qualitative risk assessment in case of an adverse outcome (a PCAOB inspector), and whether this depends on documentation requirements (H1). In contrast, the quantified and qualitative conditions allow me to test for the incremental effects of adding documentation requirements within each (H2 and H3). Thus, in H2 and H3, I test the effects of documentation requirements on quantified and qualitative audit risk assessments in general. While the qualitative-ERD condition is used in testing H1, it is irrelevant for testing H2 and H3.

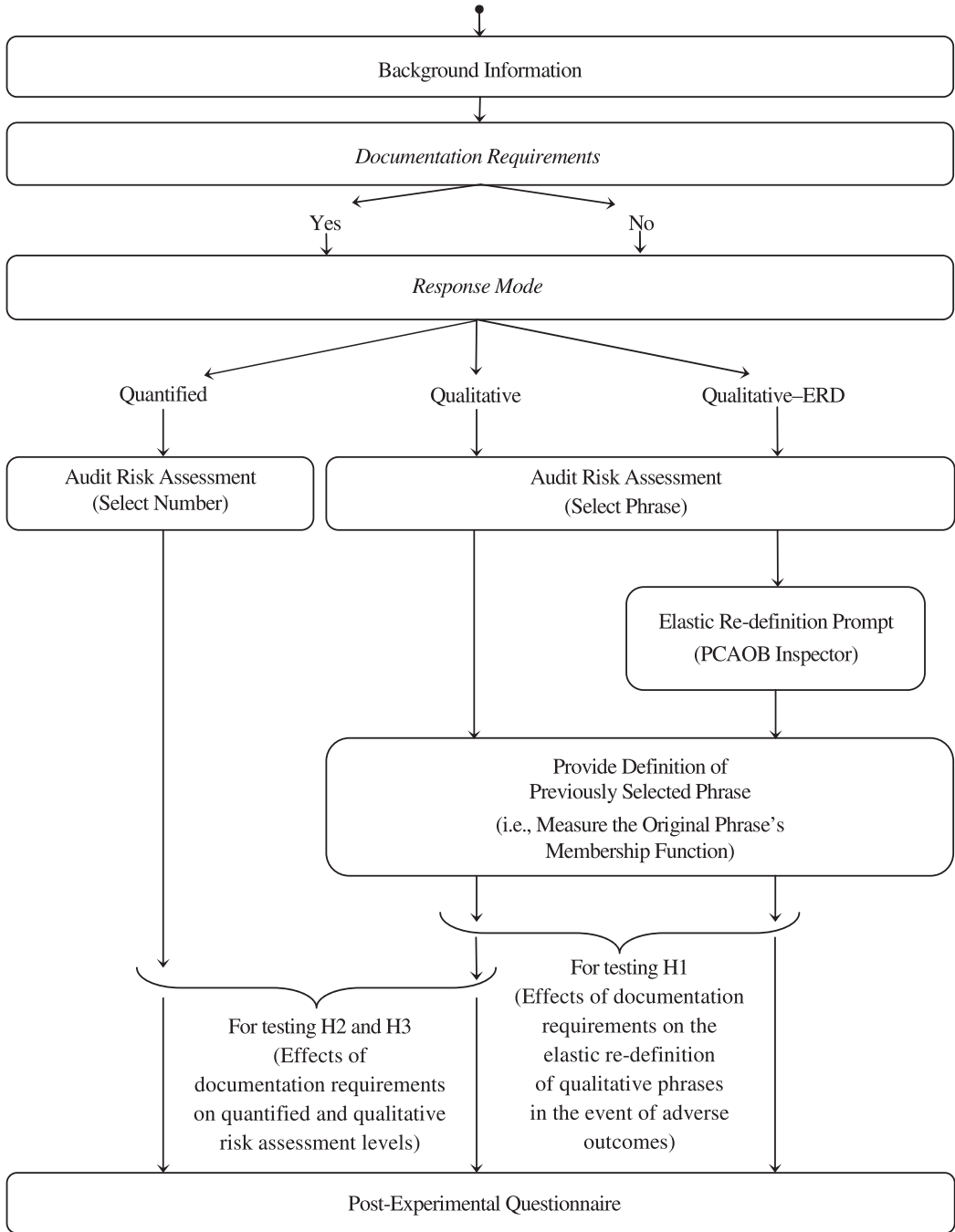
Notice that there are no "quantified-ERD" conditions. Within the quantified conditions, participants have precisely specified their audit risk assessments in numerical terms as they formed their assessment (e.g., "30 percent"). In contrast, qualitative phrases (e.g., "somewhat possible") leave risk unspecified, allowing one the latitude to plausibly re-define in their minds what they originally meant by that previously selected phrase ("somewhat possible"). This simply is not an applicable feature of quantified risk assessments. Operationally, it simply would not make sense to ask participants, after reading about a PCAOB inspector, which 0 percent to 100 percent probability they originally meant by "30 percent" when they selected that number earlier. Thus, *response mode* is a three-level factor (quantified, qualitative, and qualitative-ERD). See, e.g., [Krische \(2005\)](#) for a similar three-level design. Combining that with *documentation requirements* yields a  $2 \times 3$  fully crossed between-subjects factorial design.

### Task Details and Procedures

To begin the task, participants read about the audit of Envista Life, a hypothetical small-cap life insurance company. Because new life insurance customers will benefit a life insurance company for many years, GAAP allows companies to treat costs of obtaining new insurance customers (e.g., sales commissions) as an asset rather than expensing them immediately (Financial Accounting Standards Board [FASB] 1982). However, companies must lower the value of this asset if some new customers are likely to drop their insurance policies early. The context was chosen because, like many auditing judgments, the risk of material misstatement is both important and subjective (cf. [Kennedy et al. 1997](#); [Peecher et al. 2010](#)).

<sup>5</sup> The qualitative-ERD condition is not intended to provide a membership function that a PCAOB inspector would actually be able to observe or use in practice. Instead, this condition exists to reveal the alternative phrase that auditors retain in reserve, if needed, in case they ever would have to justify themselves to a PCAOB inspector or another outside observer. In motivated reasoning theory, people justify a preferred conclusion in their minds *as if* justifying themselves to an outside observer. As the "as if" point implies, the theory does not require that there ever comes a day in which decision makers *actually* explain their judgments to someone in order for this justification process to influence their judgments in the direction of their preferred conclusions. Similarly, elastic re-definition is a justification process in which people contemplate the alternative meanings of their qualitative judgments "in case" they ever have to defend their judgments (e.g., to a PCAOB inspector or some other outsider). Even if this inspection never occurs, elastic re-definition is still of practical import because it leads auditors with client pressures, for now, to justify reaching more lenient audit conclusions.

**FIGURE 3**  
**Experimental Design**



To provide all participants with a source of client pressure, the case described the preferences of the audit engagement partner as preferring generally optimistic (client-preferred) assessments of Envista, held constant across all conditions. As with similar supervisor preferences in prior research (e.g., [Peecher 1996](#); [Brown et al. 1999](#); [Wilks 2002](#)), this language operationalizes the client pressures that auditors face.

The case familiarized participants with the client and with the accounting rules, and gave them reasons why and why not the asset in question may be materially misstated. For the current-year audit, the company's valuation of the asset increased substantially due to (1) a successful marketing campaign that raised substantial new insurance policies, and (2) less impairment of those policies than in prior years. Without the lower rate of impairment, the company would just miss analysts' earnings forecasts. Management justified the lower rate of impairment with several recent internal control enhancements designed to reduce the rate of new customers dropping their policies. They also insisted that their valuation was independent of analysts' forecasts, citing their history of forthrightness with the auditors.

### **PCAOB Inspector and Documentation Requirements**

The case then told participants in all conditions that the PCAOB was conducting work paper reviews of other audits at their firm, and had not yet chosen the Envista audit to examine, but might.<sup>6</sup> Immediately after this, participants then encountered the first of two experimental factors, manipulated between subjects: *documentation requirements* (documented versus not documented). For testing H1–H3, the critical feature of documentation requirements is that they potentially open auditors' assessments up to greater *ex post* scrutiny (see, e.g., H2). The manipulation told participants in the documented (not documented) conditions that the assessment of the probability of material misstatement they would make on the next page would (would not) be documented in the work papers and, thus, subject (not subject) to a PCAOB review, if (even if) one were to occur. Because data collection occurred after the PCAOB's "not documented, not done" standard had already become effective, each condition included an explanation for why their risk assessment would (would not) be documented, so as to help obtain systematic differences between conditions sufficient to observe the hypothesized results ([Kerlinger and Lee 2000](#)). Specifically, the instrument explained to participants in the documented (not documented) conditions that this preliminary judgment would (would not) be documented because of the new "not documented, not done" standards that the PCAOB is enforcing (because, like many professional judgments, it is a preliminary assessment to be discussed between them and the partner).

### **Response Mode**

On the next page, participants in all conditions assessed the probability that management materially overstated the asset and justified their assessments. While participants in the quantified conditions made this assessment on a scale from 0 percent to 100 percent, participants in the qualitative and qualitative-ERD conditions did so by choosing from a list of 18 phrases or by

<sup>6</sup> Both the client pressures from the partner and this description of a PCAOB inspector at the audit firm were held constant before subjects' judgments in all conditions in this experiment. Prior theory suggests that the quality of quantified versus qualitative judgments relative to one another depend on auditors' "net" *ex ante* pressures ([Piercey 2009](#)). The purpose of this study is to examine the incremental effects of documentation requirements on auditors, given client pressures and a PCAOB regulator. As a result, the hypotheses of this study focus on the effects of documentation requirements given one of two response modes common in practice. Note that H2 and H3 do not simply compare qualitative assessments to quantified assessments directly. Rather, H2 predicts the incremental effect of documentation requirements on qualitative assessment levels, and H3 predicts the effect of documentation requirements on quantified assessment levels.

specifying their own.<sup>7</sup> Immediately after making these (quantified or qualitative) audit risk assessments, participants in all conditions read a page asking them not to change any prior answers after this point.

Participants in the qualitative conditions then provided the membership functions for their previously selected phrase on the next page. They did so by indicating the extent to which the phrase describes each of the 11 probabilities—0 percent, 10 percent, . . . , 90 percent, 100 percent, on separate scales, a method of measuring membership functions validated by Budescu et al. (2003). In contrast, participants in the qualitative-ERD conditions first read about a PCAOB inspector with strong preferences against lenient audit judgments and who now has selected their audit for upcoming inspection. Immediately after this, they provided the membership function for their previously selected phrase on the next page. As mentioned previously, the sole reason for the PCAOB inspector is for testing H1. As a result, the PCAOB inspector's preferences against lenient judgments were intentionally strong.<sup>8</sup> The qualitative-ERD condition elicits participants' alternative definition for a phrase in the event of adverse outcomes.

### Dependent Variables and Post-Experimental Questionnaire

The overall dependent variable of interest is the assessed probability of material misstatement, and for participants making that assessment qualitatively, the membership function of that phrase. *Peak* is defined as the probability with the largest membership value (i.e., the probability a phrase describes best) or the single probability value provided by a participant. *Center* is the first moment or "average" probability of the membership function or the single probability provided by a participant. *Vagueness* measures the second moment of the membership function provided by a participant. Finally, *Skewness* measures the negative third moment provided by a participant. Using the *negative* third moment makes it so that positive (negative) values indicate emphasis (de-emphasis) of the higher probabilities implied by a phrase, relative to its lower probabilities. *Vagueness* and *Skewness* are undefined for a single probability value. See Appendix C for formulas.

After completing the experimental case, participants in all conditions then responded to a post-experimental questionnaire. This included demographic data and participants' self-assessments of their own persuasiveness and frankness in general contexts, for use in testing H4. *Persuasiveness* asked participants to assess on an 11-point Likert scale whether, in general, they tend to be more or less persuasive. *Frankness* (adapted from Carter and Russell 2001) asked participants to assess on an 11-point Likert scale whether, in general, they tend to be (1) more frank, straightforward, outspoken, blunt, up-front, forthright, open, direct, and candid, or (2) more strategic, planned, purposeful, tactical, deliberate, careful, calculated, and intentional.

<sup>7</sup> The 18 phrases in the list (Appendix C) are based on a survey that I administered to professional auditors, who indicated which phrases they see or use in practice most frequently. I also constructed the list of phrases in order, from lowest-probability phrases to highest-probability phrases, and alternating between phrases that focus on occurrence versus non-occurrence of events. Juanchich et al. (2010) discuss methodological advantages of this approach.

<sup>8</sup> In this experiment, the partner preferences, the client details, and the qualitative and quantified response scales come from Piercey (2009). However, there was no mention in Piercey (2009) of the PCAOB or its inspector, either before or after participants' judgments; neither was there any manipulation or mention of documentation requirements, all of which are in this study. While I use a PCAOB inspector to test elastic re-definition, Piercey (2009) used a similarly strong, but different, stimulus (his qualitative-ERD participants learned about an undetected material misstatement right before defining what they originally meant by their qualitative risk assessment phrase in terms of its membership function). I expect that an *ex post* regulatory inspection of one's judgments is likely a vivid stimulus for empirically testing and measuring these effects (PCAOB 2008). For example, one participant in a firm training session asked me privately and rather timidly if I was really collecting their judgments for the PCAOB, or even from the PCAOB myself. The important element is that the stimulus used would be seen as an adverse outcome for overly lenient risk assessments.

## RESULTS

### Manipulation Checks

Seven participants did not provide an audit risk assessment and are, therefore, dropped from further analyses. For the response mode manipulation, I controlled directly whether the participants chose a number or a phrase, and whether they provided the membership function for their previously selected phrase before or after learning about the PCAOB inspector. The purpose of the PCAOB inspector was to provide subjects with a significant shift away from their original preferences for lenient judgments *after* they had already made their original (qualitative) audit risk assessments. Participants rated the preferences of their partner at the beginning of the task for client-preferred auditing judgments on an 11-point scale centered at zero, and after making their original risk assessments, subsequently rated the preferences of the PCAOB inspector for client-preferred judgments along the same scale. Participants' ratings of the PCAOB inspector indicated a significant shift away from the partner's preferences for client-preferred assessments ( $-2.7$  versus  $1.6$ ,  $t = -13.42$ ,  $p < 0.01$ ). Finally, participants rated on a five-point scale the extent to which they agreed that this risk assessment would not have been documented in audit work papers. Participants in the not-documented conditions responded with significantly greater agreement than did those in the documented conditions ( $3.1$  versus  $1.8$ ,  $t = 6.85$ ,  $p < 0.01$ ). Thus, the manipulations achieved significant variation between experimental conditions (Kerlinger and Lee 2000; Levine and Parkinson 1994).

A potentially competing explanation for any elastic re-definition observed in the tests of H1 is that participants in the qualitative-ERD conditions do not mentally re-define their selected phrases as in Figure 2, but rather choose different initial phrases than did participants in the qualitative conditions to begin with. There is little reason to expect this, since the experimental tasks for these two conditions are identical up to and through participants' selection of a phrase (see Figure 3). Thus, there is no reason to expect that the two groups would choose systematically different phrases. *Post hoc* tests and sensitivity analyses confirm this ( $p$ 's  $> 0.36$ ; Sidak-corrected  $p$ 's  $> 0.93$ ).

### ANOVA Models and Hypothesis Tests

I analyzed the dependent variables *Peak*, *Center*, *Vagueness*, and *Skewness* using ANOVA models and contrast tests. *Peak* and *Center* are alternative proxies for the probability level of a (qualitative or quantified) probability expression (Budescu and Wallsten 1995). Results for *Peak* and *Center* are statistically similar in every respect; therefore, I just present the results for *Peak* for brevity.

### ANOVA Models

Table 2 shows ANOVAs of the dependent variables *Peak*, *Vagueness*, and *Skewness*. *Persuasiveness*, *Frankness*, and *Participant Type* are measured variables. *Persuasiveness* and *Frankness* were not predicted to influence *Peak* or *Skewness* and do not (either in main effect or in any higher-order interactions). They are, therefore, dropped from those analyses (Neter et al. 1996). As expected, *Persuasiveness* and *Frankness* significantly influence *Vagueness* in two main effects, but not in any higher-order interactions. Since no interactions were predicted or significant, none are included in the model. Although theory would not predict *Participant Type* to interact with response mode or documentation in any of the hypothesized effects (Budescu and Wallsten 1987; Peecher and Solomon 2001; Libby et al. 2002), I measure and control for any differences attributable to *Participant Type*. As I mentioned in the description of the participants (above), auditors' assessments are, on average, 7.6 percentage points higher in *Peak* than those of students,

**TABLE 2**  
**Analyses of Variance**

**Panel A: Peak Probability of Material Misstatement**

| Source                            | Sum of Squares | df  | Mean Square | F    | p-value |
|-----------------------------------|----------------|-----|-------------|------|---------|
| <i>Documentation requirements</i> | 420.4          | 1   | 420.4       | 0.87 | 0.35    |
| <i>Response mode</i>              | 2367.4         | 2   | 1183.7      | 2.46 | 0.09    |
| <i>DR × RM</i>                    | 5622.5         | 2   | 2811.3      | 5.84 | <0.01   |
| <i>Participant Type</i>           | 2774.8         | 1   | 2774.8      | 5.76 | 0.02    |
| Error                             | 96264.0        | 200 | 481.3       |      |         |

**Panel B: Vagueness of Probability of Material Misstatement**

| Source                            | Sum of Squares | df  | Mean Square | F    | p-value           |
|-----------------------------------|----------------|-----|-------------|------|-------------------|
| <i>Documentation requirements</i> | 14.7           | 1   | 14.7        | 0.34 | 0.56              |
| <i>Response mode</i>              | 5.5            | 1   | 5.5         | 0.13 | 0.72              |
| <i>DR × RM</i>                    | 206.8          | 1   | 206.8       | 4.75 | 0.02 <sup>a</sup> |
| <i>Persuasiveness</i>             | 199.9          | 1   | 199.9       | 4.59 | 0.02 <sup>a</sup> |
| <i>Frankness</i>                  | 240.0          | 1   | 240.0       | 5.52 | 0.01 <sup>a</sup> |
| Error                             | 4785.2         | 110 | 43.5        |      |                   |

**Panel C: Skewness of Probability of Material Misstatement**

| Source                            | Sum of Squares | df  | Mean Square | F    | p-value           |
|-----------------------------------|----------------|-----|-------------|------|-------------------|
| <i>Documentation requirements</i> | 10.9           | 1   | 10.9        | 0.09 | 0.77              |
| <i>Response mode</i>              | 73.9           | 1   | 73.9        | 0.58 | 0.45              |
| <i>DR × RM</i>                    | 391.6          | 1   | 391.6       | 3.10 | 0.04 <sup>a</sup> |
| Error                             | 14409.6        | 114 | 126.4       |      |                   |

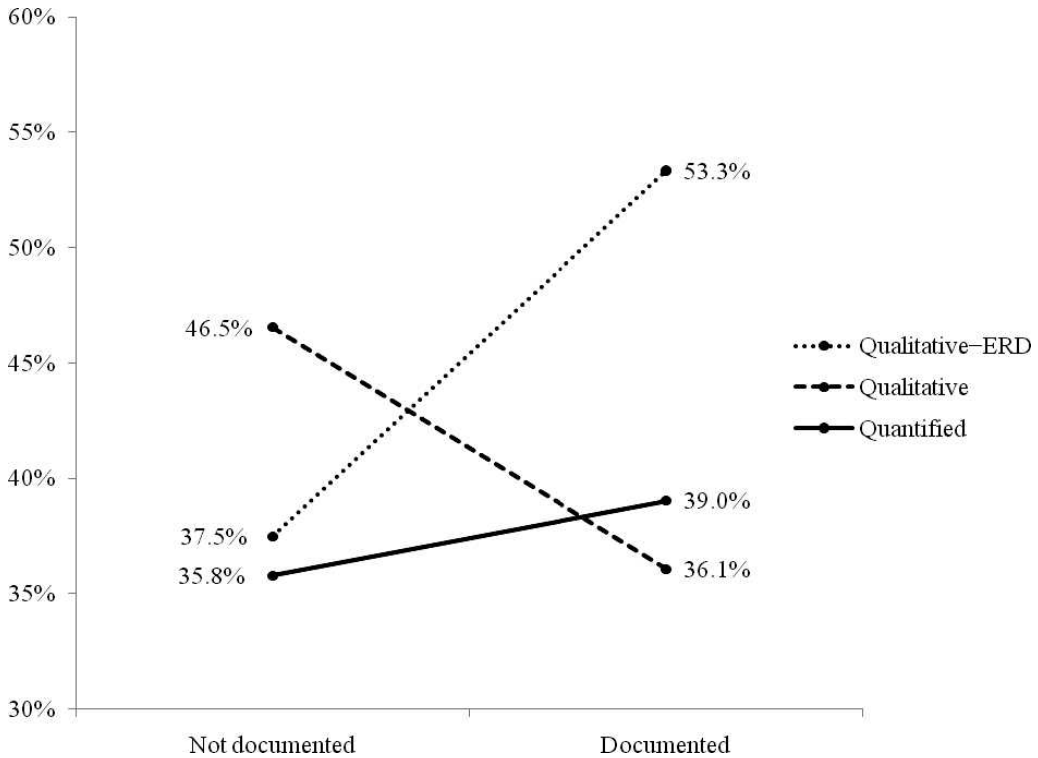
<sup>a</sup> The p-values for these terms test directional predictions and are the one-tailed test of the signed t-statistic associated with this F-test (as in, e.g., Kachelmeier and Williamson 2010). McNeil et al. (1996) discuss one-tailed tests of interactions with directional expectations. Variable definitions are in Appendix C.

across experimental conditions ( $F = 5.76$ ,  $p = 0.02$ ). However, besides this main effect, there are no significant interactions involving *Participant Type* in the analyses of *Peak* (or *Center*), as expected, and, therefore, none are included in the model. *Participant Type* is insignificant in either main effect or in any higher-order interactions in the analyses of *Vagueness* and *Skewness*, and is, therefore, dropped from those models. I do not detect any other significant measured covariates (e.g., gender, time).<sup>9</sup>

<sup>9</sup> Tests of normality and variance constancy for all dependent variables are well below levels that challenge the assumptions of ANOVA or elevate Type I error rates (Neter et al. 1996; Tabachnick and Fidell 2001). One participant provided a nonsensical membership function that was flagged by Cook's distance as statistically outlying for *Vagueness* and was, therefore, dropped from those analyses (Hardin and Hilbe 2007; Neter et al. 1996). The observation was not flagged as statistically outlying in *Peak* or *Skewness*. Results are statistically unchanged when it is dropped from those analyses, as well.



**FIGURE 4**  
**Peak Probability of Material Misstatement**



Variable definitions are in Appendix C.

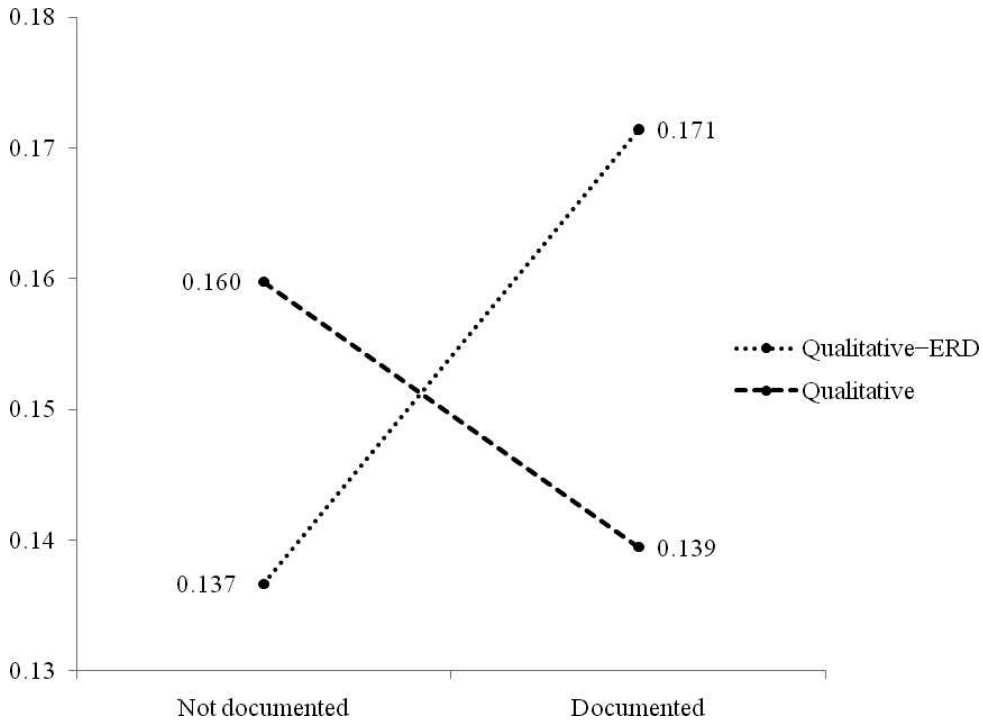
### Tests of H1

Figures 4, 5, and 6 show the means for *Peak*, *Vagueness*, and *Skewness* (respectively) from the ANCOVA models in Table 2.<sup>10</sup> H1 predicts that elastic re-definition will occur in case of a PCAOB inspection for documented—but not undocumented—qualitative risk assessments. That is, the simultaneous increases in *Peak*, *Vagueness*, and *Skewness* depicted in Figure 2 will occur within the documented conditions, but they will not occur within the undocumented conditions.

Participants' documented qualitative risk assessments conveyed a 36.1 percent *Peak* probability of material misstatement within the qualitative conditions. However, in the qualitative-ERD

<sup>10</sup> These least-squares means in Figures 4, 5, and 6 are the best linear unbiased estimates (BLUE) of population means, and control for the effects of significant covariates in the ANCOVA models (Neter et al. 1996; Searle et al. 1980). Because *Participant Type* does not interact significantly with *documentation requirements* or *response mode*, results for *Peak* by *Participant Type* are statistically similar to those collapsed across all participants in Figure 4 in almost every respect. The only difference is that auditors' *Peaks* are higher than nonauditors across each of the experimental conditions pictured in Figure 4, by the above-mentioned main effect. Otherwise, this difference does not vary by condition and has no effect on the interaction pictured in Figure 4 or the hypothesis tests (cf. Peecher and Solomon 2001, 201). Analogously, results for *Vagueness* and *Skewness* by *Participant Type* are statistically similar to those in Figures 4 and 5 in every respect.

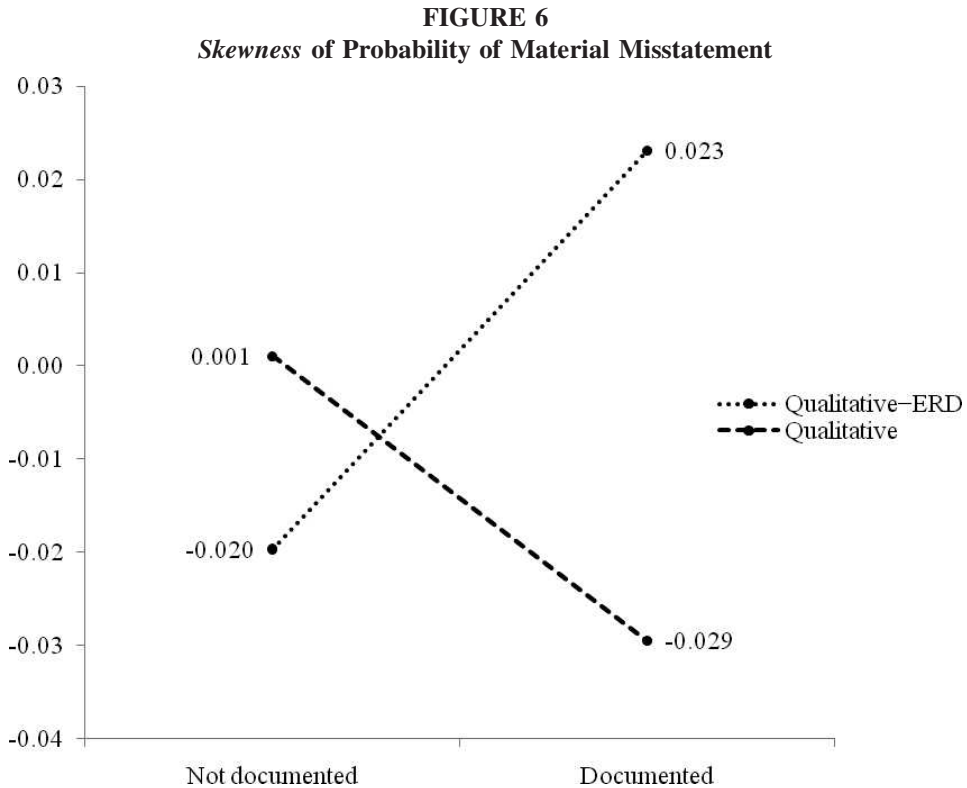
**FIGURE 5**  
**Vagueness of Probability of Material Misstatement**



Variable definitions are in Appendix C.

conditions, participants acted as if that same previous, documented qualitative risk assessment phrase really implied a 53.3 percent *Peak* probability of misstatement, a significant increase ( $t = 3.14$ ,  $p < 0.01$ ) (Figure 4). Similarly, participants re-defined their documented phrase as more vague (0.139 versus 0.171,  $t = 1.78$ ,  $p = 0.04$ ) (Figure 5) and more skewed to emphasize its higher probabilities more ( $-0.029$  versus  $0.023$ ,  $t = 1.78$ ,  $p = 0.04$ ) (Figure 6) in the qualitative-ERD conditions than in the qualitative conditions. Thus, participants' documented qualitative risk assessments were fairly lenient, while holding an alternative, more skeptical, and vague re-definition in reserve in case of PCAOB inspection. However, these same increases did not occur in the not-documented conditions. *Vagueness* and *Skewness* did not change significantly (0.160 versus 0.137,  $t = -1.33$ ,  $p = 0.19$ ) (Figure 5); (0.001 versus  $-0.020$ ,  $t = -0.71$ ,  $p = 0.48$ ) (Figure 6). *Peak* also did not increase and, if anything, decreased marginally from 46.5 percent to 37.5 percent ( $t = -1.68$ ,  $p = 0.10$ ) (Figure 4). This suggests that, unlike their documented counterpart, *undocumented* qualitative risk assessments were not relatively lenient, and did not hold an alternative, more skeptical, and vague re-definition in reserve in case of PCAOB inspection.<sup>11</sup> These results support H1.

<sup>11</sup> Furthermore, the increases in the *Peak*, *Vagueness*, and *Skewness* of qualitative risk assessments with addition of the PCAOB inspector within the documented conditions are significantly larger than are the lack of such increases within the not-documented conditions ( $t = 3.42$ ,  $p = 0.01$ ;  $t = 2.18$ ,  $p = 0.02$ ;  $t = 1.76$ ,  $p = 0.04$ , respectively).



Variable definitions are in Appendix C.

The marginally significant *decrease* in *Peak* (from 46.5 percent to 37.5 percent,  $p = 0.10$ ) suggests a potentially interesting *post hoc* finding, which I interpret as follows. H1 predicted no increases (like the ones in Figure 2) within the not-documented conditions. As expected, the *undocumented* qualitative assessments (which do not retain an alternative re-definition, as in Figure 2, to help rationalize lower risk assessments) are relatively high (46.5 percent). Within these not-documented conditions, arrival of the PCAOB inspector likely may have simply served to remind these participants that, according to the case, this judgment is undocumented and, therefore, safe from this PCAOB inspector's scrutiny. If so, such a reminder of this safe harbor would tend to make them marginally more willing to characterize the risk leniently according to their partner's preferences, compared to other not-documented participants who did not have this incremental reminder.<sup>12</sup>

<sup>12</sup> However, this marginally significant decrease ( $p = 0.10$ ) is not significant in any sense after modest adjustment for the *post hoc* nature of the finding (Sidak-corrected  $p = 0.86$ ). Whether this lack of an increase in the undocumented condition is also a marginal decrease is a somewhat interesting *post hoc* question, but relatively unimportant for my purposes. The critical part of H1 is that elastic re-definition (i.e., the three increases in Figure 2) occurred in the documented conditions, but did not occur in the undocumented conditions. That leads to the prediction in H2, and its important practical implications. The incremental effects of a PCAOB inspection on *undocumented* risk assessments may be an avenue that future researchers wish to explore further.

### Tests of H2 and H3

Building off of H1, H2 predicts that adding documentation requirements will (ironically) cause qualitative audit risk assessments to become more lenient, while H3 predicts that documentation requirements will not have this effect on quantified assessments. The rationale for these predictions is as follows: Documentation requirements increase a risk of *ex post* scrutiny that is important, but fairly small and deferred to the future. In contrast, client pressures present a more immediate need to reach a preferred audit conclusion. As a result, documentation requirements may, ironically, prompt more defensive bolstering (via elastic re-definition) of lenient, client-preferred audit conclusions in case of PCAOB inspection (H1). Thus, documentation requirements prompt a mental rationalization of how defensible a lenient audit risk assessment could be, ironically leading to more lenient qualitative risk assessments (relative to undocumented qualitative assessments, H2). In contrast, documentation requirements would not have this effect on quantified assessments (H3), because quantified assessments do not allow for word-smithing behavior.

To test H2 and H3, I examine the effects of documentation on *Peak* probability assessments within the quantified and qualitative conditions. As shown in Figure 4, participants' *Peak* qualitative assessments of the probability of material misstatement decreased significantly, from 46.5 percent to 36.1 percent (i.e., became more lenient), with the addition of documentation requirements ( $t = -1.83$ ,  $p = 0.04$ ). In contrast, participants' quantified assessments did not decrease with documentation requirements, instead increasing from 35.8 percent to 39.0 percent, although not significantly ( $t = 0.64$ ,  $p = 0.52$ ). This effect of documentation requirements on qualitative assessments is significantly larger than its (lack of an) effect on quantified assessments ( $t = -1.80$ ,  $p = 0.04$ ). These results support H2 and H3.

### Tests of H4

H4 predicts that participants who identify themselves as more persuasive (less persuasive) and more tactical (more frank) in general contexts will form more vague (less vague) risk assessments. As covariates, *Persuasiveness* and *Frankness* appear homogeneously distributed across experimental conditions and participant type (in means and in variances,  $p$ 's  $> 0.49$ ) (Neter et al. 1996). The signed t-tests for the regression coefficients of *Persuasiveness* and *Frankness*, in the general linear model from the ANCOVA on *Vagueness* in Table 2, are significantly positive and negative, respectively ( $t = 2.14$ ,  $p = 0.02$ ;  $t = -2.35$ ,  $p = 0.01$ ). These results support H4. This finding provides preliminary evidence for audit firms of a potential cost to more persuasive and verbally calculated auditors (Rich et al. 1997), since more vague risk assessments give the illusion of having adequately assessed a risk, without actually having considered it in a meaningful way (cf. Budescu and Wallsten 1995).

## CONCLUSION

The PCAOB (2004) increased auditors' documentation requirements, facilitating regulatory scrutiny of auditor's judgments *ex post*. This study examines the effect of such documentation requirement on auditors who already face conflicting pressures to reach client-preferred audit judgments. Because documentation requirements increase auditors' potential accountability to regulatory parties, one would ordinarily not expect that adding them would make auditors with client pressures become *more* lenient. However, findings from this study suggest that when auditors assess risk in non-quantified terms, adding documentation requirements prompts a specific word-smithing strategy that Piercey (2009) shows can help rationalize reaching more lenient audit judgments. Because documentation requirements prompt this behavior, they can, ironically, cause more defensive bolstering of lenient qualitative risk assessments, even though they simultaneously

open auditors' judgments up to potential *ex post* regulatory scrutiny. Such defensive bolstering can potentially help an auditor meet the more immediate need to address client pressures, while coping with the deferred chance of regulatory inspection. As a result, documentation requirements can, ironically, lead to more lenient audit risk assessments when auditors assess risk in qualitative terms. On the other hand, documentation requirements do not have this effect on quantified audit risk assessments, which are not subject to this word-smithing strategy in case of regulatory inspection. More lenient risk assessments lead auditors to perform fewer tests, rely more on internal controls, and collect less substantive evidence in support of the audit opinion (AICPA 2008; IFAC 2009). This potentially has adverse implications for audit effectiveness, contrary to the stated intentions of the regulatory requirement (PCAOB 2004). This is especially important given the general preference for qualitative risk assessment in practice (e.g., Simon 2002).

Kachelmeier and King (2002) argue that researchers should investigate potentially unintended consequences of accounting standards and mandates, and that they should use the advantages of experiments in their ability to create concurrent conditions with and without a standard in effect to evaluate the effects of the standard, while holding other factors constant. As they note, such an approach can also help regulators better understand why their mandates have the effects that they do, and potentially adjust their standard-setting processes. The findings of this study provide evidence about the potential costs and benefits of the PCAOB's "not documented, not done" standard. The unintended consequences shown here should be of interest to auditors, regulators, and others interested in potentially unintended consequences of documentation requirements. I also provide evidence that these effects are due to documentation requirements interacting with the way people naturally assess risk in an audit setting, rather than behaviors that auditors and only auditors exhibit. Regulators (as well as researchers) should understand the basic psychology of how people naturally assess risk in order to better predict when and why their regulations may have unintended consequences on auditor effectiveness (e.g., Kadous et al. 2003). This paper also extends Piercey (2009), who did not consider the effects of documentation requirements, how they prompt elastic re-definition, or how they concurrently affect the leniency of auditing judgments. Additionally, that study only considered auditors with one prevailing incentive (i.e., skepticism, objectivity, or leniency), whereas this study considers auditors with conflicting *ex ante* incentives (i.e., client pressures to be lenient, and the presence of a PCAOB inspector examining other audits at the firm). No prior psychology study (or other study) of which I am aware investigates the effects of documenting versus not documenting verbal and numerical probability assessments (including the potential for an *ex post* inspection of one's judgments that documentation entails). This study can inform regulators, auditors, and academics with a better understanding of how and why documentation requirements can affect auditor judgment in unanticipated and unintended ways. It can also provide empirical evidence to auditors, regulators, and academics of a potential cost of qualitative (versus quantified) audit risk assessments. Currently, auditing standards view the choice of qualitative versus quantified risk assessment as inconsequential (e.g., AICPA 2008; Stone and Dilla 1994).

Like all empirical studies, this study has limitations. First, I do not consider all factors of the audit setting that may moderate my results. For example, many auditors commonly assess risk on scales of high, medium, or low, whereas I allowed them to choose from a breadth of 18 phrases or by providing their own. However, my approach is also representative of audit practice. For example, Simon (2002, 608–609) and Amer et al. (1994, 129) identify (and have auditors interpret) 47 distinct risk assessment phrases that are common to auditing and accounting, professional standards, and audit firm manuals, similar to mine. In selecting the phrases used for this study, I developed a list of 69 risk assessment phrases directly from professional standards and other accounting literature before surveying professional auditors (see footnote 7) about which they use most frequently in practice, and the auditors indicated frequent use of such

phrases. Conversations with partners also suggest that the phrases were realistic, and that common phrases such as these would tend to appear in auditors' statements as they recursively revise their beliefs about the risk of misstatement throughout an audit (cf. Bell et al. 2005). Nevertheless, since high-medium-low scales are also common in auditing, future research could examine the use of such scales. Second, my instrument measured auditors' perceptions of the risk of misstatement, but not planned audit work. In general, lower risk assessments imply that auditors would perform fewer tests, rely more on internal controls, and collect less evidence to form an unqualified opinion than higher risk assessments do. Nevertheless, future research could confirm whether this is true in this setting. Third, I predict and find evidence that the judgmental effects in this study are not confined to only auditors, using accounting students as nonauditors. However, accounting students likely bear more in common with auditors than nonaccounting nonauditors. Future research could determine whether differences emerge by using individuals who have no accounting knowledge.

This study provides evidence of a potentially unintended consequence of documentation requirements with important practical implications. Given the widespread application of documentation requirements for virtually all auditing judgments, as required by the "not documented, not done" standard, future research should investigate how documentation influences auditors' judgments in other ways that may have beneficial effects under some circumstances, as well as unintended consequences under others.

## REFERENCES

- Amer, T., K. Hackenbrack, and M. Nelson. 1994. Between-auditor differences in the interpretation of probability phrases. *Auditing: A Journal of Practice & Theory* 13 (1): 126–136.
- American Institute of Certified Public Accountants (AICPA). 1988. *Rule 102: Integrity and Objectivity*. AICPA Rules of Conduct and Interpretations. New York, NY: AICPA.
- American Institute of Certified Public Accountants (AICPA). 2008. *Audit Risk and Materiality in Conducting an Audit*. Statements on Auditing Standards, AU Section 312. New York, NY: AICPA.
- Bazerman, M. H., D. A. Moore, P. E. Tetlock, and L. Tanlu. 2006. Reports of solving the conflicts of interest in auditing are highly exaggerated. *Academy of Management Review* 31 (1): 43–49.
- Bell, T. B., M. E. Peecher, and I. Solomon. 2005. *The 21st Century Public Company Audit: Conceptual Elements of KPMG's Global Audit Methodology*. KPMG International. Available at: <http://www.business.illinois.edu/kpmg-uiuccases/monograph2.pdf>
- Boiney, L. G., J. Kennedy, and P. Nye. 1997. Instrumental bias in motivated reasoning: More when more is needed. *Organizational Behavior and Human Decision Processes* 72 (1): 1–24.
- Brown, C. E., M. E. Peecher, and I. Solomon. 1999. Auditors' hypothesis testing in diagnostic inference tasks. *Journal of Accounting Research* 37 (1): 1–26.
- Budescu, D. V., T. M. Karelitz, and T. S. Wallsten. 2003. Predicting the directionality of probability words from their membership functions. *Journal of Behavioral Decision Making* 16: 159–180.
- Budescu, D. V., and T. S. Wallsten. 1987. Subjective estimation of precise and vague uncertainties. In *Judgmental Forecasting*, edited by Wright, G., and P. Ayton, 63–82. Hoboken, NJ: John Wiley & Sons Ltd.
- Budescu, D. V., and T. S. Wallsten. 1995. Processing linguistic probabilities: General principles and empirical evidence. *Psychology of Learning and Motivation* 32: 275–318.
- Carter, P., and K. Russell. 2001. *Psychometric Testing*. West Sussex, U.K.: John Wiley & Sons.
- Financial Accounting Standards Board (FASB). 1982. *Accounting and Reporting by Insurance Enterprises*. Statement of Financial Accounting Standards No. 60. Norwalk, CT: FASB.
- Friestad, M., and P. Wright. 1999. Everyday persuasion knowledge. *Psychology & Marketing* 16 (2): 185–194.

- Hardin, J. W., and J. M. Hilbe. 2007. *Generalized Linear Models and Extensions, 2nd Edition*. College Station, TX: StataCorp LP.
- Hsee, C. K. 1996. Elastic justification: How unjustifiable factors influence judgments. *Organizational Behavior and Human Decision Processes* 66 (1): 122–129.
- International Federation of Accountants (IFAC). 2009. *Overall Objectives of the Independent Auditor and the Conduct of an Audit in Accordance with International Standards on Auditing*. International Standard on Auditing 200. New York, NY: IFAC.
- Juanchich, M., K. H. Teigen, and G. Villejoubert. 2010. Is guilt “likely” or “not certain”? Contrast with previous probabilities determines choice of verbal terms. *Acta Psychologica* 135 (3): 267–277.
- Kachelmeier, S. J., and R. R. King. 2002. Using laboratory experiments to evaluate accounting policy issues. *Accounting Horizons* 16: 219–232.
- Kachelmeier, S. J., and M. G. Williamson. 2010. Attracting creativity: The initial and aggregate effects of contract selection on creativity-weighted productivity. *The Accounting Review* 85 (5): 1669–1692.
- Kadous, K., J. Kennedy, and M. E. Peecher. 2003. The effect of quality assessment and directional goal commitment on auditors’ assessment of client-preferred accounting methods. *The Accounting Review* 78 (3): 759–778.
- Kennedy, J., D. N. Kleinmuntz, and M. E. Peecher. 1997. Determinants of the justifiability of performance in ill-structured audit tasks. *Journal of Accounting Research* 35 (Supplement): 105–123.
- Kerlinger, F. N., and H. B. Lee. 2000. *Foundations of Behavioral Research, 4th Edition*. New York, NY: Thomson Learning.
- Krische, S. 2005. Investors’ evaluations of strategic prior-period benchmark disclosures in earnings announcements. *The Accounting Review* 80: 243–268.
- Kunda, Z. 1990. The case for motivated reasoning. *Psychological Bulletin* 108 (3): 480–498.
- Laswad, F., and Y. T. Mak. 1997. Interpretations of probability expressions by New Zealand standard setters. *Accounting Horizons* 11: 16–23.
- Lerner, J. S., and P. E. Tetlock. 1994. Accountability and social cognition. *Encyclopedia of Human Behavior* 1: 1–10.
- Lerner, J. S., and P. E. Tetlock. 1999. Accounting for the effects of accountability. *Psychological Bulletin* 125 (2): 255–275.
- Levine, G., and S. Parkin. 1994. *Experimental Methods in Psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Libby R., R. Bloomfield, and M. W. Nelson. 2002. Experimental research in financial accounting. *Accounting, Organizations and Society* 27 (8): 755–811.
- Martinov, N., and P. Roebuck. 1998. The assessment and integration of materiality and inherent risk: An analysis of major firms’ audit practices. *International Journal of Auditing* 2: 103–126.
- McFadgen, D. N. 1994. Discussion of: When numbers are better than words: The joint effects of response representation and experience on inherent risk judgments. *Auditing: A Journal of Practice & Theory* 13 (Supplement): 20–23.
- McNeil, K., I. Newman, and F. J. Kelley. 1996. *Testing Research Hypotheses with the General Linear Model*. Carbondale, IL: Southern Illinois University Press.
- Moore, D. A., P. E. Tetlock, L. Tanlu, and M. H. Bazerman. 2006. Conflicts of interest and the case of auditor independence: Moral seduction and strategic issue cycling. *Academy of Management Review* 31 (1): 10–29.
- Moxey, L. M., and A. J. Sanford. 2000. Communicating quantities: A review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology* 14: 237–255.
- Nelson, M. 2006. Ameliorating conflicts of interest in auditing: Effects of recent reforms on auditors and their clients. *Academy of Management Review* 31 (1): 30–42.
- Neter, J., M. H. Kutner, C. J. Nachtsheim, and W. Wasserman. 1996. *Applied Statistical Linear Models*. Boston, MA: MCB McGraw-Hill.
- Olson, M. J., and D. V. Budescu. 1997. Patterns of preference for numerical and verbal probabilities. *Journal of Experimental Decision Making* 10: 117–131.

- Peecher, M. E. 1996. The influence of auditors' justification processes on their decisions: A cognitive model and experimental evidence. *Journal of Accounting Research* 34 (1): 125–140.
- Peecher, M. E., M. D. Piercey, J. S. Rich, and R. M. Tubbs. 2010. The effects of a supervisor's active intervention in subordinates' judgments, directional goals, and perceived technical knowledge advantage on audit team judgments. *The Accounting Review* 85 (5): 1763–1786.
- Peecher, M. E., and I. Solomon. 2001. Theory and experimentation in studies of audit judgment and decisions: Avoiding common research traps. *International Journal of Auditing* 5: 193–203.
- Piercey, M. D. 2009. Motivated reasoning and verbal vs. numerical probability assessment: Evidence from an accounting context. *Organizational Behavior and Human Decision Processes* 108 (2): 330–341.
- Public Company Accounting Oversight Board (PCAOB). 2004. *Audit Documentation*. PCAOB Auditing Standard No. 3. Washington, D.C.: PCAOB.
- Public Company Accounting Oversight Board (PCAOB). 2007. *An Audit of Internal Control over Financial Reporting that is Integrated with an Audit of Financial Statements*. PCAOB Auditing Standard No. 5. Washington, D.C.: PCAOB.
- Public Company Accounting Oversight Board (PCAOB). 2008. Report on the PCAOB's 2004, 2005, 2006, and 2007 Inspections of Domestic Annually Inspected Firms. PCAOB Release 2008-008. Washington, D.C.: PCAOB.
- Pyszczynski, T., and J. Greenberg. 1987. Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In *Advances in Experimental Social Psychology, Volume 20*, edited by Berkowitz, L. New York, NY: Academic Press.
- Rich, J. S., I. Solomon, and K. T. Trotman. 1997. The audit review process: A characterization from the persuasion perspective. *Accounting, Organizations and Society* 22 (5): 481–505.
- Searle, S. R., F. M. Speed, and G. A. Milliken. 1980. Population marginal means in the linear model: An alternative to least squares means. *The American Statistician* 34 (4): 216–221.
- Sedor, L. M. 2002. An explanation for unintentional optimism in analysts' earnings forecasts. *The Accounting Review* 77 (4): 731–754.
- Shaw, D. 2006. Dilemma of interest. *American Journalism Review* 28 (1): 56–61.
- Simon, J. 2002. Interpretation of probability expressions by financial directors and auditors of U.K. companies. *The European Accounting Review* 11 (3): 601–629.
- Smith, J. F., and T. Kida. 1991. Heuristics and biases: Expertise and task realism in auditing. *Psychology Bulletin* 109 (3): 472–489.
- Stone, D. N., and W. N. Dilla. 1994. When numbers are better than words: The joint effects of response representation and experience on inherent risk judgments. *Auditing: A Journal of Practice & Theory* 13 (Supplement): 1–19.
- Tabachnick, B. G., and L. S. Fidell. 2001. *Using Multivariate Statistics, 4th Edition*. Needham Heights, MA: Allyn and Bacon.
- Teigen, K. H. 1988. The language of uncertainty. *Acta Psychologica* 68: 27–38.
- Teigen, K. H., and W. Brun. 1995. Yes, but it is uncertain: Direction and communicative intention of verbal probabilistic terms. *Acta Psychologica* 88: 233–258.
- Teigen, K. H., and W. Brun. 1999. The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes* 80 (2): 155–190.
- Teigen, K. H., and W. Brun. 2000. Ambiguous probabilities: When does  $p = 0.3$  reflect a possibility, and when does it express a doubt? *Journal of Behavioral Decision-Making* 13: 345–362.
- Tetlock, P. E. 1999. Accountability theory: Mixing properties of human agents with properties of social systems. In *Shared Cognition in Organizations: The Management of Knowledge*, edited by Thompson, L. L., J. M. Levine, and D. M. Messick, 117–138. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Wallsten, T. S., D. V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth. 1986. Measuring the meanings of probability phrases. *Journal of Experimental Psychology: General* 115 (4): 348–356.
- Wilks, J. 2002. Predecisional distortion of evidence as a consequence of real-time audit review. *The Accounting Review* 77 (1): 51–72.



**APPENDIX A**  
**Quantified Audit Risk Assessments**

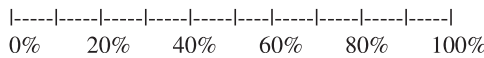
**Your Assessment, continued**

Instructions: Please respond to the following questions from the *partner* in charge of the audit:

3. “Considering all of the risks and audit evidence thus far, what is the probability that Envista management materially overstated its policy acquisition assets in the financial statements?”

Please answer by marking an X on the scale below the probability that *BEST* describes your assessment:

- A: “There is a \_\_\_% probability that Envista management materially overstated its policy acquisition assets.”



**APPENDIX B**  
**Qualitative and Qualitative—ERD Audit Risk Assessments**

**Your Assessment, continued**

Instructions: Please respond to the following questions from the *partner* in charge of the audit:

3. “Considering all of the risks and audit evidence thus far, what is the probability that Envista management materially overstated its policy acquisition assets in the financial statements?”

Please answer by placing a check mark next to the phrase that *BEST* describes your assessment (select only *one*):

- A: “. . . that Envista management materially overstated its policy acquisition assets.”

- |     |                  |                      |
|-----|------------------|----------------------|
| ___ | It is            | almost impossible    |
| ___ | It is            | not quite impossible |
| ___ | There is         | only a chance        |
| ___ | There is         | a chance             |
| ___ | It is            | very unlikely        |
| ___ | It is            | slightly likely      |
| ___ | It is            | quite uncertain      |
| ___ | There is         | some possibility     |
| ___ | There is         | substantial doubt    |
| ___ | It is            | somewhat likely      |
| ___ | It is            | somewhat uncertain   |
| ___ | It is            | reasonably possible  |
| ___ | It is            | somewhat unlikely    |
| ___ | It is            | very likely          |
| ___ | It is            | slightly improbable  |
| ___ | It is            | probable             |
| ___ | It is            | not quite certain    |
| ___ | It is            | almost certain       |
| ___ | Other (specify): | _____                |

## APPENDIX C

*Peak, Center, Vagueness, and Skewness*

For a membership function that assigns membership values ( $\mu(p)$ ) to each of the probabilities  $p \in [0, 0.1, \dots, 0.9, 1]$ , *Peak*, *Center*, *Vagueness*, and *Skewness* are defined as follows:

$$\text{Peak} = \mathbf{arg\,max}_p \mu(p), \quad (1)$$

$$\text{Center} = \bar{p} = \frac{\sum_{p=0}^1 \mu(p) \cdot p \cdot \Delta p}{\sum_{p=0}^1 \mu(p) \cdot \Delta p}, \quad (2)$$

$$\text{Vagueness}^2 = \frac{\sum_{p=0}^1 \mu(p)(p - \bar{p})^2 \cdot \Delta p}{\sum_{p=0}^1 \mu(p) \cdot \Delta p}, \quad (3)$$

$$\text{Skewness}^3 = -1 \times \frac{\sum_{p=0}^1 \mu(p)(p - \bar{p})^3 \cdot \Delta p}{\sum_{p=0}^1 \mu(p) \cdot \Delta p}. \quad (4)$$

For single probability values, *Peak* and *Center* equal the numerical probability itself, and *Vagueness* and *Skewness* are undefined (Budescu and Wallsten 1995). Note that *Peak*, *Center*, *Vagueness*, and *Skewness* are analogous to the mode, first, second, and negative third moments of a probability density, respectively, except that, for probability densities, the denominator in each expression always equals 1. The limit of these expressions as  $\Delta p \rightarrow 0$  is given in Piercey (2009).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.