

Documentation Requirements and Audit Risk Assessment

M. DAVID PIERCEY

University of Massachusetts at Amherst

Assistant Professor

Isenberg School of Management

Accounting & Information Systems Department

121 Presidents Drive

Amherst, MA 01003

piercey@som.umass.edu

September 24, 2009

This study 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 have also benefited from the comments of Efrim Boritz, Bill Dilla, Anne Farrell, Jim Gaa, Mike Gibbins, Gary Hecht, Josh Herbold, Frank Hodge, Karim Jamal, Jane Kennedy, Ron King, Natalia Kotchetova, Susan Krische, Joan Luft, Bill Messier, Kim Moreno, Mark Nelson, Derek Oler, Andrea Piercey, Jamie Pratt, Jim Smith, Chris Wolfe, workshop participants at the University of Alberta, Georgia State University, the University of Illinois, Indiana University, the University of Massachusetts, Michigan State University, the University of Washington, Washington University, and the University of Waterloo, and conference participants at the American Accounting Association Annual Meeting. I thank two public accounting firms for providing access to participants. I gratefully acknowledge the financial assistance of the Deloitte & Touche Doctoral Fellowship and the Marvin and Nancy Dennis Doctoral Fellowship. I also thank the American Accounting Association, who awarded an earlier version of this manuscript the Accounting, Behavior, & Organizations Outstanding Dissertation Award.

Documentation Requirements and Audit Risk Assessment

Abstract

Standard setters have increased auditors' documentation requirements to facilitate regulatory scrutiny of auditors' judgments ex post. This study examines the effect of such documentation requirements 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 cause auditors assessments of the client to become more *lenient*. However, this study predicts and finds the following. When auditors assess risks in non-quantified terms, adding documentation requirements prompts a specific word-smithing strategy that allows auditors to rationalize more lenient auditing judgments, while coping with the increased risk of regulatory inspection. As a result, documentation requirements can ironically lead to more lenient audit risk assessments when auditors assess risk in qualitative — but not in quantified — terms. Thus, depending on how auditors assess risk, recent documentation requirements may have unintended effects with adverse implications for audit effectiveness, contrary to their original regulatory objectives.

Keywords: Documentation requirements, audit workpapers, audit risk assessment, quantification, auditor judgment and decision making

JEL Descriptors: C91, M40, M41, M42, M49

1. 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 (Jamal 2008; Salterio 2008). 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 itself states that a primary intention is to improve audit effectiveness and efficiency by enhancing the collection of evidence in support of the auditor’s judgments (PCAOB 2004, 321). The standard requires that auditors now document all of their judgments in such detail that an experienced auditor reviewing the audit could reconstruct every auditing judgment (PCAOB 2004). Since this standard was released alongside new regulatory inspections of public auditing firms (PCAOB 2008), it appears that an *at least secondary* intention of the auditing standard was to provide regulatory inspectors — not just experienced reviewing auditors — with much more material for scrutinizing auditors’ judgments *ex post* (Ricchiute 2006, 222). As Salterio (2008, 115) puts it:

“However, PCAOB AS No. 3 on ‘Documentation’, later adopted by both the CICA’s Auditing and Assurance Standards Board (AASB) and the International Auditing and Assurance Standards Board (IAASB), was not written from the viewpoint of better evidence documentation for users, but with the view of making it easier for the PCAOB/CPAB inspector to review the files (clothed in the rhetoric of better evidence documentation).”

This study examines the effects of adding such documentation requirements to the judgments of auditors who concurrently face pressures to reach client-preferred audit conclusions. While many of the new auditing standards (like AS No. 3) are intended to improve

audit quality, 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. As more years drift from initial implementation of new regulatory mandates, auditors may become more likely to find ways of coping with the new mandates while finding ways to see subjective accounting estimates more from their clients' point of view (e.g., Jamal 2008; Kadous, Kennedy, and Peecher 2003; Salterio 2008; Sunder 2005).

Thus, auditors continue to face conflicting incentives to evaluate their clients leniently, objectively, and skeptically. Because documentation requirements open auditors' judgments up to regulatory inspection, one would ordinarily expect that adding them to auditors' client pressures would cause their audit risk assessments to become more skeptical, if anything. However, this study finds evidence that the opposite can occur, depending on how auditors assess their clients' audit risks. Professional standards give auditors the option of quantifying audit risks (i.e., putting them into numbers; Kadous, Koonce, and Towry 2005) or expressing them in qualitative terms (IFAC 2009; PCAOB 2007). I find that documentation requirements prompt use of a specific word-smithing motivated reasoning strategy documented in Piercey (2009). This word-smithing strategy can help auditors rationalize more lenient, client-preferred qualitative audit risk assessments (Piercey 2009), while simultaneously coping with the increased risk of PCAOB or CPAB inspection. As a result, adding documentation requirements can ironically cause auditors' qualitative — but not their quantified — audit risk assessments to become more lenient. 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).

An implication of my findings is that documentation requirements may not reduce the tendency for auditors to evaluate their clients leniently, and may even increase it under common conditions. In fact, auditors are more likely in practice to assess risk in qualitative than in quantified terms (Piercey 2009; PCAOB 2007, 419; Simon 2002; Olson and Budescu 1997; Martinov and Roebuck 1998). Therefore, documentation requirements can have unintended consequences contrary to both the implicit and explicit original objectives of regulators (PCAOB 2004). This new insight extends the accounting literature that examines other aspects of audit workpaper documentation (e.g., Agoglia, Hatfield, and Brazel 2009; Brazel, Agoglia, and Hatfield 2004; Ricchiute 1999; Tan and Tan 2008; Tan and Trotman 2003).

This study also provides new evidence of *how* auditors facing conflicting incentives can mentally justify client pressures to be lenient while simultaneously coping with increased accountability to regulators. Supplemental findings in this study suggest that these effects are due to institutional features of the judgment setting interacting with the way that humans naturally assess risk, rather than undesirable behaviors that auditors and only auditors exhibit. Regulators should understand how their regulations can interact with how people assess risk within an audit judgment setting, in unintended ways that achieve neither more objective nor more skeptical auditing judgments (e.g., Kadous et al. 2003).

In the remainder of this paper, section 2 develops theory and hypotheses, sections 3 and 4 describe the experimental method and results, and section 5 discusses conclusions.

2. Theory and Hypotheses

Audit risk assessment and motivated reasoning

During an audit, auditors continuously form and revise their beliefs about the likelihood of material misstatements in a client's financial statement accounts and of material weaknesses

in its internal controls (Ashton and Ashton 1988; Bell, Peecher, and Solomon 2005). Professional standards refer to these likelihoods as audit risks (AICPA 2008; IFAC 2009). Auditors assess such risks both formally (e.g., during initial planning of the audit) and informally throughout the audit (e.g., assessments and re-assessments of the risk of material misstatement in the conclusions of individual workpapers; Bell et al. 2005).

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; IFAC 2009). When client pressures motivate auditors to reach client-preferred audit conclusions, auditors can use biased information search, interpretation of audit evidence, ambiguous decision rules, and word-smithing to justify reaching more lenient auditing judgments (e.g., Brown, Peecher, and Solomon 1999; Kadous et al. 2003; Peecher 1996; Piercey 2009; Salterio and Koonce 1997; 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 (Kunda 1990, 483; Kadous et al. 2003).

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 encourage using qualitative assessments (e.g., PCAOB 2007, 419). 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; Piercey 2009; Simon 2002).

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.

[Insert Figure 1 here]

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 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 (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), accounting research has only recently begun to use membership functions to better understand qualitative risk assessments in accounting and auditing. 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 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 can help maintain a mental illusion of justifiability.

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), such as a PCAOB or CPAB regulatory inspector, if ever necessary. One can easily re-think 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 mean lots of things. This re-thought 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 (Piercey 2009; cf. Kunda 1990).

Figure 2 illustrates this specific type of motivated reasoning. The membership function on the left represents 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 re-cast 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 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 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.¹

¹ Elastic re-definition is a specific type of “word-smithing” that can assist in motivated reasoning. Following its non-scientific meaning in common usage (Webster’s 2005), Piercey (2009) defines word-smithing as contemplating multiple semantic attributes of a phrase (e.g., its peak, vagueness, and perhaps some other linguistic characteristic) when using it to express one’s beliefs or judgments to oneself or others.

Piercey (2009) used negative outcome information (learning later that there had, in fact, been a material misstatement, cf. Peecher and Piercey 2008) rather than a PCAOB or CPAB inspector, to illustrate this phenomenon, hypothesize it, and test it empirically in an experiment. However, I expect that a regulatory inspection (as opposed to a negative outcome of the risk assessment itself) will produce this same effect.²

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 (Piercey 2009). In contrast, quantified risk assessments (e.g., “30%”) are precisely specified and more difficult for an auditor to mentally re-define as meaning something else while maintaining an illusion of objectivity (e.g., it is harder to justify to oneself how 30% could really be taken to mean 60%, if necessary).

[Insert Figure 2 here]

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. In this study, I consider auditors who simultaneously face conflicting pressures to be lenient on the client on one hand, and the threat of PCAOB/CPAB inspection on the other. This characterization of auditors is especially interesting in light of the regulatory environment of recent years, because the inherent pressures on auditors to support lenient evaluations of their

² 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 to imply suspicion.

clients have not gone away just because regulatory mandates have increased. Rather than these two incentives merely canceling out, auditors must continue to deal with client pressures while simultaneously coping with the chance of regulatory inspection.

In this study, I posit that such documentation requirements *prompt* elastic re-definition as a justification strategy. Specifically, requiring auditors to document every single material audit judgment in such detail as to facilitate potential inspection 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., as depicted in Figure 2. On the other hand, when auditors do not have such documentation requirements, there is no 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.

Thus, I hypothesize that auditors with client pressures to be lenient will, in case of regulatory inspection, act as though their original *documented* qualitative risk assessments really means something more skeptical (1st moment), more vague (2nd moment), and with more emphasis on the new meaning (3rd moment). On the other hand, these three increases will not occur for undocumented qualitative risk assessments. Put differently, elastic re-definition (Figure 2) will occur for documented but not undocumented qualitative risk assessments:

HYPOTHESIS 1: Auditors with client pressures will, in the event of a regulatory inspection, re-define a previous *documented* qualitative audit risk assessment as though it (1) implies higher probabilities of material misstatement, (2) is more vague, and (3) is skewed to emphasize its higher-probability meaning more (and de-emphasize its lower-probability meaning). However, these three increases (Figure 2) will not occur for *undocumented* qualitative risk assessments in case of regulatory inspection.

Documentation requirements open auditors' judgments up to ex post scrutiny by regulatory authorities. Ordinarily, one would expect that adding such documentation requirements to auditors' client pressures would cause their judgments to become more skeptical, if anything, but not more lenient.

However, the more that individuals engaged in motivated reasoning 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 motivated reasoning (Kunda 1990). For example, Kadous et al. (2003) demonstrate that asking auditors to provide evaluations of their clients' financial reporting quality actually amplifies the effects of their goals to reach client-preferred judgments. Peecher, Piercey, Rich and Tubbs (2009) demonstrate that when supervising auditors provide outwardly objective guidance to their subordinates, the supervisors form auditing judgments more influenced by 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. Elastic re-definition is a specific word-smithing strategy that can enable auditors to accommodate the more immediate and certain client pressures to be lenient while maintaining an illusion of justifiability in case of a potential future PCAOB inspection. Hypothesis 1 predicts that documentation requirements *prompt* this elastic re-definition behavior. Thus, documentation requirements may ultimately prompt auditors' more vigilant rationalization of their lenient audit risk assessments as justifiable and defensible against the chance of a future regulatory inspection. 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 pre-emptively rationalize more lenient audit risk assessments. Therefore, they would not engage in as much defensive bolstering and motivated reasoning (see Boiney, Kennedy, and Nye 1997). Thus, adding documentation requirements to existing client pressures may ironically cause *qualitative* audit risk assessments to become more *lenient*, because of the word-smithing behavior it prompts.

However, documentation requirements would not prompt elastic re-definition of quantified assessments (since precisely specified numbers are not subject to word-smithing). As a result, adding documentation requirements would not tend to cause quantified audit risk assessments to become more lenient. This suggests the following hypothesis:

HYPOTHESIS 2: Adding documentation requirements to client pressures to be lenient will cause *qualitative* audit risk assessments to become more lenient, but not *quantified* audit risk assessments.

Individual auditor attributes

Finally, Hypothesis 2 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. Psychology research would suggest that the effect predicted in Hypotheses 1 and 2 are due to documentation requirements interacting with features of human risk assessment in the judgment setting rather than undesirable behavior unique to 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) notes, the characteristics of qualitative (vs. quantified) risk assessment and elastic re-definition 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 2 of this study). Additionally, motivated reasoning theory would suggest that the effects posited in Hypotheses 1 and 2 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 (implied by Hypothesis 2) of auditors reaching more lenient judgments are created by the regulatory mandate interacting with the way that humans naturally assess uncertainties in words vs. 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; Peecher and Solomon 2001).

However, there is likely individual variance in how a single 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; cf. Carter and Russell 2001). Rich, Solomon, and Trotman (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 auditors who are more persuasive (less persuasive) and more tactical (more frank) in general contexts to form more (less) vague qualitative auditing judgments. In general, increasingly vague qualitative risk

assessments would tend to give auditors a false sense that they have mentally assessed a specific risk, but without actually forming a useful mental characterization of the risk that can be used in any precise way. Budescu and Wallsten (1995) refer to this false sense of risk assessment that vague assessments can create as the “illusion of communication.” That is, the more vague one’s mental depiction of a risk, the less one has actually assessed the risk. This warrants predicting the following effect:

HYPOTHESIS 3: Individuals who identify themselves as more persuasive (less persuasive) and more tactical (more frank) in general contexts will, in an auditing context, form more vague (more precise) qualitative assessments of their clients’ audit risks.

If supported, this hypothesis may provide insights to auditors and audit firms of which auditors are more or less likely to have adequately considered a risk when performing qualitative risk assessment. It would also extend our understanding of how auditors strategically form audit risk assessments, and link behavior in general contexts to behavior in auditing contexts.

3. Method

Participants

Recall my prediction that the hypothesized effects would be driven by documentation requirements interacting with the way humans naturally assess risks in words vs. numbers, rather than behavior unique to auditors (section 2). Although behavioral accounting theorists call for the use of student subjects in such situations (e.g., Peecher and Solomon 2001, 199-201; Libby, Bloomfield, and Nelson 2002, 802-803; Sedor 2002, 744-745), I use both auditors and students to test this assertion. This also increases the power of the hypothesis tests while measuring and controlling for any significant differences attributable to subject type (Neter et al. 1996). One hundred thirty-eight auditors from two large public accounting firms in the U. S. and seventy-six

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 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 non-auditors (Smith and Kida 1991). However, this main effect does not vary significantly by experimental condition (i.e., regardless of whether the assessments were quantitative, qualitative, documented, or undocumented), and therefore does not affect the tests of hypotheses (Peecher and Solomon 2001; Neter et al. 1996). 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. Subjects were randomly assigned to experimental conditions.

Experimental task

Participants read a case 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 (FASB 1982). However, companies must impair 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 (Kennedy et al. 1997).

To provide all participants with a source of client pressure, the case described the preferences of audit engagement partner as preferring generally optimistic (client-preferred) assessments of Envista, held constant across all conditions. I used the language from the optimistic experimental conditions of Piercey (2009), adapted very slightly to match the current

study. 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. These case details also came from Piercey (2009).

Documentation requirements. The case then told participants in all conditions that the PCAOB was conducting workpaper reviews of other audits at their firm, and that the inspector had not yet chosen the Envista audit to examine, but might (not in Piercey 2009).³ At this point, participants then encountered the first of two experimental factors, manipulated between subjects: *documentation requirements* (documented vs. not documented). For testing Hypotheses 1 and 2, the critical feature of documentation requirements is that they open auditors' assessments up to potential regulatory scrutiny. The case told participants in the documented (not-documented) conditions that the preliminary assessment of the probability of material misstatement they would make on the next page would (would not) be documented in the

³ Because there was no mention of a PCAOB inspector in Piercey (2009), subjects in this study had two stimuli providing conflicting ex ante prompts (i.e., client pressures to be lenient and a PCAOB inspector examining other audits at the firm), whereas subjects in Piercey (2009) had only one prevailing ex ante stimulus (i.e., the client pressures).

workpapers, 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; Levine and Parkinson 1994). Specifically, the instrument explained to participants in the documented (not-documented) conditions that their judgments 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. At this point, participants encountered the second of two factors, manipulated between subjects at three levels: *response mode* (quantified, qualitative, or qualitative-PCAOB). Participants in the quantified conditions assessed the probability of material misstatement on a scale from 0% to 100%. Participants in the qualitative conditions made their probability assessments by choosing from a list of 18 phrases or by specifying their own.⁴ These participants then provided the membership functions for their selected phrase on the next page. They did so by indicating the extent to which the phrase describes each of the 11 probabilities 0%, 10%, ..., 90%, 100% on separate scales, a method of measuring membership functions validated by Budescu et al. (2003). Comparing the

⁴ The 18 phrases are the same used in Piercey (2009). Their external validity is validated by both prior research (see Simon 2002 for a review) and a survey I administered to professional auditors who indicated the phrases they would use most frequently in professional practice. The list began with lower-probability phrases (e.g., almost impossible), then intermediate probability phrases (e.g., somewhat likely) and ended with higher-probability phrases (e.g., almost certain), reflecting methodological considerations noted by Piercey (2009; e.g., Hamm 1991; Teigen and Brun 1999).

quantified and qualitative conditions (and how this depends on documentation) allows me to test Hypothesis 2.

The qualitative–PCAOB conditions were the same as the qualitative conditions, except in one respect. *After* making their audit risk assessment but immediately *before* defining its membership function, participants in the qualitative–PCAOB conditions learned that the PCAOB inspector had selected their audit for inspection. Thus, comparing the qualitative and qualitative–PCAOB conditions allows me to observe how an audit risk assessment changes definition in the case of a PCAOB inspection. This re-definition (and how it depends on documentation requirements) allows me to test Hypothesis 1.⁵

At the end of the instrument, participants in all conditions responded to a post-experimental questionnaire. This included demographic data and subjects’ self-assessments of their own persuasiveness and frankness in general contexts, for use in testing Hypothesis 3. *Persuasiveness* is an 11-point Likert-scale variable that asked participants to self-assess whether, in general, they tend to be more or less persuasive. *Frankness* is an 11-point Likert-scale variable that asked participants to self-assess whether, in general, they are best described as frank (straightforward, outspoken, blunt, up front, forthright, open, direct, candid) or strategic (planned, purposeful, tactical, deliberate, careful, calculated, intentional), adapted from Carter and Russell (2001).

Dependent variables

The overall dependent variable of interest is the assessed probability of material misstatement, and, for participants making that assessment qualitatively, the membership

⁵ Piercey (2009) used a different stimulus to test for elastic re-definition. However, I expect that a 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 timorously if I was really collecting their judgments for the PCAOB, or even from the PCAOB myself.

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 subject. *Center* is the first moment or “average” probability of the membership function or the single probability provided by a subject. *Vagueness* measures the second moment of the membership function provided by a subject. Finally, *skewness* measures the negative third moment provided by a subject. 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 (following Piercey 2009). *Vagueness* and *skewness* are undefined for a single probability value. See Appendix A for formulas.

4. Results

Manipulation checks

For the response mode manipulation, the instrument provided the scale (quantitative or qualitative) that participants could use to express their audit risk assessments. Seven subjects did not provide an audit risk assessment and are therefore dropped from further analyses. For the qualitative vs. qualitative–PCAOB comparison, the purpose of the PCAOB inspector was to provide subjects in the latter condition with a significant shift in accountability to show how the definition of a previous qualitative risk assessment will change in case of a PCAOB inspection. Subjects in all conditions rated the preferences of their partner for optimistic vs. pessimistic evaluations of the client on an 11-point scale centered at zero, and subjects in the qualitative–PCAOB also rated the preferences of the PCAOB inspector along the same scale. Subjects’ ratings of the PCAOB inspector indicated a significant shift away from the partner’s preferences for client-preferred assessments (-2.8 vs. 1.6 , $t = -13.4$, $p < 0.01$). Finally, subjects rated on a five-point scale the extent to which they agreed that this risk assessment would not

have been documented in audit workpapers. Subjects in the not-documented conditions responded with mixed but significantly greater agreement than did those in the documented conditions (3.1 vs. 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).

Analysis of variance models and hypothesis tests

I now turn to the tests of dependent variables *peak*, *center*, *vagueness*, and *skewness*. *Peak* and *center* constitute alternative proxies for the probability level of a (qualitative or quantified) probability expression (Piercey 2009). Results for *peak* and *center* are statistically similar in every respect; therefore, I just present the results for *peak* for brevity.

Table 1 shows ANOVAs of the dependent variables *peak*, *vagueness*, and *skewness*. *Persuasiveness*, *frankness*, and *subject 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 subject type to interact with response mode or documentation in any of the hypothesized effects (Peecher and Solomon 2001), I measure and control for any differences attributable to subject type. As mentioned in section 3, auditors' assessments are, on average, 7.6 percentage points higher in *peak* than those of students, across experimental conditions ($t = 5.76$, $p = 0.02$). However, besides this main effect, there are no significant interactions involving *subject type* in the analyses of *peak* (or *center*), as expected, and therefore none are included in the model. *Subject 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 co-variates (e.g., gender, time).⁶

[Insert Table 1 here]

Tests of Hypothesis 1. Figures 3, 4, and 5 show the means for *peak*, *vagueness*, and *skewness* (respectively) from the ANCOVA models in Table 1.⁷ Hypothesis 1 predicts that elastic re-definition will occur in case of a PCAOB inspection for documented — but not undocumented — qualitative risk assessments. That is, moving from the qualitative to the qualitative-PCAOB conditions, the simultaneous increases in *peak*, *vagueness*, and *skewness* depicted in Figure 2 will occur in the documented conditions, but they will not occur in the undocumented conditions. Within the documented conditions, subjects' qualitative risk assessments conveyed a 36.1% *peak* probability of material misstatement. However, in the event of a PCAOB inspection, subjects acted as if that same previous, documented qualitative risk assessment phrase really implied at 53.3% *peak* probability of misstatement, a significant increase ($t = 3.14$, $p_{\text{one-tailed}} < 0.01$; Figure 3). Similarly, subjects re-defined their documented qualitative risk assessment phrase as more vague (0.139 vs. 0.171, $t = 1.78$, $p = 0.04$; Figure 4) and more skewed to emphasize its higher probabilities more (-0.029 vs. 0.023 , $t = 1.78$, $p = 0.04$; Figure 5) in the event of a PCAOB inspection. These results show that documented

⁶ Tests of normality and variance constancy for all dependent variables are well within levels that challenge the assumptions of ANOVA or elevate Type I error rates (Neter et al. 1996; Tabachnick and Fidell 2001). One subject 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.

⁷ The least-squares means in Figures 3, 4, and 5 are the best linear unbiased estimates (BLUE) of population means and control for the effects of measured covariates (Neter et al. 1996; Piercey 2009; Searle, Speed, and Milliken 1980). Because *subject type* does not interact significantly with *documentation requirements* or *response mode*, results for *peak* by *subject type* are statistically similar to those collapsed across all subjects in Figure 3 in almost every respect. The only difference is that auditors' *peaks* are higher than non-auditors across each of the experimental conditions pictured in Figure 3, by the above-mentioned main effect. Otherwise, this difference does not vary by condition and has no effect on the interaction pictured in Figure 3 or the hypothesis tests (cf. Peecher and Solomon 2001, 201). Analogously, results for *vagueness* and *skewness* by *subject type* are statistically similar to those in Figures 4 and 5 in every respect.

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 vs. 0.137, $t = -1.33$, $p_{\text{two-tailed}} = 0.19$, Figure 4; 0.001 vs. -0.020 , $t = -0.71$, $p_{\text{two-tailed}} = 0.48$, Figure 5). *Peak* also did not increase, and, if anything, *decreased* marginally from 46.5% to 37.5% ($t = -1.68$, $p_{\text{two-tailed}} = 0.10$; Figure 3). This suggests that, unlike their documented counterpart, *undocumented* qualitative risk assessments were not relatively lenient, and did not hold and an alternative, more skeptical and vague re-definition in reserve in case of PCAOB inspection. These results support Hypothesis 1.⁸

[Insert Figures 3, 4, and 5 here]

Hypothesis 1 predicted no increases within the not-documented conditions. The marginally significant *decrease* in *peak* (from 46.5% to 37.5%, $p = 0.10$) suggests a potentially interesting post hoc finding, which I interpret as follows. 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%). Within these not-documented conditions, arrival of the PCAOB inspector likely would have simply served to remind these subjects that, according to the case, this judgment is undocumented and therefore safe from this PCAOB inspector's scrutiny. Such a reminder would tend to make them marginally more willing to characterize the risk leniently according to their partner's preferences, compared to other not-documented subjects who did not have this incremental reminder.⁹

⁸ 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). This also supports Hypothesis 1.

⁹ However, this marginally significant decrease 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. It suffices

Tests of Hypothesis 2. Hypothesis 2 predicts that qualitative — but not quantified — audit risk assessments will become more lenient with documentation requirements. The basis for this prediction is that documentation requirements prompt defensive bolstering (via elastic re-definition) of the leniency of a qualitative assessment, whereas undocumented assessments do not experience this defensive bolstering. Thus, by prompting a mental rationalization of how defensible a lenient audit risk assessment could be, documentation requirements may ironically lead to more lenient qualitative risk assessments (relative to undocumented qualitative assessments). In contrast, documentation requirements would not have this effect on quantified assessments, because they do not prompt such word-smithing of quantified assessments.

As shown in Figure 3, subjects' *peak* qualitative assessments of the probability of material misstatement decreased significantly from 46.5% to 36.1% with documentation requirements ($t = -1.83$, $p = 0.04$). In contrast, subjects' quantified assessments did not decrease with documentation requirements, increasing from 35.8% to 39.0%, 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 Hypothesis 2.

Tests of Hypothesis 3. Hypothesis 3 predicts that individuals who identify themselves as more persuasive (less persuasive) and more tactical (more frank) in general contexts will form evaluations of their clients that are more vague (less vague). As co-variates, *persuasiveness* and *frankness* appear homogenously distributed across experimental conditions and subject type (in both mean and variance, 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

for my theory and hypotheses that elastic re-definition (i.e., the three increases in Figure 2) occurs for documented but not for undocumented qualitative risk assessments. The incremental effects of a PCAOB inspection on undocumented risk assessments may be an avenue that future researchers wish to explore further.

vagueness in Table 1 are significantly positive and negative, respectively ($t = 2.14$, $p = 0.02$; $t = -2.35$, $p = 0.01$). These results support Hypothesis 3.

5. Conclusions

This study examines the effects of documentation requirements that explicitly open up auditors' judgments to ex post regulatory scrutiny on the judgments of auditors with conflicting pressures to reach client-preferred conclusions. I predict and find that such documentation requirements prompt a word-smithing strategy (documented in Piercey 2009) that auditors can use to reach client-preferred judgments. Specifically, this word-smithing strategy allows auditors to rationalize how, in case of regulatory inspection, a lenient qualitative audit risk assessment could be taken to mean something more skeptical, and mean more things, with some de-emphasized connection to the original, lenient meaning of the phrase. This perceived justifiability of the qualitative audit risk in case of regulatory inspection allows auditors, for now, to assess the risk more leniently. 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. 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.

The perceived justifiability of this word-smithing strategy is likely very illusory. In motivated reasoning theory, individuals are more likely to let a preference (e.g., for client-preferred audit conclusion) influence their judgment only to the extent that they can maintain an *illusion* of justifiability for doing so (Kunda 1990). For example, if a regulatory inspection were to occur, the leniency of the original risk assessment may manifest itself in other areas, such as relatively low levels of tests of controls and substantive evidence collected as a consequence of the auditor's lenient mindset. Thus, the elastic re-definition of qualitative audit risk assessments may only give auditors a false sense of justifiability (Piercey 2009). Nevertheless, this illusion of justifiability would tend to give auditors the perception of more comfort to meet the immediate needs of satisfying the client for now, while mentally coping with the deferred need to defend this judgment in case of a potential PCAOB inspection. As Kunda (1990, 483) puts it,

“[T]he objectivity of this justification construction process is illusory because people do not realize that the process is biased by their goals, that they are accessing only a subset of their relevant knowledge, that they would probably access different beliefs and rules in the presence of different directional goals, and that they might even be capable of justifying different conclusions on different occasions.”

Thus, because documentation requirements prompt this elastic re-definition behavior, they may prompt what is ultimately only a false sense of defensibility, while simultaneously prompting more lenient qualitative assessments of the client. Such false justifiability for leniency would tend to create adverse effects for audit quality across audits.

This unintended consequence of documentation requirements extends the accounting literature that examines other aspects of audit workpaper documentation and its effects on auditor judgment (e.g., Agoglia et al. 2004; Ricchiute 1999; Tan and Trotman 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, whereas Piercey (2009) only considered auditors with one prevailing incentive (i.e., skepticism, objectivity, or leniency), this study considers how 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) cope with these pressures. 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. I also present evidence linking the persuasive behavior of individuals in general contexts to the attributes of their qualitative risk assessments of an audit client within an auditing context. 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.

Like all empirical studies, this study has limitations. I do not consider all factors of the audit setting that may moderate my results. For example, subjects in this task evaluated an asset valuation necessary for the client to meet earnings targets. Future research may investigate other factors relevant to the task and context (e.g., group decision making). Given the widespread application of documentation requirements of virtually all auditing judgments, as required by recent “not documented, not done” auditing standards, 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.

Appendix A: *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:

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

$$Vagueness^2 = \frac{\sum_{p=0}^1 \mu(p)(p - \bar{p})^2 \cdot \Delta p}{\sum_{p=0}^1 \mu(p) \cdot \Delta p} \quad (3), \quad Skewness^3 = -1 \times \frac{\sum_{p=0}^1 \mu(p)(p - \bar{p})^3 \cdot \Delta p}{\sum_{p=0}^1 \mu(p) \cdot \Delta p} \quad (4).$$

For single probability values, *peak* and *center* equal the numerical probability itself, and *vagueness* and *skewness* are undefined (Piercey 2009; 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).

References

- AICPA. 2008. AU Section 312: Audit risk and materiality in conducting an audit. *Statements on Auditing Standards*, New York, NY: American Institute of Certified Public Accountants.
- AICPA. 1988. Rule 102: Integrity and objectivity. *AICPA Rules of Conduct and Interpretations*, New York, NY: American Institute of Certified Public Accountants.
- Agoglia, C. P., R. C. Hatfield, and J. F. Brazel. 2009. The effects of audit review format on review team judgments. *Auditing: A Journal of Practice & Theory* 28 (1): 95-111.
- Anderson, U., K. Kadous, and L. Koonce. 2004. The role of incentives to manage earnings and quantification in auditors' evaluations of management-provided information. *Auditing: A Journal of Practice & Theory* 23 (1): 11-28.
- Ashton, A. H. and R. H. Ashton. 1988. Sequential belief revision in auditing. *The Accounting Review* 63 (4): 623-641.
- 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.
- 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.
- Bowlin, K. O., J. Hales, and S. J. Kachelmeier. 2009. Experimental evidence on how prior experience as an auditor influences managers' strategic reporting decisions. *Review of Accounting Studies* 14: 63-87.
- Brazel, J. F., Agoglia, C. R., and R. C. Hatfield. 2004. Electronic versus face-to-face review: The effects of alternative forms of review on auditors' performance. *The Accounting Review* 79 (4): 949-966.
- Brown, C. E., M. E. Peecher, and I. Solomon. 1999. Auditor's 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. *Judgmental Forecasting*, ed. G. Wright and P. Ayton, John Wiley & Sons Ltd, 63-82.
- Budescu, D. V. and T. S. Wallsten. 1995. Processing linguistic probabilities: General principles and empirical evidence. *Decision Making from a Cognitive Perspective, Psychology of Learning and Motivation: Advances in Research and Theory*, ed. J. Busemeyer, R. Hastie, and D. L. Medin, 32: 275-318.

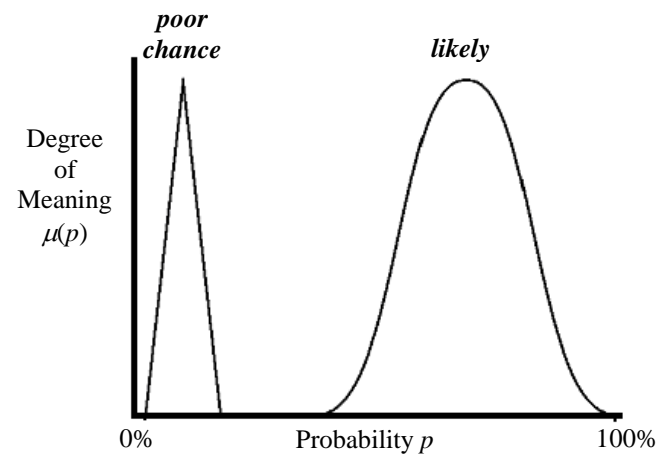
- Carter, P. and K. Russell. 2001. *Psychometric Testing*, West Sussex, England: John Wiley & Sons.
- FASB. 1982. Financial Accounting Standards Board Statement of Financial Accounting Standards No. 60: *Accounting and Reporting by Insurance Enterprises*, Norwalk, CT: Financial Accounting Standards Board.
- Friestad, M. and P. Wright. 1999. Everyday persuasion knowledge. *Psychology & Marketing* 16(2): 185-194.
- Hamm, R. M. 1991. Selection of verbal probabilities: A solution for some problems of verbal probability expression. *Organizational Behavior and Human Decision Processes* 48: 193-223.
- Hardin, J. W. and J. M. Hilbe. 2007. *Generalized Linear Models and Extensions*, 2nd ed., 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.
- IFAC. 2009. ISA 200: Overall objectives of the independent auditor and the conduct of an audit in accordance with international standards on auditing. *International Standards on Auditing*, New York, NY: International Federation of Accountants.
- Jamal, K. 2008. Mandatory audit of financial reporting: A failed strategy for dealing with fraud. *Accounting Perspectives* 7 (2): 97-110.
- 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.
- Kadous, K., L. Koonce, and K. L. Towry. 2005. Quantification and persuasion in managerial judgement. *Contemporary Accounting Research* 22 (3): 643-86.
- 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 ed. New York, NY: Thomson Learning.
- Kunda, Z. 1990. The case for motivated reasoning, *Psychological Bulletin* 108 (3): 480-498.
- 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.

- 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.
- McNeil, K., I. Newman, and F. J. Kelley. 1996. *Testing Research Hypotheses with the General Linear Model*. Carbondale, IL: Southern Illinois University Press.
- 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.
- 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 auditor's justification processes on their decisions: A cognitive model and experimental evidence. *Journal of Accounting Research* 34 (1): 125-140.
- Peecher, M. E. and M. D. Piercey. 2008. Judging audit quality in light of adverse outcomes: Evidence of outcome bias and reverse outcome bias. *Contemporary Accounting Research* 25 (1): 243-274.
- Peecher, M. E., M. D. Piercey, J. S. Rich, and R. M. Tubbs. 2009. The effects of a supervisor's active intervention in subordinates' judgments, directional goals, and perceived technical knowledge advantage on audit team judgments. Working paper, University of Illinois at Urbana-Champaign, University of Massachusetts Amherst, Illinois State University, and University of Iowa.
- 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.
- PCAOB. 2004. Audit documentation. *PCAOB Auditing Standard No. 3*, Washington, DC: U. S. 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, DC: U. S. 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, DC: U. S. Public Company Accounting Oversight Board.

- 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. Figure 2 adapted with permission from Elsevier.
- Ricchiute, D. N. 1999. The effect of audit seniors' decisions on working paper documentation and on partners' decisions. *Accounting, Organizations and Society* 24 (2): 155-171.
- Ricchiute, D. N. 2006. *Auditing*, 8th edition, Thomson South-Western.
- 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.
- Salterio, S. E. 2008. A strategy for dealing with financial reporting fraud: Fewer mandates, more auditing. *Accounting Perspectives* 7 (2): 111-122.
- Salterio, S. E. and L. Koonce. 1997. The persuasiveness of audit evidence: The case of accounting policy decisions. *Accounting, Organizations and Society* 22 (6): 573-587.
- 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 UK 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.
- Sunder, S. 2005. Minding our manners: Accounting as social norms. *The British Accounting Review* 37: 367-368.
- Tabachnick, B. G. and L. S. Fidell. 2001. *Using Multivariate Statistics*, 4th ed. Needham Heights, MA: Allyn & Bacon.
- Tan, S. K., and H. T. Tan. 2008. Effects of exposure to subsequently invalidated evidence on judgements of audit workpaper preparers and reviewers. *Contemporary Accounting Research* 25 (3): 921-946.
- Tan, H. T. and K. T. Trotman. 2003. Reviewers' responses to anticipated stylization attempts by preparers of audit workpapers. *Accounting Review* 28 (2): 581-604.
- 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.

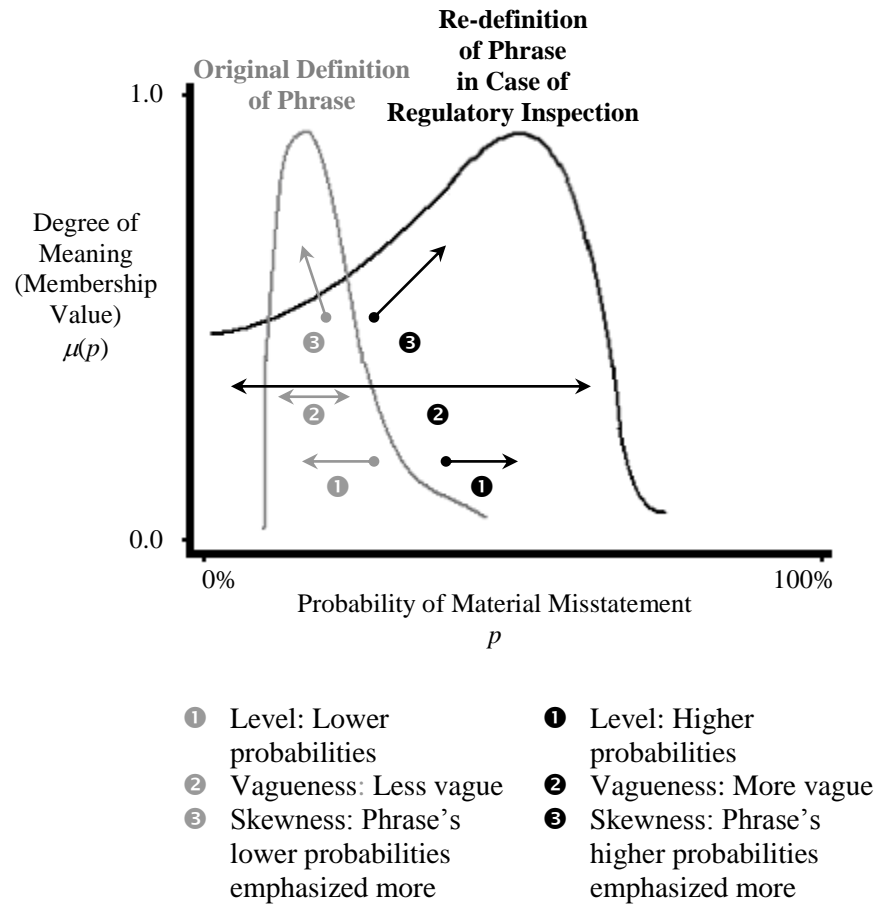
- Teigen, K. H. and W. Brun. 1999. The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes* 80 (2): 155-190.
- Teigen, K. H. and W. Brun. 2000. Ambiguous probabilities: When does $p = 0.3$ 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. *Shared Cognition in Organizations: The Management of Knowledge*, ed. L. L. Thompson, J. M. Levine, and D. M. Messick. Mahwah, NJ: Lawrence Erlbaum Associates, Inc, 117-138.
- 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. Figure 1 used with permission from the American Psychological Association.
- Webster's. 2005. *Webster's New World College Dictionary*, Wiley Publishing, Inc.: Cleveland, OH.
- Wilks. J. 2002. Predecisional distortion of evidence as a consequence of real-time audit review. *The Accounting Review* 77 (1): 51-72.

Figure 1 Membership functions for two probability phrases



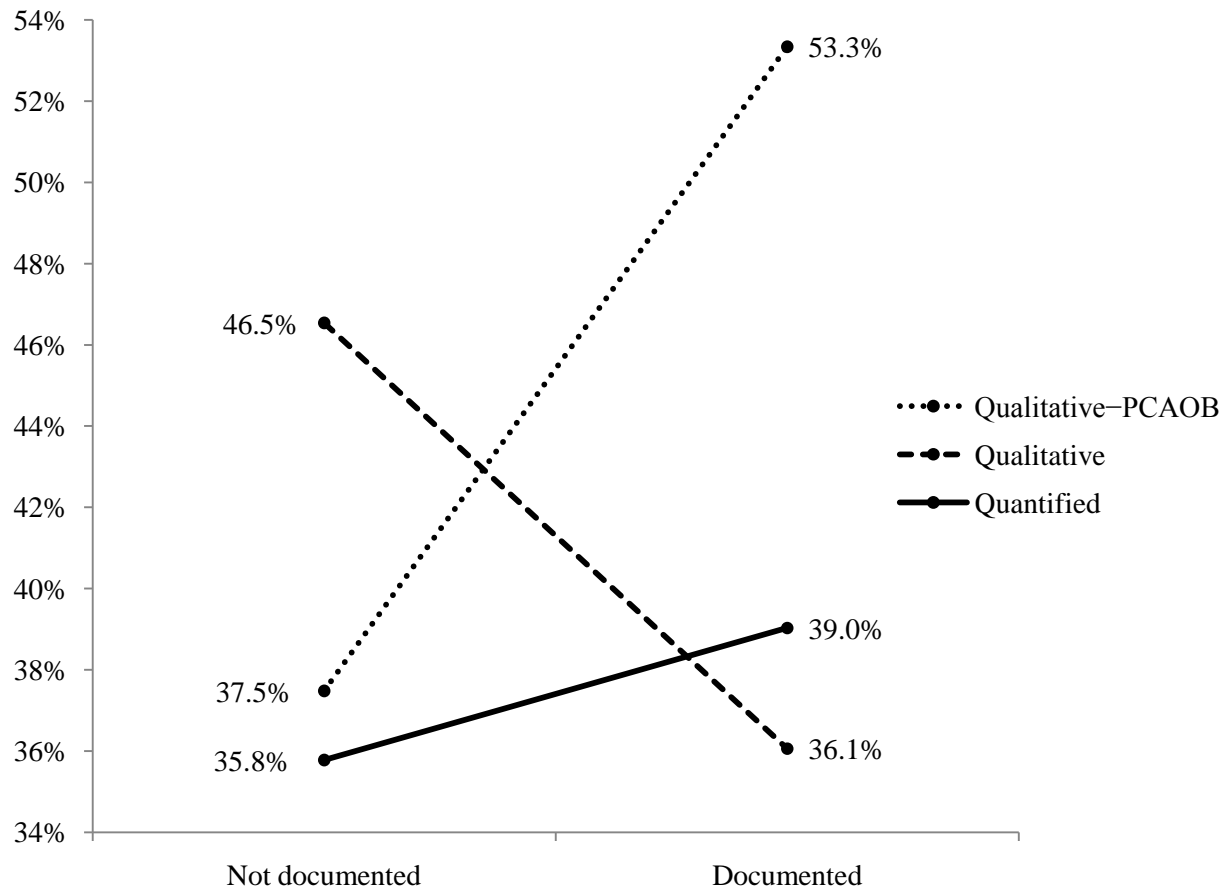
Source: Wallsten et al. (1986).

Figure 2 Elastic re-definition



Source: Adapted from Piercey (2009).

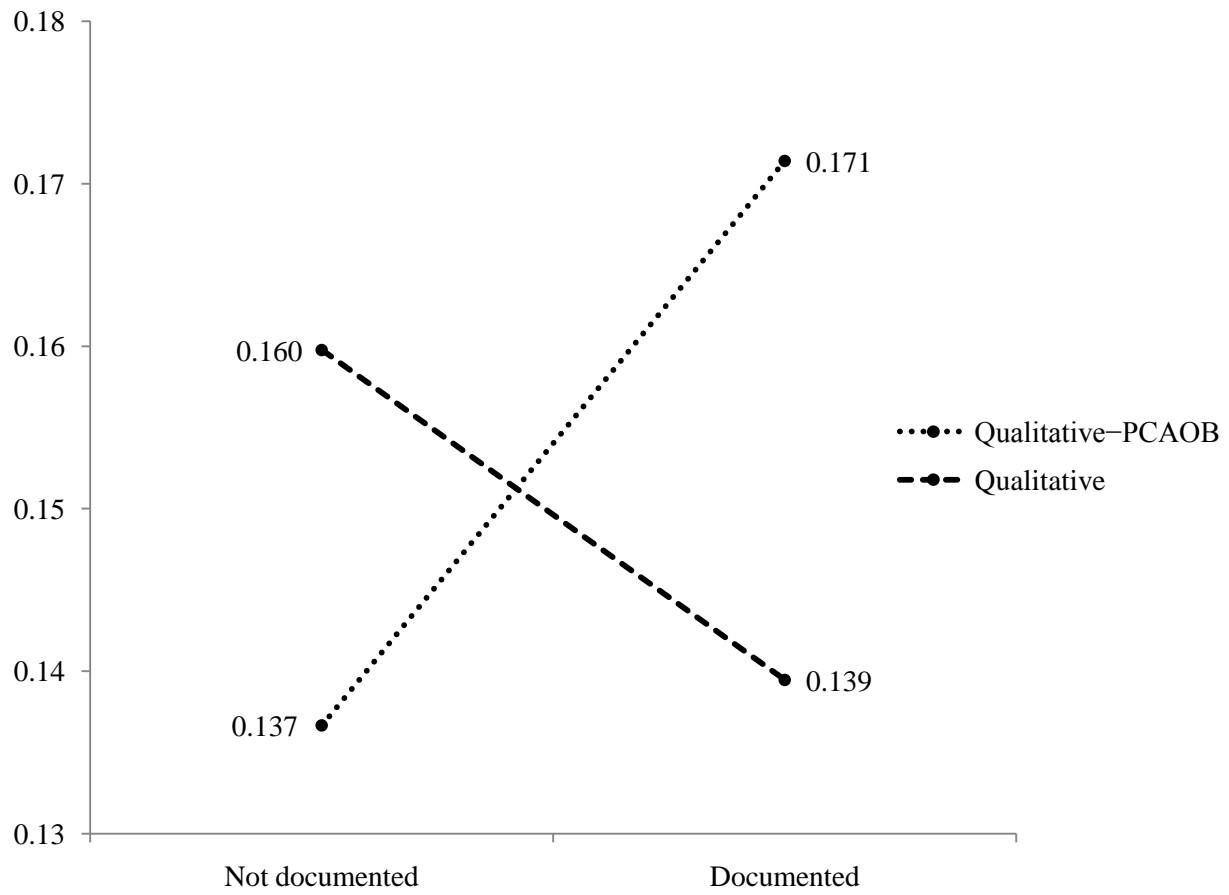
Figure 3 *Peak probability of material misstatement*



Note:

Variable definitions are in Appendix A.

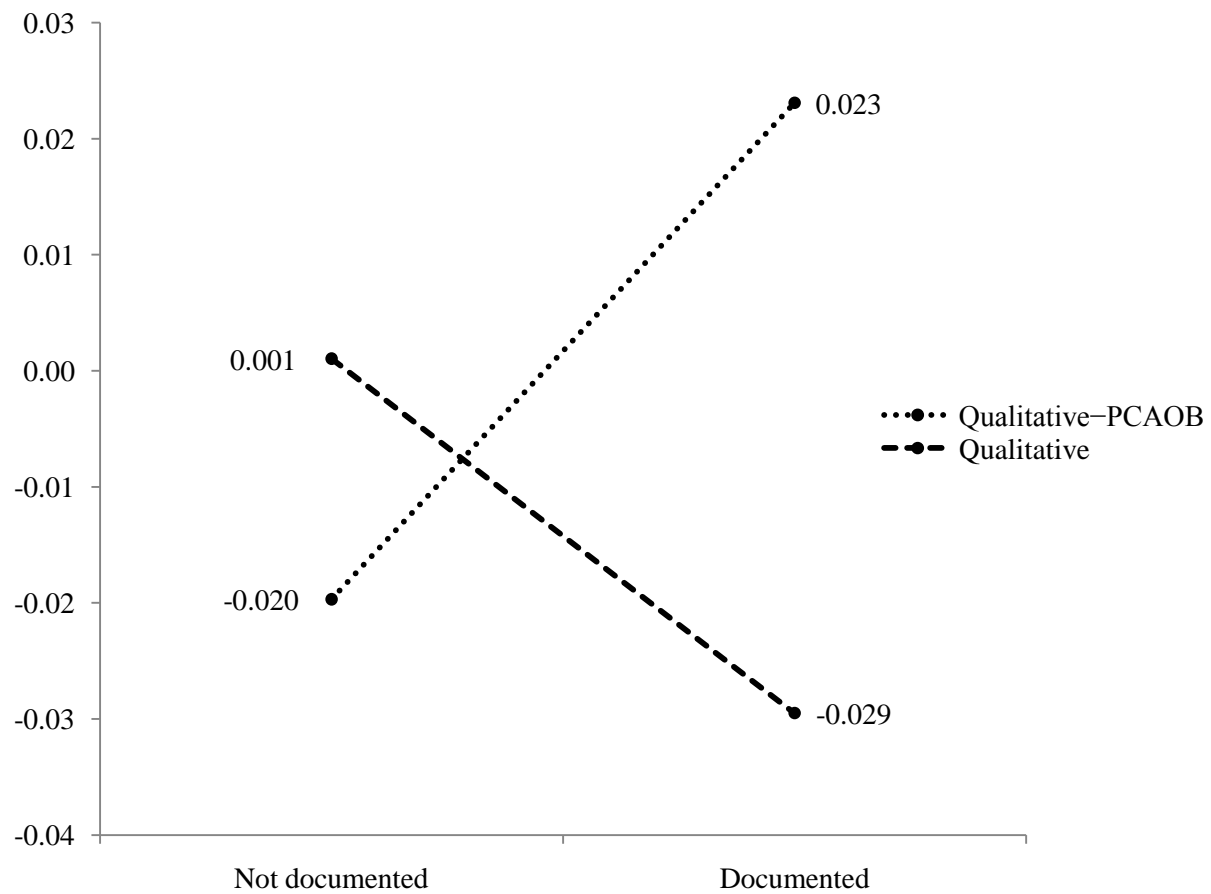
Figure 4 *Vagueness of probability of material misstatement*



Note:

Variable definitions are in Appendix A.

Figure 5 *Skewness of probability of material misstatement*



Note:

Variable definitions are in Appendix A.

TABLE 1
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	0.00
<i>Subject 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 ^a
<i>Persuasiveness</i>	199.9	1	199.9	4.59	0.02 ^a
<i>Frankness</i>	240.0	1	240.0	5.52	0.01 ^a
Error	4785.2	110	43.5		

Panel C: *Vagueness* 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 ^a
Error	14409.6	114	126.4		

Notes:

Variable definitions are in Appendix A.

^a 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., Bowlin, Hales, and Kachelmeier 2009, 76). For a discussion of one-tailed tests of directionally predicted interactions, see McNeil, Newman, and Kelley (1996, 137-139).