Decision Aids In Control Risk Assessments: Does The Metric Matter?

Somnath Bhattacharya, (E-mail: sbhatt@fau.edu), Florida Atlantic University
Ronald E. Marden, (E-mail: mardenre@appstate.edu), Appalachian State University

Abstract

Auditors have witnessed rapid growth in judgment facilitating decision aids. In light of the litigious nature of the profession, this research sounds a cautionary note with regard to the development of control risk assessment decision aids. Our results indicate that the linguistic and numeric techniques can yield different risk assessments depending on factors such as the strength of the control environment, the type of account, and the management assertion being examined. While neither technique is consistently superior to the other, the results suggest further exploration of the inconsistencies in these assessments prior to their development and use.

Introduction

Audit risk assessment has increasingly come to depend on decision aids designed to guide and structure such judgments (Boritz 1985; Dirsmith and Haskins 1991; Graham, et al. 1991; Houghton and Fogarty 1991; Jannell and Wright 1991; Dilla and Stone 1997). This is because the assessment and communication of uncertainty and probability occupies an expanding role in a domain characterized by increasing complexity and technological sophistication. In such a world decision aids are often designed to complement and compensate for human cognitive limitations. Dilla and Stone (1997) have asserted that audit decision aids often restructure the information used as inputs to decision making, as well as the form in which decisions are expressed and communicated (Ashton and Willingham 1989; Dirsmith and

Haskins 1991; Messier 1995). Some of this restructuring frequently involves converting decision inputs from qualitative (i.e. linguistic) to quantitative (i.e. numeric) representations (Libby and Libby 1989; Dilla and Stone 1997).

Linguistic to numeric conversions are usually rooted in the conviction that decision makers can improve the precision of their judgments by expressing them in numbers instead of words (Birnberg 1964; Boritz, et al. 1987a; Lichtenstein and Newman 1967; Chesley 1979; Chesley 1985; Argenti 1976; Argenti 1980; Nakao and Axelrod 1983; Dirsmith and Haskins 1991; Graham, et al. 1991; McFadgen 1994; Rittenberg and Schwieger 1994; Dilla and Stone 1997). Sometimes decision aids capture judgments in qualitative terms and then convert them to numerical equivalents for subsequent use (Eining, et al. 1994). However, little is known about the effects of information representation

Readers with comments or questions are encouraged to contact the authors via e-mail.

changes on decision makers' risk judgments (Dilla and Stone 1997). And yet, such knowledge is of fundamental importance to the development of adequate decision aids. This is because prior research suggests that small changes in information representation have large effects on decision processes and judgments (Payne, et al. 1992; Kleinmuntz and Schkade 1993; Dilla and Stone 1997). This leads to the possibility that to the extent that we are unable to determine the effects of information representation on various facets of audit judgment (for example, areas such as audit risk assessment), auditors may be vulnerable to changes in their judgments shaped simply by changes in the representation metrics themselves. This possibility is further advanced by prior studies that provide evidence that decision input representation does affect judgment processes and performance (Bell 1984; Svenson and Karlsson 1986; Stone and Schkade 1991; Schkade and Kleinmuntz 1994). Other studies have indicated that the representation used to communicate decision outputs can also affect (auditors') risk judgments (Waller 1991; Emby 1993; Reimers, et al. 1993).

The above studies, therefore, suggest that differences in risk assessments may be wrought simply as a function of the representation metrics themselves. Not only does this hold strong implications for the development of risk assessment decision aids in auditing, but furthermore, the (in)adequacy of such decision aids has implications for related issues such as the suitability of one metric over another in different risk assessment contexts, audit sample composition and size, audit strategy effectiveness and efficiency, legal liability, etc. All of these issues, however, hinge on whether audit risk assessment is indeed vulnerable to differences in the risk assessment metrics themselves. Hence, the research premise for this paper is simple: "Does the Metric Matter?" Our contention is that this question must be addressed prior to further decision aid development in the area(s) of audit risk assessment.

Premise

We find that a majority of the largest accounting firms use linguistic scales for assessing components of audit risk (Reimers, et. al. 1993). Contrasting this prevalent audit practice is prior research that argues that numeric scales are better than linguistic scales for improving experienced auditors' risk assessments (Stone and Dilla 1994). We set out to examine whether numeric and linguistic risk assessment scales, given the same audit decision-making context, yield significantly different risk assessments. This research question is posed against the backdrop of a seeming contradiction between prior research and prevalent practice. Our contention is that if they do, then such differences will have been fashioned simply as a function of the scales themselves. If such were the case, then auditors would need to contend with the possibility that audit judgment, as it pertains to risk assessment, may be manipulated simply by varying the metrics themselves. This would further lead to the issue of suitability of one metric (scale) over another in various audit contexts. Such judgment differences would hold implications for risk-assessment decision aid development and use and other related audit issues such as audit sample selection, audit strategy, etc.

Dilla and Stone (1997) have indicated that the processes by which auditors combine cues into risk judgments have been of keen interest to audit researchers (Jiambalvo and Waller 1984; Daniel 1988; Libby and Libby 1989). They have further shown how this body of research fits well into phase 2 of Einhorn and Hogarth's (1981) decision-making model (see Messier 1995, for a succinct discussion of the four phases of this model). Furthermore, they have asserted the importance of assessing the audit implications of risk assessment cues (phase 1 of the Einhorn and Hogarth 1981 model) and the subsequent communication of risk judgments (phase 3 of the same model) (Graham 1985; Arens and Loebbecke 1991). This study takes its cue from that assertion and focuses on the

communication of risk judgments. To frame the study we use five management assertions across several objective and subjective types of accounts and *a priori* information on the state of a financial institution's internal controls. Auditors' assessments occur under both weak and strong control environment conditions. Our analyses involve comparing subjects' numerical risk assessments to the midpoints of the numeric ranges representing their self-reported linguistic assessments. Our results suggest that control risk assessment differences may surface only under complex conditions of interactions between different account types, management assertions, and environmental conditions.

The next section discusses the assessment of audit risk and the use of qualitative and quantitative scales for control risk assessment. Because control risk is a specific component of overall audit risk, we believe that a discussion of audit risk prior to the discussion on control risk assessment is apropos for framing of the control risk assessment issue used in this paper. Sections on the experimental framework, research question development, research methodology, results, discussion, and concluding remarks follow in their stated order.

The Assessment of Risk

A succession of professional pronouncements (Statements on Auditing Standards (SAS) No. 22, *Planning and Supervision*, AICPA 1978; SAS No. 39, *Audit Sampling*, AICPA 1981; SAS No. 47, *Audit Risk and Materiality in Conducting an Audit*, AICPA 1983; SAS No. 55, *Consideration of the Internal Control Structure in a Financial Statement Audit*, AICPA 1988ⁱ and SAS No. 78, *Consideration of Internal Control in a Financial Statement Audit: An Amendment to SAS No. 55*, AICPA 1996) have established the importance of considering audit risk in planning and executing financial statement audits (Mock and Wright 1993). Furthermore, these pronouncements have considered audit risk through its component parts of inher-

ent risk, control risk, and detection risk (Haskins and Dirsmith 1993). Martinov and Roebuck (1998) have further suggested that the inherent risk, control risk, and detection risk framework is used to focus on the initial risk and materiality assessments of individual accounts or classes of transactions. Furthermore, given that the level of inherent risk and control risk is a function of the client, and independent of the audit function, the levels of these risks must be evaluated prior to planning materiality and the establishment of detection risk. This will ensure that the achieved level of audit risk is kept within acceptable limits (Martinov and Roebuck 1998).

According to SAS No. 47 (AICPA 1983), the components of audit risk (e.g., control risk, inherent risk, detection risk) may be assessed in quantitative (numeric) terms such as percentages or in qualitative (linguistic) terms that range from a minimum to a maximum. Dilla and Stone (1997) ask: "Does it matter whether decision makers use numbers or words to: (1) represent the information used in making risk judgments and (2) express and communicate risk judgments?" Prior literature suggests an affirmative answer to both of the above questions. For example, Libby and Libby (1989) found that mechanically combined judgments used in control reliance decisions yielded greater consensus than unaided control reliance judgments. Answering the latter question, Waller (1991) had auditors express combined control and inherent risk assessments using both a continuous (0-100) and a discrete (low, moderate, high) scale. Waller (1991) found that auditors had lower confidence in their judgments when using numbers as opposed to words to communicate their judgments.

Dilla and Stone (1997) assert that whether practicing auditors should use numeric or linguistic risk expressions is an important issue. This is because academics have repeatedly argued that auditors could increase the precision of their judgments by using numerical scales instead of linguistic ones (Boritz, et al. 1987a,

1987b; Chesley 1979, 1985). On the other hand, efforts to quantify risk judgments have met with opposition from practicing auditors (Dirsmith and Haskins 1991). And while a large majority of Canadian (Boritz et al. 1987b) and United States (Jannell and Wright 1991) firms continue to use words instead of numbers to express risk judgments (Dilla and Stone 1997), Reimers, et al. (1993) note that at least two of them appear to involve some numerical assessments. The authors add that the reason the majority of firms prefer the linguistic scale is that they perceive the numerical assessments to be unjustifiably precise. Stone and Dilla (1994) add that while many of these risk assessments are made in linguistic terms, firms often use sampling decision aids which convert linguistic risk judgments back to numeric values, but without explicitly disclosing the numeric vs. linguistic equivalencies to the auditor in the field (Taylor 1997). This conversion is often intended to increase agreement among auditors and audit engagements.

The use of agreement among auditors and engagements as a metric is based on the argument that the weight given to a judgment cue depends on its ability to convey valence or evaluative meaning (Artz and Tybout 1990). A particular advantage of representing cues in words may be their ability to directly communicate the evaluative meaning of data. For example, trained auditors may share a well-defined lexicon that conveys specific meaning to scenarios painted in the shared vocabulary. Thus, the phrase: "...the financial institution's accounting system and control procedures...are in place and operating effectively" (the scenario used in this study) may convey a more complete description of the audit scenario than the accounting system and control procedures' effectiveness conveyed through numerical percentages. This is because a numerical statement of 70% effectiveness, for instance, may yet leave unanswered questions such as whether 70% is common in the financial institution industry; whether it is less (more) than what is expected in the industry, etc. Therefore,

even if information cues represented in words are less precise than those represented in numbers (Stone and Schkade 1991), the meaning of particular words may be interpreted more consistently within a given judgment context. Therefore, using words as representation for inputs to audit decisions may increase judgmental consensus because auditors agree more on the relative importance (evaluative meaning) of cues represented in words rather than in numbers (Dilla and Stone 1997).

Countering the above argument, however, is research that indicates that individuals interpret cues fairly consistently over a series of cases, regardless of whether the cues are stated in words or numbers (Fillenbaum, et al. 1991; Wallsten, et al. 1986). Therefore, according to this school, changes in information cue representation should not affect the consistency of auditors in interpreting cues over a set of cases (Dilla and Stone 1997).

Reimers, et al. (1993) investigated whether control risk assessments made using linguistic expressions differed from assessments made using numeric probabilities. Their results indicated that the types of risk assessments differed in two ways. First, the numeric assessments were significantly lower than the linguistic assessments. Second, there was a higher level of agreement among the linguistic assessments. However, Reimers, et al. (1993) suggest that this latter finding may have been a function of the fact that the numeric scale had a greater number of data points from which to choose compared to the four linguistic labels used in the experiment. In fact, Reimers, et al. (1993) concluded that audit firms should consider the potential for excessive conservatism induced by linguistic assessments in the design of their risk assessment procedures.

Reimers, et al.'s (1993) results are consistent with those of prior studies (Beyth-Marom 1982; Chesley 1979, 1985; Hamm 1991; Lichtenstein and Newman 1967; Nakao and Axelrod

tenstein and Newman 1967; Nakao and Axelrod 1983; Reimers 1992) that suggest that words tend to be ambiguous in communicating uncertainty. For example, if a linguistic scaling label of "high" control risk assessment covers a numeric range of 70 percent -to- 100 percent risk, then two auditors providing the same "high" risk assessment could have a theoretical difference of as much as 30 percent points between them. This difference in control risk assessment could have a direct bearing on the calculation of detection risk in audit practice. Furthermore, other audit issues such as sample size determination for collecting evidence may be significantly affected merely as a function of the assessment method (quantitative versus qualitative). Differences in control risk assessments wrought simply as a function of the assessment method employed hold further implications for issues such as the development of risk assessment decision aids (the subject of this paper), that incorporate and subsume issues of audit strategy, sample size composition and determination, etc., in them.

Experimental Framework

Given the wide spectrum of results obtained from prior risk assessment studies, and specifically those obtained by Reimers, et al. (1993), we seek to examine whether or not differential control risk assessment metrics are capable of yielding differential control risk assessments simply as a function of the metrics themselves.

To that end, we ask our subjects to assess control risk for a risky industry (the financial institutions industry), for both subjective (more risky accounts such as Loans and Foreclosed Real Estate) and objective (less risky accounts such as Cash and Deposits) accounts, over strong and weak control environments, and five specific management assertions. We believe that a comparison of the quantitative and the qualitative risk assessment methods over two types of accounts, two control environments, and five management assertions may yield complex

interactions between the control risk assessment methods, accounts, assertions, and control environment conditions. The manipulation of a risky industry, differential control environments, accounts, and assertions are based on the following discussion.

Financial Institutions

We follow Marden, et. al. (1997) in using the financial institution industry because of the industry's problems with failures over the last decade. Estimates as high as \$500 billion over the next 30 years to fund the industry bailout (Bowsher 1990) have already reached roughly \$200 billion (Day 1998). Some critics claim that the crisis might either have been averted or substantially diminished had auditors been more sensitive to signs such as management lacking integrity or management philosophy and attitudes placing companies at high risk for fraud and insider abuse (GAO 1989). Given the recent notoriety of financial institutions and their use in an associated context by Marden, et al. (1997), we chose this industry as representative of a risky audit environment for this study.

Control Environment

There has been little published research on the control environment's (CE) effect on the assessment of control and inherent risk (Marden, et. al. 1997). Yet the control environment can have a pervasive effect on the entire audit process because it is the primary component in the evaluation of control risk and shares many of the same factors that the auditor considers in assessing control risk. The control environment may be evaluated through the consideration of its primary components.

Subjective-Objective Accounts

Some accounts, particularly accounts that are based in part on management estimates, tend to be more subjective than others. SAS No. 57, *Auditing Accounting Estimates*, states that

the auditor is responsible for evaluating the reasonableness of accounting estimates made by management in the context of financial statements taken as a whole (AICPA 1996, AU 342). However, because of the uncertainties associated with estimates and valuations, the audit risks associated with these types of accounts are often much greater than for accounts which can be objectively verified (e.g., with third parties). Evaluating the reasonableness of management estimates (and the associated internal controls over how these estimates are made) is one of the most difficult areas of audit evidence, and it requires the greatest skill on the part of the auditor (Mautz and Sharaf 1961).

In this study four accounts were used: two that were considered to be more subjective in audit evaluation, and two that were considered to be less subjective (objective). Accounts were chosen from opposite ends of an objective/subjective continuum in order to examine whether changes in CE condition had a differential effect on the risk assessment of one account *type* versus another. The accounts were selected based on the results of a pilot test.²

The pilot test indicated that of 12 commonly used financial institution accounts, the accounts "Loans" and "Real Estate Owned" were most subjective in nature, and the accounts "Cash and Due From Banks" and "Deposits and Savings Accounts" were considered the least subjective. The choice of "Loans" and "Real Estate Owned" as the most subjective accounts is reasonable given that one of the more difficult judgment areas in auditing financial institutions is determining the net realizable value of loans which includes an estimated allowance for uncollectible loans (i.e., loan loss reserves). Estimating the correct valuation for the loan-loss reserve account is a critical part of the audit and aids in understanding an institution's financial condition. This is because the loans account is often the largest asset in the financial statements and loan loss reserves are a major indicator of a depository institution's loss exposure.

As for the Real Estate owned account (REO), generally the largest component of REO held by financial institutions is foreclosed real estate. Risks associated with foreclosed real estate are often significant because of the high degree of subjectivity involved in determining real estate values. For example, given the same piece of property, several real estate appraisals will likely differ. Choosing which appraised value is fairest often becomes a subjective judgment debated between the client and auditor. To complicate matters further, the property assessments themselves are subject to outside and potentially volatile factors such as interest rates. Similarly, it also seems reasonable that Cash and Deposits would be considered more objective (less subjective) in nature because they are less likely to require auditors to address the subjective issues of estimations and valuations frequently associated with the other two accounts.

Management Assertions

The third independent variable consisted of the five management assertions: existence and occurrence, completeness, rights and obligations, valuation and allocation, and presentation and disclosure (SAS No. 31 AICPA 1980). Marden, et al. (1997) found that on average, the valuation assertion was assessed at a very high risk, but only for more subjective accounts. For the more objective accounts, this assertion was assessed at a lower risk than the other assertions. The non-valuation assertions were not as sensitive to changes in the control environment condition and account types, when compared to the valuation assertion. This pattern of a higher risk assessment for the valuation assertion is also consistent with Braun (1998), who found it "readily apparent" that subjects considered the valuation assertion to be of higher risk than the other assertions. Given the exploratory nature of this study, we consider the effect of the valuation assertion both in conjunction with the other assertions, and also separately by itself.

Dependent Variables

We use two dependent variables in this study. The first is the numeric assessment of control risk. The control risk assessment was made on a 0-100 percent scale where 0 percent = no risk and 100 percent = maximum risk. For the second dependent variable, subjects were asked to convert their firm's linguistic labels of "High," "Moderate," and "Low" risk into numerical values on a 0-100 percent numeric scale. On average, the subjects believed that the label "High Risk" started at 73.2 percent (s.d. = 6.6) and rose to 100 percent (median 86 percent). The label "Moderate Risk" started at 33.8 percent (s.d. = 6.8) and rose to 73.1 percent (median = 53 percent). The label "Low Risk" started at 0 percent and rose to 33.7 percent (s.d. = 6.2, median = 17 percent). These mid-range values are consistent with the median numbers collected by Amer, et. al. (1994) in their study of the interpretations of probability phases. In that study, the terms "High," "Moderate," and "Low," which were taken from the 1991 KPMG Peat Marwick Audit Manual, were interpreted by 49 audit managers as being 85 percent, 50 percent, and 20 percent, respectively.

Research Questions

There are few prior accounting/auditing studies that have examined the effects of response mode differences on accounting/audit judgment and communication. Dilla and Stone (1992) investigated response mode judgment differences in auditors' inherent risk assessments. Their results suggest a marginally higher degree of auditor consensus in inherent risk judgments communicated numerically than those conveyed linguistically. Auditors in that study also took longer to make numerical risk assessments than to make linguistic assessments. However, that study did not address whether the magnitude of inherent risk assessments was affected by response mode.

In a similar vein, Dilla and Stone (1997)

examined the asymmetric effects wrought from the use of words versus numbers when used to represent risk cue information and when used to communicate risk judgments. They found that presenting judgment cues in words as opposed to numbers increased agreement on cue weights but did not affect judgment consistency. On the other hand, numeric response representation increased judgment consistency relative to the linguistic response representation, but agreement on cue weights was not affected by response representation. However, this study too did not directly examine whether the magnitude of risk assessments was affected by response mode.

Reimers, et al. (1993) did examine the response mode effects on the magnitude of control risk assessments. They based their study on the results of Budescu, et al. (1988) that suggested that linguistic responses should be more extreme (biased) than numerical responses – especially in a loss domain where more extreme judgments are judgments of higher audit risks. Reimers, et al. (1993) used a between-subjects two-phase design for a designer/manufacturer/retailer of construction machinery.

The Reimers, et al. (1993) results suggest that the response-mode effects (bias) elicited therein should become more pronounced in a risky industry environment. This would also be in keeping with Budescu, et al. (1988). Therefore, given our use of the financial institutions industry (representative of a risky industry), we expect (in alternate form) that:

RQ: In the financial institution industry, linguistic control risk assessments will yield higher risk assessments than numeric control risk assessments.

The Reimers, et al. (1993) study along with others such as Beyth-Marom 1982, Chesley 1979, Chesley 1985, Hamm 1991, Lichtenstein and Newman 1967, Nakao and Axelrod 1983, and Reimers 1992 suggest that words tend to be ambiguous in communicating uncertainty.

Therefore, if the numeric assessment technique indeed allows for greater precision in the risk assessment process, then for auditors faced with a risky (less risky) control environment, the numeric method should allow for a more precise assessment of control risk. This would occur because the numeric technique would allow auditors to reduce the vagueness of any linguistic assessment category and instead focus on a single point within the restricted interval of probabilities deemed representative of that category. For example, a linguistic label of “high” risk may cover a risk-range between 70 percent and 100 percent. Using the numeric assessment technique would allow the auditor to pinpoint a more precise risk level within this restricted interval of probabilities.

We use the control environment as a manipulation variable to surface the differences, if any, between the two risk assessment methods. Marden, et al. (1997) suggests that the control environment can have a pervasive effect on the assessment of control and inherent risk. Kreutzfeldt and Wallace (1990) found that a client’s control environment was significantly related to the incidence of account error. Furthermore, Haskins and Dirsmith (1993) found that independent auditors ascribe similar relevance to a relatively large number of specific client attributes in assessing the control environment and inherent risk. And, for those client attributes viewed as equally relevant to assessing control environment and inherent risk, auditors tend to ascribe similar relevance to them for control (and inherent) risk assessments. These prior studies, therefore, suggest that the control environment should have a singular impact on auditors’ control risk assessments. We further expect that the control environment condition will interrelate with the response mode of the risk assessments. However, there is scant *a priori* evidence to suggest the manner in which the qualitative bias referred to earlier in the discussion leading up to RQ1 would now work in conjunction with the control environment. In other words, faced with a strong control envi-

ronment, the qualitative bias may prejudice the qualitative risk assessment lower than its numeric counterpart. On the other hand, it is possible that the qualitative bias only works in a loss (high-risk) domain (as suggested by a weak control environment). Therefore, following the discussion preceding RQ1, we would expect to see higher (than quantitative) qualitative control risk assessments in a weak control environment condition. For the strong control environment condition, however, we are unable to generate any *a priori* directional expectation. Therefore, we posit (again in alternate form) that:

RQ2A: Given a strong control environment, auditors’ numeric control risk assessments will *differ* from their linguistic control risk assessments.

RQ2B: Given a weak control environment, auditors’ numeric control risk assessments will be lower than their linguistic control risk assessments.

Because of the uncertainties and subjectivity associated with management estimates and valuations, the risks associated with these types of accounts are often much greater than for accounts which can be objectively verified (for example, with outside third parties). Thus, for auditors working with subjective types of accounts, the ability to express risk assessments as precisely as possible becomes paramount. This precision may be compromised by linguistic risk assessments.

On the other hand, the discussion leading up to RQ1 above suggests that linguistic responses tend to be more extreme (biased) than numerical responses – especially in a loss domain where more extreme judgments are judgments of higher audit risks. This would lead us to expect that faced with subjective (more risky) accounts, auditors’ qualitative risk assessments would be higher than their quantitative risk assessments. For the more objective type of accounts, however, it is again difficult to frame

any *a priori* directional expectation of the manner in which the qualitative bias may function. Hence, we posit (in the alternate form) that:

RQ3A: For more objective type of accounts, auditors' numeric control risk assessments will *differ* from their linguistic control risk assessments.

RQ3B: For more subjective type of accounts, auditors' numeric control risk assessments will be lower than their linguistic control risk assessments.

Finally, the valuation assertion is often the most difficult to assess of the five management assertions because it is frequently based on management estimates and other assumptions. Furthermore, a stable valuation metric may not hold over different account types and environmental conditions. In fact, Marden, et al. (1997) found that when auditors were asked to assess control risk across several account types and control environment conditions, the assessed risks for the valuation assertion tended to be the highest of the management assertions for subjective accounts, but lowest for the objective accounts. Based on this interaction, we believe that significant variability in risk assessments for the valuation assertion may become lost in the averaging that the linguistic risk assessment method involves.

On the other hand, however, given that linguistic responses tend to be more extreme (biased) than numerical responses in high-risk domains, auditors' qualitative control risk assessment for the valuation assertion for subjective accounts would be expected to be higher than the numerical assessments. For the more objective accounts, however, it is again difficult to generate any *a priori* directional expectations for the valuation assertion. Hence we posit (in alternate form) that:

RQ4A: For more objective type of accounts, auditors' numeric control risk assessments for

the valuation assertion will *differ* from their linguistic control risk assessments.

RQ4B: For more subjective type of accounts, auditors' numeric control risk assessments for the valuation assertion will be lower than their linguistic control risk assessments.

Research Methodology

The Consistency Metric

The risk-based approach to auditing requires the nature, timing, and extent of audit tests to be determined by assessing and evaluating the risk that financial statements are materially misstated (Cushing, et. al. 1995). This approach is based on the concern for *consistency* across audit engagements. "Thus, while the procedures to be performed might vary with the circumstances of each engagement, the process of determining which procedures would be performed, on which assertions they would be based, and to what extent they would be performed, would not vary" (Cushing, et. al. 1995). The concept of "risk," therefore, is involved in the quest for *consistent* selection of the required auditing procedures.

Consistency as a metric is useful not only *between* auditors, but also *among* auditors. In other words, one could study the consistency patterns (or lack thereof) of auditors' judgments between auditors on the same engagement, or between auditors on different engagements (a measure of consensus). Alternatively, one could study the consistency patterns of auditors' judgments by the same auditors on different engagements, or by the same auditors on the same engagement (a measure of consistency). This last metric (measuring the same auditors on the same engagement) is the one used in this study to examine whether auditors' judgments on a given audit scenario is also a function of the metric(s) used to gauge such judgments. Hence, we intend to examine the implications that the two audit judgment scales (linguistic versus numeric) hold

for auditor consistency with regard to control risk assessments.

Subjects

The experiments were administered to 40 practicing auditors attending Big-Six national financial institutions training seminar. The subjects were in-charge seniors with two to four years of financial institution audit experience.

Task

Information on the condition of the internal control structure of two fictitious firms was provided to subjects in a case packet that described (a) the Control Environment (CE) components and their relative strengths or weaknesses (pre-rated on a 1-10 control environment condition scale where 1 = "very weak" and 10 = "very strong") and (b) the conditions of the accounting system and control procedures (held constant by informing subjects they were "in place and operating effectively").

Each subject received two different cases with each case referring to a different financial institution. For each financial institution case, the strengths and weaknesses of the control environment components were varied so that one financial institution had a relatively weak control environment and the other financial institution had a relatively strong control environment. Subjects were then asked to assess control risk for four different accounts at the individual assertion level.³

The Control Environment Components

The individual CE component rating scores were varied by providing either: high ("7", "8" or "9" on the "1 to 10" rating scale) or low scores (e.g., "2", "3" or "4") which indicated their contribution in defining the CE. The higher scores indicated a stronger CE and the lower scores indicated a weaker CE condition. It is important to note that in each case the

extreme ends of the weak to strong scale (i.e., 1 or 10) were not used so that the component manipulations would not be as obvious to the subjects.

Experimental Design

The data were analyzed using MANOVA with a (2x2x2x5) within-subjects factorial design: Control Environment condition (Strong, Weak) x Response Mode (Numeric, Linguistic) x Type (Objective, Subjective) x Management Assertion (Existence, Completeness, Rights and Obligations, Valuation, Presentation and Disclosure). The reasons for selecting a totally within-subjects design are two-fold. First, there was the practicality of obtaining enough subjects (experienced in the financial institution industry) to fill the necessary number of cells that would be required for a between-subjects design and the complexity of the experiment. Second, the within-subjects design provides more power to the analysis by controlling for the individual differences between the subjects' risk assessments (Keppel 1992). And while carryover and practice effects are always a concern with within-subjects designs, we assert that auditors frequently make repeated judgments (Pany and Reckers 1987) and that our subjects were experienced in this decision making task.

Each subject first provided a numeric assessment of control risk for 4 different accounts at 5 assertion levels. This procedure was repeated twice, once for each CE condition. The order in which the subjects received the accounts and control environment conditions was randomized using a digram-balanced Latin square. This is useful in research designs where repeated measures are involved, because of the potential presence of nuisance factors such as carryover and practice effects (Keppel 1992). This risk is particularly high when experimental measures are not randomly sequenced (Harsha and Knapp 1990; Pany and Reckers 1987). An advantage of using a digram-balanced Latin square is that it allows for a different ordering of the treatments

using a minimum number of orders where each condition immediately precedes and follows the other conditions only once (Wagenaar 1969).

In order to analyze the data each numeric assessment of control risk was matched with a converted linguistic-to-numeric assessment. For example, subject "A" may have considered the linguistic label of "High" risk assessment to encompass the numeric probability range of 70-100 percent. Now, if subject A's original numeric assessment fell within that range (for instance at 82 percent), it would be matched with the *mid-point* or median of the 70-100 percent range or 85 percent. The use of median points to represent ranges associated with linguistic labels has been used successfully in prior research in this area (e.g., Dusenbury, et. al. 1996; Reimers, et. al. 1993). Thus, for every individual's numeric assessment of risk, there was a matched *mid-point* "converted" score based on that subject's linguistic-to-numeric conversion.

Our reason for matching converted linguistic scores *after* the numeric scores had been assessed was our concern (because of the within-subjects design) that the subjects might be influenced by the process of using both assessment methods together. As was stated earlier, firms often use sampling decision aids without explicitly disclosing the linguistic to numeric equivalencies to the auditor in the field (Taylor 1997). Thus, by asking the subjects to separately indicate their own numeric range for each linguistic label, and then matching those individual labels with their original numeric assessments, we have attempted to mitigate the differences *between* subjects in interpreting the size of the numeric ranges associated with each linguistic label.

Further, we were also concerned that as experimental conditions changed (i.e., as the control environment changed from weak to strong and vice versa), there was a potential for individual subjects to modify their own interpretations of the ranges that defined each linguistic

label as they made their numeric assessments. We believed that having subjects identify the ranges associated with each linguistic label *after* using the numeric risk assessment method in the experiment would mitigate the potential *within* subject bias of formulating a separate (and somewhat different) scale each time a numeric assessment was made.

Lastly, the Account *Type* variables were constructed by averaging the loans and foreclosed real estate accounts into the "subjective" account type, and by averaging the cash and deposits accounts into the "objective" account type. Combining two accounts into one account type variable mitigated the possibility of one particular account driving the results. Separate analyses of the output revealed no significant differences between the two subjective accounts or between the two objective accounts.

Results

Our first research question proposed that based on the Reimers, et al. (1993) and the Budescu, et al. (1988) studies, the qualitative bias should become more pronounced in a risky industry environment. Therefore, given our use of the financial institutions industry (representative of a risky industry), we expected (in alternate form) that:

RQ1: In the financial institution industry, linguistic control risk assessments will yield higher risk assessments than numeric control risk assessments.

Table 1 supports research question 1. We find that the average control risk assessment in a numeric mode is at 47.78 on a scale of 0-100, whereas the average linguistic assessment stands at 48.59. The response mode means differ significantly [$p=.001$] and the mean linguistic control risk assessments are indeed higher than their numeric counterparts.

Both research questions 2a and 2b pro-

Table 1

RQ1: In the financial institution industry, linguistic control risk assessments will yield higher risk assessments than numeric control risk assessments.

Response Mode	Mean	F	Sig of F
Linguistic	48.59	4.70	.001
Numeric	47.78		

pose a relation between response-mode (qualitative versus numeric) and the control environment (weak versus strong). Again, based on Budescu, et al. (1988) and Reimers, et al. (1993), we propose that the risky nature of the financial institutions industry would be further compounded by a weak control environment, resulting in a greater manifestation of the qualitative bias referred to earlier. Faced with a strong control environment, however, we have no *a priori* expectation of the manner in which the control environment may work with the response-mode effects. Hence, we proposed that:

RQ_{2A}: Given a strong control environment (CE), auditors' numeric control risk assessments will *differ* from their linguistic control risk assessments. And

RQ_{2B}: Given a weak control environment (CE), auditors' numeric control risk assessments will be lower than their linguistic control risk assessments.

Table 2 supports RQ_{2A} in that faced with a strong control environment, auditors' linguistic control risk assessments do differ significantly from their numeric counterparts [$p=0.031$]. More specifically, the linguistic mode average control risk assessment at 32.23 is higher than its numeric counterpart at 29.56. Table 3, however, shows results that run contrary to RQ_{2B} and suggests that given a weak control environment, auditors' numeric risk assessments at 66 are significantly higher than their qualitative counterparts at 64.94 [$p=0.010$].

Research Questions 3 A and 3 B are

TABLE 2

RQ_{2A}: Given a strong control environment (CE), auditors' numeric control risk assessments will *differ* from their linguistic control risk assessments.

Response Mode	Mean	F	Sig of F
Linguistic	32.23	5.02	.031
Numeric	29.56		

Table 3

RQ_{2B}: Given a weak control environment (CE), auditors' numeric control risk assessments will be lower than their linguistic control risk assessments.

Response Mode	Mean	F	Sig of F
Linguistic	64.94	7.31	.010
Numeric	66.00		

similar in vein to Research Questions 2A and 2B. They test the manner in which subjective (objective) types of accounts work in conjunction with response mode (linguistic versus numeric). Following similar logic to RQ_{2A} and RQ_{2B}, we propose that:

RQ_{3A}: For more objective type of accounts, auditors' numeric control risk assessments will *differ* from their linguistic control risk assessments. And

RQ_{3B}: For more subjective type of accounts, auditors' numeric control risk assessments will be lower than their linguistic control risk assessments.

In support of RQ_{3A}, table 4 suggests that faced with more objective types of accounts, the linguistic mode control risk assessments do indeed significantly differ from their numeric counterparts [$p=0.000$]. More specifically, given more objective type of accounts, the qualitative mode yields significantly lower control risk assessments (42.74) than the numeric mode (49.37). Table 5, however, suggests that when faced with more subjective type of accounts, the numeric mode yields significantly [$p=0.000$] higher control risk assessments (55.65) than the qualitative mode (54.43).

Table 4

RQ_{3A}: For more objective type of accounts, auditors' numeric control risk assessments will *differ* from their linguistic control risk assessments.

Response Mode	Mean	F	Sig of F
Linguistic	42.74	68.61	.000
Numeric	49.37		

Table 5

RQ_{3B}: For more subjective type of accounts, auditors' numeric control risk assessments will be lower than their linguistic control risk assessments.

Response Mode	Mean	F	Sig of F
Linguistic	54.43	49.20	.000
Numeric	55.65		

Prior studies like Marden, et al. (1997) and Braun (1998) have suggested that the valuation assertion is more susceptible to changes in the control structure than are other assertions. However, we have no basis for any *a priori* expectation of the manner in which the valuation assertion's sensitivity will play in conjunction with more objective accounts and response-mode effects. On the other hand, given a high-risk environment (as characterized by subjective type of accounts) we have posited that the qualitative bias discussed earlier will become more accentuated in the qualitative response mode. Therefore, formalizing our expectations, we proposed that:

RQ_{4A}: For more objective type of accounts, auditors' numeric control risk assessments for the valuation assertion will *differ* from their linguistic control risk assessments. And

RQ_{4B}: For more subjective type of accounts, auditors' numeric control risk assessments for the valuation assertion will be lower than their linguistic control risk assessments.

Table 6 indicates that for more objective types of accounts, the auditors' numeric control risk assessments for the valuation assertion do indeed significantly differ from the linguistic

mode [$p=0.000$]. More specifically, the numeric risk assessment for this assertion (33.05) faced with objective accounts is lower than that in the linguistic mode (37.35). On the other hand, Table 7 suggests that for more subjective type of accounts, auditors' numeric control risk assessment (70.5) for the valuation assertion is significantly higher [$p=0.000$] than that for the linguistic mode (65.3).

In addition to the main effect research questions tested above, Table 8 shows a number of significant higher order interactions between the two response modes (qualitative and quantitative), the two levels of the control environment (strong and weak), account type (subjective and objective), and the five management assertions. The results from Table 8 indicate significant main effects for the two control environment conditions [$p < .0001$]; the two response modes [$p < .008$]; the two account types [$p < .024$]; and the five management assertions [$p < .0001$]. More interestingly, there are a number of significant response mode interaction effects. There are two-way interaction effects regarding response mode and assertions [$p < .001$], and response mode and account type [$p < .056$]. In addition there are several higher order interactions: assertion x account type x response mode

Table 6

RQ_{4A}: For more objective type of accounts, auditors' numeric control risk assessments for the valuation assertion will *differ* from their linguistic control risk assessments.

Response Mode	Mean	F	Sig of F
Linguistic	37.35	48.20	.000
Numeric	33.05		

Table 7

RQ_{4B}: For more subjective type of accounts, auditors' numeric control risk assessments for the valuation assertion will be lower than their linguistic control risk assessments.

Response Mode	Mean	F	Sig of F
Linguistic	65.3	153.43	.000
Numeric	70.5		

TABLE 8
Interaction Effects of Numeric Assessments vs. Linguistic Mid-Point Assessments

Tests involving 'ASSERTION' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
ASSERTION	235614.58	4	58903.64	144.36	.000
WITHIN+RESIDUAL	63653.80	156	408.04		

Tests involving 'ACCOUNT TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
TYPE	286.88	1	286.88	5.55	.024
WITHIN+RESIDUAL	2016.01	39	51.69		

Tests involving 'CONTROL ENVIRONMENT' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
CE	14247.41	1	14247.41	63.77	.000
WITHIN+RESIDUAL	8713.11	39	223.41		

Tests involving 'RESPONSE MODE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
MODE	2035.14	1	2035.14	7.80	.008
WITHIN+RESIDUAL	10172.71	39	260.84		

Tests involving 'ASSERTION BY ACCOUNT TYPE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
ASRT BY TYPE	140136.95	4	35034.24	70.51	.000
WITHIN+RESIDUAL	77515.25	156	496.89		

Tests involving 'ASSERTION BY CONTROL ENVIRONMENT' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
ASRT BY CE	55820.19	4	13955.05	76.58	.000
WITHIN+RESIDUAL	28429.39	156	182.24		

Tests involving 'ASSERTION BY RESPONSE MODE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
ASRT BY MODE	1949.41	4	487.35	4.92	.001
WITHIN+RESIDUAL	5466.22	156	99.14		

Tests involving 'ACCOUNT TYPE BY CONTROL ENVIRONMENT' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
TYPE BY CE	518.13	1	518.13	13.18	.001
WITHIN+RESIDUAL	1533.61	39	39.32		

Tests involving 'ACCOUNT TYPE BY RESPONSE MODE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
TYPE BY MODE	80.33	1	80.33	3.88	.056
WITHIN+RESIDUAL	807.99	39	20.72		

Tests involving 'CONTROL ENVIRONMENT BY RESPONSE MODE' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F
CE BY MODE	564.66	1	564.66	2.21	.145
WITHIN+RESIDUAL	9958.09	39	255.34		

Tests involving 'ASSERTION BY ACCOUNT TYPE BY CONTROL ENVIRONMENT' W-S Effect.

Source of Variation	SS	DF	MS	F	Sig of F
ASRT BY TYPE BY CE	135510.36	4	33877.59	118.38	.000
WITHIN+RESIDUAL	44644.99	156	286.19		

Tests involving 'ASSERTION BY ACCOUNT TYPE BY RESPONSE MODE' W-S Effect.

Source of Variation	SS	DF	MS	F	Sig of F
ASRT BY TYPE BY MODE	9026.84	4	2256.71	10.57	.000
WITHIN+RESIDUAL	33317.06	156	213.57		

Tests involving 'ASSERTION BY CONTROL ENVIRONMENT BY RESPONSE MODE' W-S Effect.

Source of Variation	SS	DF	MS	F	Sig of F
ASRT BY CE BY MODE	11579.22	4	2894.81	16.02	.000
WITHIN+RESIDUAL	28183.00	156	180.66		

Tests involving 'ACCT TYPE BY CONTROL ENVIRONMENT BY RESPONSE MODE' W-S Effect.

Source of Variation	SS	DF	MS	F	Sig of F
TYPE BY CE BY MODE	200.58	1	200.58	15.37	.000
WITHIN+RESIDUAL	508.89	39	13.05		

Tests involving 'ASSERTION BY ACCT TYPE BY CE BY RESPONSE MODE' W-S Effect.

Source of Variation	SS	DF	MS	F	Sig of F
ASRT BY TYPE BY CE					
BY MODE	11442.55	4	2860.64	15.83	.000
WITHIN+RESIDUAL	28190.70	156	180.71		

Source of Variation	SS	DF	MS	F	Sig of F
SUBJECTS	3714148.02	1	3714148.0	395.71	.000
WITHIN+RESIDUAL	366052.70	39	9385.97		

[$p < .0001$]; assertion x control environment x response mode [$p < .0001$]; account type x control environment x response mode [$p < .0001$]; and assertion x account type x control environment x response mode [$p < .0001$].

Discussion

The above results yield some interesting insights. The results do suggest that given equivalent client scenarios in the financial institutions industry, linguistic control risk assessments are significantly higher than their numeric assessment counterparts. This is in keeping with prior studies such as Reimers, et al. (1993) and Budescu, et al. (1988) that have suggested that especially in loss (high risk) domains linguistic responses tend to be more extreme than numeric

responses. Lower assessments of risk translate into less audit work because higher assessed risks translate into larger sample sizes and additional audit procedures. This suggests that faced with a risky industry, qualitative control risk assessments tend to be more conservative (higher) than their numeric counterparts.

Our results further suggest that given a strong control environment, auditors' linguistic control risk assessments tend to be higher than their numeric counterparts. On the other hand though, given a weak control environment, auditors' linguistic control risk assessments tend to be lower than their numeric counterparts. These somewhat counter-intuitive results suggest that qualitative control risk assessments tend to be more conservative (higher) when faced with a

strong control environment and tend to be less conservative (lower) when faced with a weak control environment. These results suggest congruence with prior studies such as Reimers, et al. (1993) and others such as Beyth-Marom 1982, Chesley 1979, Chesley 1985, Hamm 1991, Lichtenstein and Newman 1967, Nakao and Axelrod 1983, and Reimers 1992, that suggest that words tend to be ambiguous in communicating uncertainty. The results further suggest that the numeric assessment technique indeed allows for greater precision in the risk assessment process, allowing auditors to reduce the vagueness of a given linguistic assessment category and instead focusing on a single point within the restricted interval of probabilities deemed representative of that category. In other words, faced with a strong control environment, the qualitative bias seems to continue, thereby prejudicing the qualitative risk assessment higher than its numeric counterpart. On the other hand, faced with a weak control environment, the qualitative bias seems to counter-intuitively fade, and the numeric assessment is actually higher (more conservative) than its qualitative counterpart. This provides the numeric control risk assessment greater intuitive appeal, given that it is more conservative (than the qualitative mode) when faced with a weak control environment and less conservative when faced with a strong control environment.

The insights from the third set of research questions are somewhat murky. One may recall that we had posited that because of the uncertainties and subjectivity associated with management estimates and valuations, the risks associated with these types of accounts are often much greater than for accounts which can be objectively verified. We find from tables 4 and 5 that irrespective of the nature of accounts, the numeric assessment mode yields more conservative (higher) control risk assessments than the qualitative mode. This runs counter to our expectation of the qualitative bias when faced with a more risky (subjective type of accounts) domain. It also runs counter to any expectation we

may have formed following the results to RQ2A and RQ2B that the numeric assessment technique would be more discriminatory and precise when faced with the two different types of accounts.

Tables 6 and 7, though, seem to turn the tide again in favor of greater precision from numeric risk assessments. We find that when faced with more objective type of accounts, and forming risk assessments for the valuation assertion alone, auditors' linguistic risk assessments are more conservative (higher) than their numeric counterparts. When faced with more subjective type of accounts, and again for the valuation assertion alone, the numeric assessment mode seems to be more conservative than its qualitative counterpart. Given our earlier assertion that the valuation assertion is often the most difficult to assess of the five management assertions, and given that linguistic responses tend to be more extreme (biased) than numerical responses in high-risk domains, we had expected auditors' qualitative control risk assessment for the valuation assertion for subjective accounts to be higher than the numerical assessments. We do not find this to be the case and the numeric assessment mode seems to be more conservative, appropriate, and precise than its qualitative counterpart. For the more objective accounts, however, it was difficult for us to generate any *a priori* directional expectations for the valuation assertion. Table 7 suggests that the qualitative bias seems to have had an effect here and the qualitative mode is actually more conservative than its numeric counterpart. As in the case of RQ2A and RQ2B, the numeric control risk assessment seems to hold greater intuitive appeal, given that it is more conservative (than the qualitative mode) when faced with more subjective type of accounts and less conservative when faced with more objective type of accounts.

While the previous discussion pertains to the main effect research questions posited in this study, Table 8 demonstrates the numerous higher order interactions between control environment, account type, and management asser-

tions. These results suggest that in developing control risk assessment decision aids, we cannot ignore the complex interactions between the control environment, the nature (type) of accounts involved, and the management assertions that pertain to a set of audit scenarios.

Implications for Control Risk Assessment Decision Aids

The above results lend support to the contention that the audit environment continues to become more complex. As suggested above, risk identification and risk analysis have a central role in this complex process (Graham, et al. 1991). Risk analysis is usually summarized in strategy documentation and provides a link to the specific testing plan for an engagement. Furthermore, in circumstances where less experienced staff formulate the audit strategy or where the engagement team has less experience with the client, the need for a systematic approach and documentation of audit risk and its components is even greater. And, while professional standards identify some indicators for the auditor to examine, there exist few documented generic rules or proven structured approaches for such assessment of risk (Graham, et al. 1991).

The generous use of judgment for control risk assessment (as one of the main components of audit risk), therefore, places a premium on experience and client knowledge. However, the transfer of knowledge between the partner and manager on the engagement and less experienced members of the audit team is difficult. In this situation, a decision aid specifically designed to aid in control risk assessment becomes paramount. To that end, Martinov and Roebuck (1998) have asserted that the use and relative importance of specific factors and the combination of these factors into an overall risk assessment (over the financial statement, account, and assertion levels) and its ultimate impact on the audit process needs to be better understood. They consider this important given the movement towards the development of risk models in

order to assist in arriving at the most efficient and effective audit strategy. Martinov and Roebuck (1998) further assert that a related domain also worthwhile of exploration is the level of necessary knowledge required to be able to proficiently make risk (and materiality) assessments and integrate them. This would include exploring the type of training necessary to gain this level of knowledge and the type of audit decision aids most useful in this decision process (Martinov and Roebuck 1998).

We believe that the current study addresses these issues in two ways. First, we have demonstrated that auditors make significantly different risk assessments, given the same audit context, simply as a function of the assessment metrics themselves. Differences in risk assessments can have differential impacts on the amounts and types of evidence that are collected. For example, under Probability Proportional to Size sampling (PPS), if the overall audit risk is kept at .05, inherent Risk at 1.0, and analytical procedures risk at .60, the sample size at a control risk of 70 percent (the low end of a 70-100 percent range) is 59. If this were to be contrasted against a linguistic assessment of "high," where "high" equals a 100 percent, PPS would yield a required sample size of 69. This 14 percent decrease in sample size from the qualitative assessment would come about simply as a result of using two different control risk assessment-scaling techniques. This would leave us with the question: "which is the better technique?" Unfortunately, the above results do not suggest that one of the two risk assessment methods is consistently superior to the other used in this study. And yet, any risk assessment decision aid development would have to be contingent on the appropriate answer to the preceding question.

Secondly, the suggestion that the suitability of a control risk assessment technique is to a large extent driven by the nature of the control environment, account type, and specific assertion being evaluated underscores the possible legal implications for any control risk assessment

decision aid. For example, plaintiffs may make the case that a "more conservative" risk assessment technique should have been used in a suit claiming negligence because an inappropriate "lesser" amount of audit evidence was collected. Buckless and Peace (1993) have suggested that jurors' attitudes can be swayed when they perceive that a standard of performance is self-serving. A shrewd lawyer may make the case that auditors took the "easy" way out by selecting the assessment technique that resulted in the least amount of acceptable work. Thus, while the utilized assessment technique may well be acceptable to auditors, it may still fail in the court of law and public opinion.

Conclusions

The above discussion suggests that any attempt to develop control risk assessment decision aids must account for the inherent difficulty and complexity of the domain. However, Blattberg and Hoch (1990) found that expert decision aid use even in situations where the decision aid does not impound all the information available to a decision-maker was beneficial to experts. This augurs well for situations where control risk assessments are made through a combination of decision aid and human user. While auditors would have to exercise caution so as to not blindly rely on the decision aid alone, it is also feasible that over time the decision aid is augmented with the auditors' superior insight into the complex factors that affect auditors' control risk assessments.

At the end we must also discuss the limiting caveats to this study. These include the fact that the accounts used in the study are mostly representative of the financial institutions industry. Furthermore, the description of the accounting system and control procedures as being "in place and operating effectively" may have appeared somewhat simplistic and artificial. However, deliberate artificiality is often necessary in experiments in order to test theories more directly (Marden, et al. 1997; Swieringa and

Weick 1982). Although the CE should be initially considered, and only then should the accounting system and control policies be evaluated, there is also a reverse-relationship in that having the accounting system and control procedures in place reflects something about the control environment itself. Experimental manipulation of the conditions of the accounting system and control procedures may have provided additional insights. However, post-experimental interviews with subjects indicated that the manner and amount of information provided on the CE, accounting system, and control features was reasonable for the task.

Suggestions for Future Research

The basic question of whether control risk assessment decisions aids should be linguistic or numeric cannot be answered until we find the cause for the lack of consistency in control risk evaluation methods. Decision aids developed without adequate consideration of this issue run legal risks and the possibility of misguiding novice auditors. Hence, it becomes imperative that we address the issue of scale differentials in control risk assessments prior to the development or release of audit decision aids in this area. Continued research on these risk assessment-scaling techniques is, therefore, of practical importance for academics and audit firms alike. ☐

Endnotes

1. SAS No. 78 was an Exposure Draft at the time that this study was conducted. It was officially released shortly after the experiment was conducted.
2. A total of 12 different financial institution accounts were preliminarily selected as those typically found in financial institutions (AICPA 1991: Audit Guide to Audits of Savings Institutions; GAAS Guide, 1996). The pilot subjects were asked to rate the extent to which the twelve accounts were considered to be more objective or more subjective in na-

- ture on a scale of "1" to "7", where "1" = more objective and "7" = more subjective. Ratings for each account were made at the individual assertion level. The subjects were all from Big Six accounting firms (nine seniors and eight managers) and all had extensive financial institution experience.
3. The original data collection also included assessments of inherent risk. A separate analysis of the inherent risk assessments indicated that they were virtually identical to the results of the control risk assessments. This may have been due in part to the fact that the CE also shares many of the same factors that the auditor considers in assessing inherent risk (e.g., Waller 1993; Dirsmith and Haskins 1991; Kinney 1989). Given the complexity of analyzing two additional dependent variables along with the two control environment conditions across four accounts and five assertions, only the results of the control risk assessments are discussed in this paper.
 4. There were no significant violations found during the testing of the statistical assumptions.
- ### References
1. AICPA, American Institute of Certified Public Accountants. 1978. Statement On Auditing Standards No. 22: *Planning and Supervision*. New York: AICPA.
 2. AICPA, American Institute of Certified Public Accountants. 1980. Statement On Auditing Standards No. 31: *Evidential Matter*. New York: AICPA.
 3. AICPA, American Institute of Certified Public Accountants. 1981. Statement On Auditing Standards No. 39: *Audit Sampling*. New York: AICPA.
 4. AICPA, American Institute of Certified Public Accountants. 1983. Statement On Auditing Standards No. 47: *Audit Risk and Materiality in Conducting an Audit*. New York: AICPA.
 5. AICPA, American Institute of Certified Public Accountants. 1988. Statement On Auditing Standards No. 55: *Consideration of the Internal Control Structure in a Financial Statement Audit*. New York: AICPA.
 6. AICPA, American Institute of Certified Public Accountants. 1988. Statement On Auditing Standards No. 57: *Auditing Accounting Estimates*. New York: AICPA.
 7. AICPA, American Institute of Certified Public Accountants. 1996. Statement On Auditing Standards No. 78: *Consideration of the Internal Control Structure in a Financial Statement Audit: An Amendment To SAS No. 55*. New York: AICPA.
 8. AICPA, American Institute of Certified Public Accountants. 1991. Audit and Accounting Guide: *Audits of Savings Institutions*. New York: AICPA.
 9. Amer, T., K. Hackenbrack, and M. Nelson. 1994. Between-Auditor Differences in the Interpretation of Probability Phrases. *Auditing: A Journal of Practice & Theory* (Spring): 126-136.
 10. Arens, A.A. and J.K. Loebbecke. 1991. *Auditing: An Integrated Approach* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.
 11. Argenti, J. 1976. Corporate Collapse: The Causes and Symptoms. New York: John Wiley.
 12. Argenti, J. 1980. *Practical Corporate Planning*. London: George Allen & Unwin Ltd.
 13. Artz, N. and A.M. Tybout. 1990. Numeric and Verbal Information: Implications for Persuasion. Working Paper, Northwestern University.
 14. Ashton, R.H. and J.J. Willingham. 1989. Using and Evaluating Audit Decision Aids. *Proceedings: Touche Ross/University of Kansas Symposium on Auditing Problems*: 1-25.
 15. Bell, J. 1984. The Effect of Presentation Form on the Use of Information in Annual Reports. *Management Science* 30 (2): 169-185.

16. Beyth-Marom, R. 1982. How Probable is Probable? A Numerical Translation of Verbal Probability Expressions. *Journal of Forecasting* (1): 257-269.
17. Birnberg, J.G. 1964. Bayesian Statistics: A Review. *Journal of Accounting Research* 2 (1): 108-116.
18. Blattberg, R.C. and S.J. Hoch. 1990. Database Models and Managerial Intuition: 50% Model + 50% Manager. *Management Science* (August): 887-899.
19. Boritz, J.E. 1985. The Effect of Information Presentation Structures on Audit Planning and Review Judgments. *Contemporary Accounting Research*, 1 (2): 193-218.
20. Boritz, J.E., B.G. Gaber, and W.M. Lemon. 1987a. An Experimental Study of Review of Preliminary Audit Strategy by External Auditors. The Canadian Academic Accounting Association: Toronto.
21. Boritz, J.E., B.G. Gaber, and W.M. Lemon. 1987b. Managing Audit Risk. *CA Magazine* 120 (1): 36-41.
22. Bowsher, C. 1990. Resolving The Savings and Loan Crisis: Billions More and Additional Reforms Needed. GAO/T-AFMD-90-15.
23. Braun, G.P. 1998. Auditors' Assertion-Level Inherent Risk Assessments: A Descriptive Analysis. Working paper, The University of Texas at El Paso.
24. Buckless, F. and R. Peace. 1993. The Influence of The Source Of Professional Standards on Juror Decision Making. *The Accounting Review* (January): 164-175.
25. Budescu, D.V., S. Weinberg, and T.S. Wallsten. 1988. Decisions Based on Numerically and Verbally Expressed Uncertainties. *Journal of Experimental Psychology: Human Performance and Perception* (14): 281-294.
26. Chesley, G.R. 1979. Procedures for the Communication of Uncertainty in Auditors' Working Papers. In T. Burns (Ed.), *Behavioral Experiments in Accounting II*. Columbus, OH.: College of Administrative Sciences, The Ohio State University.
27. Chesley, G.R. 1985. Interpretation of Uncertainty Expressions. *Contemporary Accounting Research*, 2(1): 179-199.
28. Cushing, B.E., L.E. Graham, Z. Palmrose, R.S. Roussey, and I. Solomon. 1995. Risk Orientation. In *Auditing Practice, Research, and Education*. T.B. Bell and A.M. Wright, eds. American Institute of Certified Public Accountants.
29. Daniel, S. 1988. Some Empirical Evidence About the Assessment of Audit Risk in Practice. *Auditing: A Journal of Practice & Theory* (Spring): 174-181.
30. Day, K. 1998. Byline. The Washington Post Financial Section, Sunday Final Edition (December 13): H 01.
31. Dirsmith, M. and M. Haskins. 1991. Inherent Risk Assessments and Audit Firm Technology: A Contrast in World Theories. *Accounting, Organizations and Society* (Spring): 61-90.
32. Dilla, W. and D. Stone. 1992. Effects of Response Representation and Experience on Risk Judgments. Working Paper, University of Illinois.
33. Dilla, W.N. and D.N. Stone. 1997. Representations as Decision Aids: The Asymmetric Effects of Words and Numbers on Auditors' Inherent Risk Judgments. *Decision Sciences* (Summer): 709-743.
34. Dusenbury, R., J. Reimers, and S. Wheeler. 1996. An Empirical Study of Belief-Based and Probability-Based Specifications of Audit Risk. *Auditing: A Journal of Practice & Theory* (Fall): 12-28.
35. Einhorn, H. and R. Hogarth. 1981. Behavioral Decision Theory: Processes of Judgment and Choice. *Annual Review of Psychology* 32: 53-88.
36. Eining, M., D. Jones, and J. Loebbecke. 1994. An Experimental Examination of the Impact of Decision Aids on the Assessment and Evaluation of Management Fraud. Working Paper, University of Utah, Presented at the USC Audit Judg-

- ment Symposium.
37. Emby, C. 1993. An Experimental Test of a Decision Aid for Sequential Belief Revision Under Uncertainty. Working Paper, Simon Fraser University.
38. Fillenbaum, S., T.S. Wallsten, B.L. Cohen, and J.A. Cox. 1991. Some Effects of Vocabulary and Communication Task on the Understanding and Use of Vague Probability Expressions. *American Journal of Psychology* 104 (1): 35-60.
39. GAAS Guide. (1996). *HBJ Miller Comprehensive GAAS Guide: A Comprehensive Restatement of Generally Accepted Auditing Standards*. L.P. Bailey and M.A. Miller, eds. New York: Harcourt Brace Janovich.
40. GAO, General Accounting Office. (1989). CPA Audit Quality: Report to the Chairman, Committee on Banking, Finance and Urban Affairs, House of Representatives. U.S. GAO Accounting and Financial Management Division.
41. Graham, L.E. 1985. Audit Risk-Parts I-V. *The CPA Journal* 55: 8-11.
42. Graham, L., J. Damens, and G. Van Ness. 1991. Developing Risk Advisor: An Expert System for Risk Identification. *Auditing: A Journal of Practice & Theory* 10 (Spring): 69-96.
43. Hamm, R. 1991. Selection of Verbal Probabilities: A solution for Some Problems of Verbal Probability Expression. *Organizational Behavior and Human Decision Processes* 50 (2): 193-223.
44. Harsha, P. and M. Knapp. 1990. The Use of Within- and Between-Subjects Experimental Designs in Behavioral Accounting Research: A Methodological Note. *Behavioral Research in Accounting* (Vol. 2): 50-62.
45. Haskins, M.E. and M.W. Dirsmith. 1993. Control and Inherent Risk Assessments in Client Engagements: An Examination of Their Interdependencies. *Journal of Accounting and Public Policy* (14): 63-83.
46. Houghton, C.W. and J.A. Fogarty. 1991. Inherent Risk. *Auditing: A Journal of Practice & Theory* (Spring): 1-21.
47. Janell, P. and A. Wright. 1991. Inherent Risk and Control Environment Assessment Practices of Major Auditing Firms: Implications for Practice and Future Research. Working Paper, Northeastern University.
48. Jiambalvo, J. and W. Waller. 1984. Decomposition And Assessments of Audit Risk. *Auditing: A Journal of Practice & Theory* (Spring): 80-88.
49. Kreutzfeldt, R.W. and W.A. Wallace. 1990. Reply and Extension to Discussion of Control Risk Assessments: Do They Relate to Errors? *Auditing: A Journal of Practice and Theory (Supplement)*: 33-48.
50. Keppel, G. 1992. *Design and Analysis: A Researcher's Handbook*. Prentice Hall, Englewood Cliffs, NJ.
51. Kinney, W. 1989. Achieved Audit Risk and the Outcome Space. *Auditing: A Journal of Practice & Theory (Supplement)*: 67-84.
52. Kleinmuntz, D.N. and D.A. Schkade. 1993. Information Displays and Decision Processes. *Psychological Science* 4 (4): 221-227.
53. Libby, R. and P.A. Libby. 1989. Expert Measurement and Mechanical Combination in Control Reliance Decisions. *The Accounting Review* (October): 729-747.
54. Lichtenstein, S. and J.R. Newman. 1967. Empirical Scaling of Common Verbal Phrases Associated with Numerical Probabilities. *Psychonomic Science*, 9 (10): 563-564.
55. Marden, R.E., G.L. Holstrum, and S.L. Schneider. 1997. Control Environment Condition and the Interaction Between Control Risk, Account Type, And Management's Assertions. *Auditing: A Journal of Practice & Theory* (Spring): 51-68.
56. Martinov, N. and P. Roebuck. 1988. The Assessment and Integration of Materiality and Inherent Risk: An Analysis of Major Firms' Audit Practices. Working Paper.

57. Mautz, R.K. and H.A. Sharaf. 1961. *The Philosophy of Auditing*. American Accounting Association: Monograph No. 6. Sarasota, FL.
58. McFadgen, D. N. 1994. Discussion of Stone and Dilla's "When Numbers are Better than Words: The Joint Effects of Response Representation and Experience on Inherent Risk Judgments." *Auditing: A Journal of Practice & Theory* (Supplement): 20-23.
59. Messier, W.F., Jr. 1995. Research in and Development of Audit Decision Aids. In R.H. Ashton and A.H. Ashton (Eds.), *Judgment and Decision-Making Research in Accounting and Auditing*. New York: Cambridge University Press.
60. Mock T. and A. Wright. 1993. An Exploratory Study of Auditors' Evidential Planning Judgments. *Auditing: A Journal of Practice & Theory* (Fall): 39-61.
61. Nakao, M.A. and S. Axelrod. 1983. Numbers are Better Than Words: Verbal Specifications of Frequency Have No Place in Medicine. *American Journal of Medicine* (74): 1061-1065.
62. Pany, K. and P. Reckers. 1987. Within- vs. Between-Subjects Experimental Designs: A Study of Demand Effects. *Auditing: A Journal of Practice & Theory* (Fall): 39-53.
63. Payne, J.W., J.R. Bettman, and E.J. Johnson. 1992. Behavioral Decision Research: A Constructive Processing Perspective. *Annual Review of Psychology* (43): 87-131.
64. Reimers, J. 1992. Additional Evidence on the Need for Disclosure Reform. *Accounting Horizons* 6 (1): 36-41.
65. Reimers, J., S. Wheeler, and R. Dusenbury. 1993. The Effect of Response Mode on Auditors' Control Risk Assessments. *Auditing: A Journal of Practice & Theory* Fall: 62-78.
66. Rittenberg, L.E. and B.J. Schwieger. 1994. *Auditing: Concepts for a Changing Environment*. Fort Worth, TX: Dryden.
67. Schkade, D. and D. Kleinmuntz. 1994. Information Displays and Choice Processes: Differential Effects of Organization, Form, and Sequence. *Organizational Behavior and Human Decision Processes* 57 (3): 319-337.
68. 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.
69. Stone, D.N. and D.A. Schkade. 1991. Numeric and Linguistic Information Representation in Multiattribute Choice. *Organizational Behavior and Human Decision Processes* 49 (1): 42-59.
70. Svenson, O. and G. Karlsson. 1986. Attractiveness of Decision Alternatives Characterized by Numerical and Nonnumerical Information. *Scandinavian Journal of Psychology* (27): 74-84.
71. Swieringa, R. and K. Weick. 1982. An Assessment of Laboratory Experiments in Accounting. *Journal of Accounting Research* (Supplement): 56-101.
72. Taylor, M.H. 1997. Bounded Rationality, Uncertainty, and Competence: The Effects of Industry Specialization on Auditors' Inherent Risk Assessments. Working paper, University of Nebraska, Lincoln, NE.
73. Wagenaar, W.A., 1969. A Note on the Construction of Digram-Balanced Latin Squares. *Psychological Bulletin* (No. 72): 384-386.
74. Waller, W.S. 1991. Auditors' Risk Assessments: Effects of Task Experience, Conditioning Information, and Precision on Second-Order Uncertainty. Working Paper, University of Arizona.
75. Waller, W. 1993. Auditors' Assessments of Inherent and Control Risk in Field Settings. *The Accounting Review*. (October): 783-803.
76. Wallsten, T.S., D.V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth. 1986.

Measuring the Vague Meanings of Probability Terms. *Journal of Experimental*

Psychology: General 115 (4): 348-365.

Notes

Notes