

STX

330
B385

COPY 2

91:113

Information Representation, Scaling, and Experience in Inherent Risk Judgements

The Library of the

APR 1 1991

University of Illinois
of Urbana-Champaign

William N. Dilla
Department of Accountancy

Dan N. Stone
Department of Accountancy



Bureau of Economic and Business Research
College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 91-0113

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

February 1991

Information Representation, Scaling, and Experience
in Inherent Risk Judgments

William N. Dilla*
Dan N. Stone

Department of Accountancy
University of Illinois at Urbana-Champaign

Sincere thanks to the University of Illinois Office for Information Management for financial support and programming assistance on the project. In addition, the financial assistance of the University of Illinois Investors in Business Education and Campus Research Board are gratefully acknowledged.

*Names are listed alphabetically. The authors contributed equally to the project.

Information Representation, Scaling, and Experience in Inherent Risk Judgments

ABSTRACT

Do audit decision aids such as standardized response scales and numeric measures improve audit risk assessments? Does the value of these aids depend upon the extent of auditor experience? This paper reports the results of two experiments exploring these issues.

In the first experiment, 42 practicing auditors and 55 auditing students assessed inherent risk using pre-established, standardized scales with either numeric or linguistic representation labels (e.g., "very high" vs. "1.0" inherent risk). Decision cues were also manipulated between numeric and linguistic representations. Students had significantly lower judgment deviation (i.e., higher consensus) with linguistic cue representation than with numeric. In addition, manipulating cue representation led to changes in relative cue weighting for both students and auditors.

The second experiment explored four issues: (1) auditors' versus students' initial risk anchors, (2) implications of standardized versus individual scaling for risk judgments, (3) the effect of 'mixed' (i.e., some numeric/some linguistic) cue information on risk judgments, and (4) the impact of cue and response representation manipulations on participants' perceived and actual cognitive effort. In Experiment 2, 60 practicing auditors and 64 auditing students created individual assessment scales by using custom-developed software to state equivalencies between numeric and linguistic risk representations. Cue representation and response representation were again manipulated using numeric and linguistic formats.

In Experiment 2, students again had lower judgment deviation with linguistic cue representation. Students' cue weightings were also dependent on cue representation. In contrast, auditors' weightings were consistent across manipulations of cue representation. However, auditors' decision processes were affected by cue representation -- auditors reexamined cue information more with numeric cue representation, but took less time per examination. All participants took significantly longer to make risk assessments using numeric response representations relative to linguistic ones. 'Mixed' cue conditions did not lead to significant increases in cognitive effort or decreases in judgment accuracy relative to pure numeric and pure linguistic conditions.

In both experiments, students' assessments of inherent risk were higher than auditors', regardless of experimental condition. Data from Experiment 2 suggest that this is because auditors use lower initial anchors for risk judgments. The use of individual scales in Experiment 2 appears to have resulted in lower inherent risk assessments for both auditors and students, and increased judgment deviation among both participant groups relative to the standardized scales used in Experiment 1. The paper concludes by suggesting that audit decision aids are not unequivocally beneficial and that the efficacy of such aids may depend upon the prior training and experience of the auditor.

MEMORANDUM

TO : [Illegible]

FROM : [Illegible]

SUBJECT : [Illegible]

[The following text is extremely faint and illegible due to the quality of the scan. It appears to be a multi-paragraph memorandum.]

Introduction

Audit firms invest considerable resources in developing standards, guidelines, and policies intended to aid and structure audit judgments (Boritz, 1985). Implicit in this approach is a presumption that unaided judgments are inferior to aided ones, and that providing structure increases consistency across audit situations and among auditors. However, little empirical evidence exists on the impact of decision aiding techniques on audit judgments and decision processes (Boritz, 1985; Libby and Libby, 1989).

Assessing audit risk is a task in which decision aids have been implemented in recent years in an attempt to improve subjective judgments. Boritz, Gaber, and Lemon (1987a, 1987b) argue that quantitative approaches to risk assessment may provide extensive benefits to an audit firm relative to qualitative approaches. These benefits include more defensible logic, improved training for staff members, easier review processes, and reduced errors due to the judgmental combination of data. Indeed, some accounting firms have moved towards standardized, quantitative approaches for risk assessment (e.g., Elliott, 1983). However, many other firms have retained qualitative approaches, despite contrary recommendations in the academic auditing literature. The intent of this paper is to provide empirical evidence on the value of quantitative versus qualitative risk assessment approaches.

The suggestion that auditors use numeric (i.e., quantitative) as opposed to linguistic (i.e., qualitative) information representation for expressing risk judgments is supported in part by Chesley's (1979, 1985) studies of accounting undergraduate students' usage and interpretation of linguistic expressions of uncertainty. He found wide variety in interpretations of linguistic expressions and therefore recommended:

"Words in large measure ambiguously communicate uncertainty. Until further study can find a reason for their use, it is suggested, based on this research, that a number scale for probability communication be adopted." (1985, p. 197)

However, other researchers have expressed skepticism regarding recommendations to move towards the exclusive use of numeric expressions of uncertainty. For example, Zimmer (1983, 1984) argues that linguistic expressions of uncertainty (but not numeric expressions) contain useful information about the precision of uncertainty expressions. Recent work by Wallsten and his colleagues (Erev and Cohen, 1990; Erev, Gonzalez, and Wallsten, 1990) provides evidence suggesting that decisions are improved when decision makers use their preferred information representation. Accordingly, it is not obvious that the use of numerical representations of uncertainty will improve audit risk judgments.

Another important issue with respect to audit risk assessment is the information representation of cues used to make these judgments. An extensive body of research suggests that the same information presented in different representations results in different decisions and decision processes (for reviews, see Bettman, 1979; Payne, 1982; Kleinmuntz and Schkade, 1990). For example, Stone and Schkade (in press) asked decision makers to choose accounting software using either numeric or linguistic representations of information describing the software. They found that decision makers chose more quickly and made more comparisons among available information when using numeric representation. When applied to auditing, such results suggest that presenting audit client information in numbers may lead to differences in decision processes and judgments compared to "equivalent" information represented in words.

Previous research suggests that the impact of decision aids on audit judgment may depend upon the prior training and experience of auditors. For example, Boritz (1985) found that structuring information presentations caused

considerably different effects depending upon the prior experience and firm position of the auditor. Similarly, Waller (1990, p. 5) argues that "A possible consequence of diversity of experience is that auditors may vary in how they interpret verbal risk descriptors (e.g., low-moderate-high), which would impair consensus." Accordingly, the same information representation may lead to different effects depending upon the training and experience of the auditor.

The remainder of this paper describes two experiments intended to provide evidence relevant to the issues of information representation, scaling, and experience in inherent risk decisions. The first uses a paper and pencil type experimental instrument and examines inherent risk judgments using standardized, pre-established risk scales. The second uses computer-assisted data collection and examines inherent risk judgments using individual, participant-established risk scales. Both manipulate the cue and response representations used by decision makers.

A Theoretical Analysis of Inherent Risk Judgments

Inherent Risk in Auditing

Inherent risk is the probability that material error has occurred in an account balance or class of transactions, before considering the effectiveness of internal accounting controls (AICPA, Statement on Auditing Standards (SAS) 47). It, along with control risk and detection risk, jointly define audit risk, or the probability that, unknown to the auditor, material error exists in the financial

statements after audit procedures are complete.¹ Inherent risk factors may affect the probability of misstatement in the financial statements in general, or may only affect specific accounts or classes of transactions.

Assessing inherent risk can be a powerful tool for increasing audit efficiency, since if the auditor documents an inherent risk level of less than 1.0, the extent of detection procedures can be reduced (Leichti, 1986, Alderman and Tabor, 1989). As a result, inherent risk assessment has become an integral part of many large accounting firms' audit practice. As mentioned in the introduction, auditors typically make either quantitative or qualitative inherent risk assessments (Boritz, Gaber, and Lemon, 1987b). Auditors using the quantitative approach gather risk-relevant information, make a qualitative judgment of inherent risk (e.g., 'high', 'medium', or 'low'), and convert this judgment to a quantitative equivalent using a firm-specified scale (i.e., 'low' = 0.5, 'medium' = 0.7, etc.). The numeric assessment is then used to assist in determining statistical sample sizes for substantive tests (e.g., Elliott, 1983). With a qualitative approach, auditors use risk information to make linguistic assessments that are used as judgmental guides in developing audit programs. A recent survey of Canadian firm practices suggests that about 33% of firms use quantitative assessments, while 77% use qualitative inherent risk assessment (Boritz, Gaber, and Lemon, 1987b).²

¹Some auditing researchers have argued that inherent risk and control risk are interdependent, and that it is therefore infeasible to separately assess audit risk components (Cushing and Loebbecke, 1983; Kinney, 1984; Waller, 1990). Some audit firms do follow the approach of making combined estimates of inherent and control risk. However, inherent risk remains in the audit risk formula currently in the Statements on Auditing Standards and many audit firms continue to make separate inherent and control risk assessments.

²The percentage totals do not equal 100% since some firms use both qualitative and quantitative methods.

For the most part, research on inherent risk has focused on documenting current audit practices and exploring relationships among audit judgments, environmental cues, and audit errors. Some of this research has used archival data to examine the ex post relationship between client characteristics, environmental factors, and audit errors (e.g., Willingham and Wright, 1984; Kreutzfeldt and Wallace, 1986; Johnson, 1987). Studies of individual auditor judgments (Boritz, Gaber, and Lemon, 1987a; Colbert, 1988; Daniel, 1988) have built upon this work by examining the extent to which such cues are actually used in auditors' inherent risk assessments. In addition, Peters (1989) has developed a descriptive expert systems model of individual auditors' combined inherent and control risk assessment processes.

Extant research on inherent risk has been clearly useful in understanding existing audit practices. A logical extension, however, is to explore the impact of alternative decision aiding approaches on inherent risk judgments. The following sections discuss cognitive strategies for audit risk assessment and the potential effects of alternative cue and response representations on these strategies.

Risk Assessment Strategies

Ashton, Kleinmuntz, Sullivan, and Tomassini (1988, p. 119-120) have recently called for the application of a cognitive cost-benefit framework to issues of auditor decision behavior. Using the most common formulation of such a framework, decision makers are assumed to choose decision strategies primarily on the basis of trade-offs between the anticipated cognitive effort and the anticipated accuracy of various strategies (Payne, 1982; Johnson and Payne, 1985; Kleinmuntz and Schkade, 1990). To illustrate the application of error and effort theory to inherent risk assessment, consider three possible inherent risk judgment strategies, as follows.

Assuming a true but unknown inherent risk for each audit client, a low effort, low accuracy strategy is to always set inherent risk at 1.0. Such an approach requires no cognitive effort and is even recommended by some audit researchers (Kinney, 1989). However, most practicing auditors would argue that inherent risk can be set at less than 1.0 in the majority of audits; as a result, assessments made with this strategy one will overstate inherent risk for the majority of clients.

Two alternative strategies employ the anchoring and adjustment heuristic (Boritz 1985, Boritz, Gaber, and Lemon, 1987a). A moderate effort, moderate accuracy strategy is to initially anchor the assessment at 1.0 and adjust it based upon information obtained in a formal inherent risk investigation of the audit client. Such a strategy requires greater cognitive effort and is more likely to provide an assessment closer to the true (but unknown) inherent risk. However, insufficient adjustment may occur from the initial anchor (Kinney and Uecker, 1982; Joyce & Biddle, 1981) of maximum inherent risk, resulting in assessments that are overstated relative to the true inherent risk .

A high effort, high accuracy strategy is to anchor the assessment based upon knowledge of the audit environment and background information of the client, and to adjust based upon information obtained during a formal inherent risk investigation. This strategy requires more cognitive effort than either of the previous two strategies, since knowledge of the audit environment and client are integrated into the assessment. However, such a strategy is the most likely of the three to produce accurate inherent risk judgments.

Information Representation & Experience in Inherent Risk Judgments

How might using different cue and response representations impact inherent risk judgments? Wallsten (1988) argues that the appropriate representation for communicating probabilities depends upon the uncertainty

associated with the information (called secondary uncertainty). For highly certain information (e.g., the probability of getting "heads" when flipping a fair coin), numeric representation is best, since it affords precise statement of an exact probability. However for less certain predictions (e.g., the probability of a job candidate accepting an offer), linguistic representation includes useful information about the uncertainty of the estimate. Accordingly, the representations chosen for information should be only as precise as the information itself.

Information used in making inherent risk judgments is frequently imprecise. If auditors are only able to discriminate a small number of categorical differences (e.g., "high", "medium" or "low") in interpreting cue information and making risk assessments, then numeric representations could exaggerate the implied precision of such estimates (Boritz and Wensley, 1988, p. 80). If auditors can only distinguish a small number of cue and risk categories, then linguistic cues and risk representations could increase accuracy by providing a small number of well-understood categories that contain implicit information about the secondary uncertainty of estimates. In contrast, numeric cues and risk representations could decrease accuracy by providing a larger number of poorly understood categories, and by omitting relevant information about secondary uncertainty.

Research suggests that experience may also affect the usefulness of cue and response representations. Wallsten and colleagues (Erev and Cohen, 1990; Erev, Gonzalez, and Wallsten, 1990) studied dyads engaged in decision making tasks who communicated probability information to one another. In two different tasks, results indicated that decision accuracy was improved when decision makers used their preferred information representation. Evidence from practice suggests that the majority of auditors use linguistic expressions of audit risk

(Boritz, Gaber, and Lemon, 1987b). As a result, greater familiarity with linguistic cue and risk response representations could produce more accurate and less effortful assessments as the result of a more direct mapping between auditors' knowledge structures and linguistically stated information. However, the relative benefits of linguistic cue and response representations may be influenced by the extent of auditors' training and experience. Audit firms increasingly provide training in quantitative risk assessment methodologies. The relative benefits of linguistic risk assessment may therefore lessen as auditors acquire knowledge of and experience with quantitative risk assessment methodologies.

Because there is little empirical evidence that explores the impact of information representation and experience on inherent risk judgments, we conducted two experiments to examine these relationships. The following sections describe the methodology, results, and implications of the experiments.

Experiment 1

Method

Experimental Task

Participants began the task by reading a short description of a hypothetical second year audit engagement of a computer hardware products manufacturer, followed by a definition of inherent risk. Subsequently, they made 17 inherent risk assessments at the overall financial statement level.³ The case materials described four cues potentially relevant in assessing inherent risk:

(1) management incentives (the percentage of management compensation derived from measures related to net income), (2) management's influence on accounting

³Participants were told to assume that aggregate errors of 5% or more of net income before taxes were to be considered material.

policies (the extent to which upper-level management makes year-end changes to accounting estimates), (3) the discovery of material errors in the prior year's audit, and (4) product complexity (the percentage of revenues and cost of goods sold that are determined by subjective estimates). The cues were chosen based on inclusion in SAS 53 and other prescriptive auditing literature as having an impact on the risk of material misstatements at the financial statement level.⁴ Empirical research suggests that all four cues influence the likelihood of material financial statement errors.⁵

The cues were manipulated at two levels in a full factorial design, resulting in 16 (2⁴) unique cases. Participants were given a brief description of each cue, along with both linguistic and numeric descriptions of the high and low cue levels (See Figure 1). For example, a linguistic description for a cue at the low level was: "In some cases, management's compensation has little relationship to net income." The equivalent numeric description immediately followed in parentheses: "(i.e., 5% of management's compensation is from bonus plans / stock options)." The last part of the instructions gave participants equivalencies between linguistic and numeric expressions of inherent risk (See bottom of Figure 1.). These equivalencies were established based upon an examination of the

⁴SAS 53 mentions management influence on accounting (i.e., an aggressive attitude toward financial reporting), discovery of material errors in previous engagements, and contentious or difficult accounting issues as inherent risk factors. While SAS 53 does not explicitly mention management incentives, this factor is often mentioned in other prescriptive auditing literature (e.g., Elliott, 1983; Arens and Loebbecke, 1988; Alderman and Tabor, 1989).

⁵Johnson(1987) found a relationship between the existence of management bonus plans and the size of financial statement errors. Kreutzfeldt and Wallace (1986) have shown that companies with aggressive accounting policies have up to three times as many errors as do other companies. Also, nearly half of the errors in their sample were classified as judgmental evaluation errors or incorrect applications of GAAP. Hylas and Ashton (1982), Willingham and Wright (1984), and Wright and Ashton (1989) all have found a relationship between the discovery of material errors in the prior year's audit and the existence of material misstatements in the current year's unaudited financial statements.

professional (e.g., Leichti, 1986) and academic (e.g., Boritz, Gaber, and Lemon, 1987a) literature on inherent risk assessment.

After completing the instructions, the participants completed a practice case, followed by the 16 experimental cases. Each case appeared on a separate page. Four different case presentation orders were randomly assigned across participants.

Experimental Design

The experimental factors were experience (student or auditor), cue representation, and response representation. These were manipulated in a 2 x 2 x 2 crossed design. Cue and response representations were either linguistic or numeric. Figure 2 illustrates the linguistic and numeric cue representations for the "low" cue levels used in the experiment. The numeric response representation was a continuous scale running from 0.6 to 1.0, while the linguistic response representation was a set of five discrete labels with accompanying boxes (Figure 3).

Participants

Fifty-five students and 42 practicing auditors participated in the study. Students completed the experiment approximately three weeks after introduction of the audit risk model in class. All auditor participants were from 'Big Six' public accounting firms. They had professional experience ranging from two to nine years, with a mean of 3.5 years and a median of three years. Thirty-two auditors from a single firm completed the instrument during a staff training session. Ten participants from three different firms completed the experimental instrument in their offices. All instruments were distributed and collected by a firm representative, who then mailed them to the experimenters.

Dependent Variables

As with many audit tasks, there are no objective standards available for assessing the accuracy of inherent risk judgments (e.g., R. Ashton, 1974, 1982; Libby, 1981). We therefore used two surrogate measures for accuracy: (1) judgment deviation (i.e., consensus) (Ashton, 1985) and (2) the proportion of variance explained by individual participants' linear judgment models (i.e., omega-squared) (Hays, 1981). To compute judgment deviation, we averaged the absolute difference between each participant's response on each case and the mean response on that case for the other individuals in the participant's experimental treatment group.^{6 7} Note that higher deviation scores therefore indicate lower consensus.

To measure explained variance, we first computed omega-squared statistics for each main effect in participants' decision models.⁸ These were summed to compute overall explained variance for main effects.⁹ Two separate analyses of omega-squared statistics were conducted.¹⁰ (1) An ANOVA on the

⁶ The judgments of participants in the the linguistic response condition were converted to numeric values using the numeric/linguistic equivalencies shown in Figure 1.

⁷Our measure is algebraically equivalent to the pairwise absolute consensus measure in A. Ashton (1985). The only difference is that Ashton first computed pairwise consensus scores across cases, then averaged the scores across subjects.

⁸Since the responses are proportions, a variance-stabilizing arcsin transformation (Neter, Wasserman, and Kutner, 1990) was applied before analysis.

⁹Since there were no theoretical reasons to expect significant interactions between the decision cues, the omega-squared values were computed based on a model with main effects only and all interactions included in the error term. Subsequent to our initial analyses, we performed analyses based on individual models incorporating main effects and two-way interactions and having only three- and four-way interactions included in the error term. No substantive differences were noted between the results of these analyses and the results reported in the paper.

¹⁰Omega-squared statistics are proportions, therefore, a variance-stabilizing arcsin transformation (Neter, Wasserman, and Kutner, 1990) was applied before further analysis.

total proportion of variance statistics compared the extent to which the decision cues in total explained judgment variance across experimental groups. (2) A MANOVA with the omega-squared values for individual cues as dependent variables compared relative cue weightings. In addition, we analyzed participants' mean risk assessment responses.

Method of Analysis

In order to perform more powerful tests of experience effects, we analyzed the data using a set of orthogonal planned comparisons, instead of the standard tests of factorial effects (see Keppel, 1982, p. 240; Anderson and Wright, 1988; Buckless and Ravenscroft, 1990). The effects tested were: students vs. auditors (same as the test of a main effect for participants), and cue representation (numeric vs. linguistic), response representation (numeric vs. linguistic), and cue by response representation within students and within auditors.

As a check on participant understanding of the task, we computed individual participant regressions using the inherent risk assessments for each case as dependent variables and the decision cue levels entered as dummy (i.e., 0,1) independent variables. Normatively, the regression weights for each cue should be positively related to risk judgments. Therefore, participants with one or more statistically significant ($p < 0.05$) negative regression weights were dropped from further analyses. Also, we dropped participants whose total variance explained by the four decision cues (total of four individual omega-squared values) was less than 0.25, under the assumption that such participants either did not understand the task or randomly responded to the cases. After these two tests, 95 participants remained in our sample. Table 1 shows the distribution of remaining participants across cells of the experimental design.

Results--Inherent Risk Judgments

Cue and Response Representation

Judgment deviation: The students' mean judgment deviation was significantly higher with numeric cue representation (0.068) than with linguistic representation (0.057) ($F(1,87) = 6.31; p = 0.014$), indicating greater consensus among students with linguistic cues (See Table 2).¹¹ The auditors' mean judgment deviation was also greater with numeric representation than with linguistic (0.055 vs. 0.050), but the difference was not statistically significant ($F(1,87) = 1.14; p = 0.288$). Response representation did not have a significant effect on judgment deviation for either students ($F(1,87) = 0.03; p = 0.855$) or auditors ($F(1,87) = 0.51; p = 0.477$). The cue by response representation interaction for judgment deviation was not significant for either group (students: $F(1,87) = 1.27; p = 0.262$; auditors: $F(1,87) = 0.29; p = 0.592$).

Proportion of variance explained: The mean total proportion of variance explained by students' judgment models was significantly higher with linguistic cue representation (0.793) than with numeric representation (0.698) ($F(1,87) = 5.91; p = 0.017$) (See Table 3). For auditors, the mean total proportion of variance explained was nearly equal across cue representation conditions (linguistic: 0.703; numeric: 0.709; $F(1,87) = 0.06; p = 0.805$). Cue representation influenced relative cue weightings for both students (Wilks $\Lambda = 0.74, F(4,84) = 7.54, p < 0.001$) and auditors (Wilks $\Lambda = 0.81, F(4,84) = 4.77, p = 0.002$). For both groups, the univariate effect for the management influence cue was significant (students: $F(1,87)$

¹¹Because the distribution of judgment deviation scores was positively skewed, a square root transformation (Neter, Wasserman, and Kutner, 1990) was applied to the data before analysis.

= 27.66; $p < 0.001$; auditors: $F(1,87) = 18.14$; $p < 0.001$), with higher mean omega-squared values in the linguistic relative to the numeric condition (See Table 3).

Response representation did not have a significant effect on the total proportion of variance explained for either students ($F(1,87) = 0.03$; $p = 0.868$) or auditors ($F(1,87) = 2.34$; $p = 0.129$), nor did it have an effect on relative cue weights for either group (students: Wilks $\Lambda = 0.92$, $F(4,84) = 1.76$, $p = 0.144$; auditors: Wilks $\Lambda = 0.94$, $F(4,84) = 1.36$, $p = 0.256$). The cue by response representation interaction for total proportion of variance explained was not significant for either group (students: $F(1,87) = 0.08$; $p = 0.782$; auditors: $F(1,87) = 0.62$; $p = 0.432$), nor was it significant for relative cue weights (students: Wilks $\Lambda = 0.95$, $F(4,84) = 1.21$, $p = 0.313$; auditors: Wilks $\Lambda = 0.94$, $F(4,84) = 1.23$, $p = 0.303$).

Experience

Risk Judgments: Students' risk judgments (0.847) were significantly higher than auditors' (0.816) ($F(1,87) = 12.38$, $p = 0.001$). Students' judgments were significantly ($p \leq 0.05$) higher on eight of the 16 individual cases. Six of the eight were relatively 'high-risk' cases, that is, cases with two or more decision cues at the high level.

Judgment deviation: The students' mean judgment deviation across all cases (0.063) was significantly higher than the auditors' (0.053) ($F(1,87) = 7.19$, $p = 0.009$), indicating greater consensus among auditors. Students' judgment deviation scores were significantly ($p \leq 0.05$) higher than auditors' on seven individual cases. Six of these cases had two or fewer cues at the high level, indicating greater consensus among auditors on moderate to low risk cases. Students' judgment deviation was less than auditors' only in the case where all four cues were at the high level, since all students responded at or very near 1.0 for this case (mean response = 0.99). The mean risk judgment and judgment deviation results suggest that students may have anchored on an initial

assessment of 1.0 for the highest risk experimental case and adjusted downwards for the other cases.

Proportion of variance explained: Experience did not affect the total proportion of variance explained by students' and auditors' judgment models ($F(1,87) = 1.59; p = 0.211$). However, there were significant differences in relative cue weighting between students and auditors (Wilks $\Lambda = 0.87, F(4,84) = 3.20, p = 0.017$). Students' mean omega-squared values were significantly higher than auditors' for both the management incentives ($F(1,87) = 6.51; p = 0.012$) and management influence on accounting policies ($F(1,87) = 4.78; p = 0.031$) cues. Auditors' mean omega-squared values were marginally higher than students for the previous audit errors cue ($F(1,87) = 3.23; p = 0.076$).

Discussion

Summary of Results: Cue representation affected relative cue weighting for both groups. Both auditors and students had higher explained variance for the management influence cue with linguistic representation. Cue representation also impacted both judgment deviation and the percentage of variance explained for students, but not for auditors. Students had lower judgment deviation (i.e., higher consensus) and their judgment models explained a higher proportion of variance with linguistic cue representation relative to numeric. Accordingly, the results support the conjecture that students make more accurate risk assessments when using linguistic cue representations. However, the benefits of linguistic cue representation appear to disappear with experience. In contrast, response representation had no effect on decision accuracy, for either students or auditors.

Experience and Risk Judgments: An unexpected result was that students' inherent risk judgments were consistently higher than auditors, especially in cases where the decision cues indicated relatively high risk. One explanation for

this result is that the auditors applied information from their personal auditing experiences about the frequency of inherent risk problems, while students had no such information available. Previous research supports this speculation. For example, Christensen-Szalanski et al. (1983) found that medical doctors' risk judgments of disease mortality were more accurate than those of students', and were highly correlated with the frequency of the doctors' personal encounters with the diseases. Libby and Frederick (in press) found that experienced auditors were more accurate at identifying the causes of financial statement errors, potentially indicating the use of base rate knowledge derived from personal encounters with errors.

Given that audit textbooks tend to focus on the causes and detection of audit errors, students may apply a representativeness heuristic and believe that the probability of material error occurring for a given type of client is greater than it actually is. On the other hand, auditors who in their experience have only infrequently detected material errors will likely have lower initial anchors for inherent risk judgments. If the adjustment processes used by students and auditors are similar, then the lower initial anchors used by auditors would produce lower inherent risk assessments.

Effects of Standardized Scaling: For the lowest risk case (i.e., all cues at the low level), both auditors' (0.67) and students' (0.66) mean risk assessments were close to the 0.60 lower boundary on the pre-established scale used in the experiment. The lower boundary may have therefore artificially restricted participants' risk assessments in low risk cases. Thus, the lower scale bound could have decreased the explained variance of participants' judgments by constraining the range of risk assessments available to participants. An approach to addressing this problem is to allow participants to establish their own (i.e., individual) risk assessment scales.

'Mixed' Cue Information: An important information representation issue not explored in Experiment 1 is the effect of 'mixed' cue information on decision making (Fennema, 1990). In practice, auditors must combine numeric (e.g., '\$150,000' net loss last year) and linguistic ('capable and experienced' management) information to make risk assessments. As a result, investigating the impact on decision making of combining 'mixed' information representations holds relevance to audit practice. It is hypothesized that using 'mixed' information results in an intermediate condition between the pure numeric and pure linguistic conditions explored in Experiment 1. That is, that using 'mixed' cue information results in judgment accuracy and cognitive effort that fall between the extremes of pure numeric and pure linguistic information.

Perceived and Actual Cognitive Effort: Experiment 1 also provides no information on the perceived and actual cognitive effort of participants. Research suggests that information representation can significantly change perceived and actual (Kotovsky, Hayes, and Simon, 1985; Schkade and Kleinmuntz, 1990; Stone and Schkade, in press) cognitive effort. The three inherent risk judgment strategies discussed previously assume accuracy and effort trade-offs in inherent risk judgments. A second experiment was undertaken to examine the relationship between judgment accuracy and cognitive effort in inherent risk judgments.

Motivation for Experiment 2: To summarize, Experiment 2 was designed to explore four issues: (1) auditors' versus students' initial inherent risk anchors, (2) the effect of standardized vs. individual scaling in risk judgments, (3) the effect of 'mixed' (i.e., some numeric/some linguistic) cue information on risk judgments, and (4) the effect of cue and response representation manipulations on participants' perceived and actual cognitive effort.

Experiment 2

Method

Experimental Task

Participants completed the experimental task using microcomputers equipped with a computer mouse and custom-built software. The software recorded traces of participants' decision processes, similar to the Mouselab program (Johnson, Payne, Schkade, and Bettman, 1988). Participants began the task by reading the inherent risk definition used in Experiment 1. They next established equivalencies between linguistic and numeric inherent risk labels. To do this, participants first indicated the lowest overall inherent risk they believed would ever exist in an audit (Figure 4). After responding to this question, they were shown a scale with endpoints of 1.0 and their stated lowest possible inherent risk value. They completed the scale by entering numeric equivalencies for three additional risk labels ("moderately low," "moderate," "moderately high") (Figure 5). These numeric/linguistic equivalencies were intended to provide data on the initial anchors used by participants.

After completing the scale, participants were presented with the case information and cue descriptions used in Experiment 1. Subsequently, participants responded to a set of three practice cases. The practice cases represented low, medium, and high inherent risk scenarios, and participants were informed of this in the task instructions. Following completion of the practice cases, participants had the opportunity to reset their scaling equivalencies, if they wished to do so. They then proceeded to the sixteen actual experimental cases. The same 2^4 within-subjects design as in Experiment 1 was used.

After finishing the computerized part of the experiment, participants completed a post-experimental questionnaire that used Likert-type scales to

gather data on participants' perceptions of the experiment. The post-experimental questionnaire included four questions that asked participants to estimate the percentage of audit engagements in which they believed high risk conditions existed for each of the four cues used in the experiment (see Figure 10).

The cues for both the practice and actual cases were "hidden" in labelled boxes. Participants obtained data by using the mouse to move a cursor into a box and then clicking. Each box remained open until another box was clicked; only one box could be open at a time. As in Experiment 1, cue representations were either numeric or linguistic. Figure 6 illustrates the numeric cue representation screen display when a participant is examining the management incentives cue. Figure 7 illustrates the linguistic representation of the management incentives cue.

When participants were ready to make an assessment, they clicked a box at the top of the display that enabled them to proceed to an assessment screen (Illustrated on the first line of Figures 6-9). Participants in the numeric response representation condition responded by clicking on a continuous response scale that ranged from the participant's lowest stated assessment of inherent risk to 1.0 (Figure 8). Intermediate risk levels were marked on the scale at equal intervals. The numeric value for any point chosen on the response scale was displayed on the screen. Participants clicked points on the scale until they reached their desired risk assessment value. Participants in the linguistic response representation condition responded by clicking in one of five boxes with linguistic labels (Figure 9).

Two 'mixed' cue conditions were added to the all numeric and all linguistic conditions used in Experiment 1. In mixed condition 1, the management incentives and management influence on accounting cues were presented using numeric representation, the other two cues (i.e., material errors, product

complexity) with linguistic representation. These representation modes were reversed in mixed condition 2 (i.e., material errors and product complexity -- numeric; management incentives and management influence -- linguistic). The three experimental factors of experience, cue representation, and response representation were manipulated in a 2 x 4 x 2 crossed design. Both case orders and cue orders were randomized across subjects in Experiment 2.

Participants

Sixty-four students from an introductory auditing class completed the experiment approximately three weeks after having been introduced to the audit risk model. In addition, 60 auditor participants completed the study. Thirty-eight auditors were from 'Big Six' firms, the remaining 22 were from other large firms. Twenty-eight auditors completed the task in a computer lab while visiting campus; the remaining 32 completed it in their practice office during regular work hours. The range of auditors' professional experience was from one-and-a-half to 15 years, with a mean of 6.4 years and a median of six years.¹²

As in Experiment 1, we examined regression coefficient and total omega-squared data to screen participants before further analysis. Fifty-nine student participants and 54 auditor participants remained in the sample. Table 4 shows the distribution of remaining participants across cells of the experimental design. Outcome and postexperimental data were analyzed using the same basic models and planned comparisons described in Experiment 1. Process data was analyzed using an ANOVA and the planned comparisons described in Experiment 1.

¹²There were three auditors in this experiment with less than two years of experience, however, they had all completed two audit busy seasons and were doing some audit planning at the time they participated in the study.

Results--Inherent Risk Judgments

Cue and Response Representation

Judgment deviation: Cue representation had a significant impact on students' mean judgment deviation ($F(3,97) = 3.09, p = 0.031$), but not auditors' ($F(3,97) = 0.52, p = 0.669$). Consistent with Experiment 1, students' mean judgment deviation was highest for the all numeric cue representation condition (0.199) and lowest for the all linguistic condition (0.111), indicating greater consensus among students with linguistic cue representation (See Table 5). Students' mean judgment deviation for the mixed cue conditions was between the all numeric and all linguistic conditions, indicating that combining "mixed" numeric and linguistic information did not lower judgment accuracy relative to uniform representations of cues (Tukey HSD (3,97), $p \leq .05$). Response representation did not have a significant effect on judgment deviation for either students ($F(1,97) = 0.19; p = 0.663$) or auditors ($F(1,97) = 1.37; p = 0.245$). In addition, there was no cue by response representation interaction for either group (students: $F(3,97) = 1.20; p = 0.312$; auditors: $F(3,97) = 0.71; p = 0.548$).

Proportion of variance explained: There were no significant cue representation effects on the total proportion of variance explained, for either students ($F(3,97) = 1.49; p = 0.222$) or auditors ($F(3,97) = 1.40; p = 0.249$) (See Table 6.). Cue representation did have an impact on students' relative cue weights (Wilks $\Lambda = 0.73, F(12,249) = 2.64, p = 0.002$), but not auditors' (Wilks $\Lambda = 0.95, F(12,249) = 0.42, p = 0.957$). For students, there were significant univariate results on both the management influence ($F(3,97) = 6.63; p < 0.001$) and accounting complexity ($F(3,97) = 3.00; p = 0.035$) cues. Consistent with Experiment 1, the mean omega-squared value for the management influence cue was highest when all cues were linguistic (0.207) and lowest when all cues were numeric (0.053) (See Table 6.). In the mixed cue conditions, management influence also explained a

higher proportion of the variance in students' judgments with linguistic representation (0.129) than with numeric (0.088) (Tukey HSD (3,97), $p \leq .05$). In contrast, students' proportion of variance explained for the accounting complexity cue was higher with numeric representation (numeric: 0.181; mixed condition 2: 0.275) and lower with linguistic representation (linguistic: 0.128 and mixed condition 1: 0.100) (Tukey HSD (3,97), $p \leq .05$).

As in Experiment 1, response representation did not significantly affect total explained variance for either students ($F(1,97) = 2.23$; $p = 0.139$) or auditors ($F(1,97) = 0.00$; $p = 0.969$), nor did it affect relative cue weights for either group (students: Wilks $\Lambda = 0.97$, $F(4,94) = 0.78$, $p = 0.542$; auditors: Wilks $\Lambda = 0.97$, $F(4,94) = 0.726$, $p = 0.577$). There were no significant cue by response representation interactions for either group for total variance explained (students: $F(3,97) = 0.35$; $p = 0.792$; auditors: $F(3,97) = 0.70$; $p = 0.556$) or relative cue weights (students: Wilks $\Lambda = 0.90$, $F(12,249) = 0.84$, $p = 0.606$; auditors: Wilks $\Lambda = 0.86$, $F(12,249) = 1.19$, $p = 0.292$).

Experience

Scaling Equivalencies & Post-experimental Base Rate Questions: Auditors set significantly lower numeric values for all four linguistic expressions than did students (See Table 7). In addition, the variance of numeric values given as equivalent to the "moderately high" linguistic label is significantly higher for auditors than students ($F(53,58) = 3.50$, $p < 0.001$), indicating lower auditor agreement on a numeric equivalent for this label.

Significant differences also existed between auditors and students in the base rate estimates of the percentage of audits in which two of the four cues were significant audit issues. Students believed that problems related to management incentives ($F(1,97) = 15.84$, $p < 0.001$) and product complexity ($F(1,97) = 7.16$, $p = 0.009$) occurred significantly more often than did auditors. These data are

consistent with the assumption that auditors' inherent risk anchors are lower than those used by students.

Risk Judgments: As in Experiment 1, students' mean risk judgments (0.73) were significantly higher than auditors' (0.57) ($F(1,97) = 27.53, p < 0.001$). Student judgments were significantly ($p \leq 0.05$) higher than auditors' on all 16 individual cases.

Judgment deviation: In contrast to Experiment 1, the auditors' mean judgment deviation (0.182) was marginally greater than students' (0.154) ($F(1,97) = 3.39, p = 0.068$), indicating greater consensus among students. Comparisons on individual cases found that auditors' deviation scores were significantly ($p \leq 0.05$) higher than students' on six individual cases. Five of the six cases had either three or four cues at the high level, and the sixth had two cues at the high level, showing lower consensus among auditors on relatively high risk cases. The low auditor agreement on a numeric equivalent for the "moderately high" linguistic label provides one explanation for the lower consensus among auditors on relatively high risk cases.

Proportion of variance explained: As in Experiment 1, the total proportion of variance explained by individual judgment models was not significantly different between students and auditors ($F(1,97) = 0.08, p = 0.774$). Unlike Experiment 1, there were no significant student/auditor differences in relative cue weighting (Wilks $\Lambda = 0.97, F(4,94) = 0.63, p = 0.645$).

Results-Process & Post-Experimental Data

Cue & Response Representation

Cognitive Effort: Auditors, and to a lesser extent students, examined information differently depending upon cue representation. Cue representation had a significant impact on the number of cue acquisitions made by auditors ($F(3,97) = 2.9, p = 0.039$) (See Table 8.). Auditors made the greatest average

number of cue acquisitions per judgment with numeric cue representation (5.5) and the smallest with linguistic representation (4.2), indicating that, on average, auditors reexamined one more cue per judgment with numeric cue representation (Tukey HSD (3,97), $p \leq .05$). Cue representation also affect the time auditors spent on each cue acquisition ($F(3,97) = 2.81, p = 0.044$). They averaged 4.7 seconds per acquisition with linguistic representation, 3.7 seconds with numeric (Tukey HSD (3,97), $p \leq .05$).

There were no differences due to cue representation in students' total cue acquisitions ($F(3,97) = 0.84, p = 0.474$), but it did have a marginal effect on their time per acquisition ($F(3,97) = 2.33, p = 0.079$). Students averaged 4.1 seconds per acquisition with linguistic representation, 3.1 seconds with numeric. For both students and auditors, the mixed cue representation condition means generally fell between the pure numeric and pure linguistic conditions for the number of cue acquisitions and time per acquisition results.

Both auditors ($F(1,97) = 13.11, p < 0.001$) and students ($F(1,97) = 8.20, p = 0.005$) took significantly longer to make assessments with numeric response scales relative to linguistic. With the numeric response scale, the auditors' mean assessment time was 89.0 seconds, while students averaged 78.0 seconds. With the linguistic response scale, auditors averaged 51.2 seconds and students averaged 49.2 seconds.

Experience

Cognitive Effort: Auditors took significantly longer for each cue acquisition than did students ($F(1,97) = 6.22, p = 0.014$) (3.97 vs. 3.48 seconds, respectively). There were no differences due to experience in total cue acquisition time ($F(1,97) = 1.93, p = 0.168$) or total number of cue acquisitions ($F(1,97) = 0.01, p = 0.928$).

Perceived Cognitive Effort & Accuracy: Auditors were more confident in their risk assessments than students ($F(1,97) = 17.13, p < 0.001$). There were no

significant differences in perceived task difficulty due to experience ($F(1,97) = 0.05$, $p = 0.825$). Interestingly, auditors rated the experimental task as more realistic than did students ($F(1,97) = 4.03$, $p = 0.047$), perhaps since auditors are more familiar with computer-assisted auditing technology.

Discussion

Cue & Response Representation: Response representation affected cognitive effort for both auditors and students; both groups expended greater effort using numeric response scales. As in Experiment 1, response representation had no effect on decision accuracy. Accordingly, numeric response representations increased cognitive effort with no corresponding increase in judgment accuracy.

As in Experiment 1, students had greater consensus with linguistic cue representation. In addition, cue representation again affected students' relative cue weights. However, in contrast to Experiment 1, there were no differences in the total proportion of variance explained by students' judgmental models. Thus, Experiment 2 provides additional evidence suggesting that, for students, linguistic cue representation leads to more accurate inherent risk judgments. However, such effects appear to be partially mitigated by the type of scaling (i.e., standardized vs. individualized) used, with more significant cue representation effects occurring with standardized scaling.

Inherent Risk Decision Strategies: Cue representation affected auditors' decision processes. They reexamined more cues with numeric cue representation, but took less time per examination. Accordingly, auditor decision processes were contingent upon the cue representation manipulation. However, consistent with Experiment 1, there were no significant cue representation effects on auditors' judgment deviation or total proportion of variance explained measures.

Auditors set significantly lower numeric equivalencies for linguistic expressions of risk than did students, and believed that inherent risk problems occurred on a smaller percentage of audits. In addition, auditors were more confident in their risk judgments. These results suggest that auditors may apply information from their personal experience to inherent risk judgments, and that this information leads them to establish lower initial risk anchors. In contrast, students appear to set initial risk anchors close to a more conservative 1.0.

Auditors in Experiment 2 had marginally higher judgment deviation than did students, primarily on cases where the majority of cues indicated high inherent risk. In addition, auditors had lower agreement than students on a numeric equivalent for the linguistic term "moderately high" inherent risk. If the base rate of audit clients with significant inherent risk problems is low, then auditors will have less personal experience in assessing inherent risk for high risk cases. As a consequence, auditors' judgments may be less accurate for such cases. In practice, consultation on high risk clients with expert auditors acting in an advisory role to the audit team may serve to mitigate this effect.

Mixed Cue Representations: Combining "mixed" cue information (i.e., some numeric and linguistic cues) did not appear to pose particular difficulties for either auditors or students. With only one exception (students' proportion of variance explained by the accounting complexity cue), the mixed cue conditions resulted in cognitive effort and proportions of variance explained measures that fell between the pure numeric and pure linguistic conditions. Accordingly, the results suggest that integrating "mixed" representation information into risk judgments does require differential cognitive effort relative to pure information representation.

General Discussion

The error and effort framework discussed earlier provided a useful theory for examining inherent risk judgments. Previous explorations of inherent risk judgment have either analyzed auditor's linear models (Colbert, 1988; Daniel, 1988), or focused on creating detailed, purely descriptive models of auditor judgment (Peters, 1989). Using the error and effort framework has enabled us to produce a descriptive theory of the inherent risk judgment strategies based upon the assumption that auditors will exhibit *contingent decision behavior* -- that is, auditors' decision processes are dependent upon the characteristics of decision maker's knowledge, task, context, and information display (Kleinmuntz and Schkade, 1990). We believe theories based upon contingent decision behavior hold great explanatory power in understanding auditor decision strategies and are deserving of greater attention than evidenced in extant auditing research.

Inherent Risk Decision Strategies

One unexpected finding of our research was that auditors' risk assessments were consistently lower than students'. The results of our second experiment suggest that this is because auditors use lower initial anchors in making risk assessments. We speculate that this occurred because auditors bring personal experience about inherent risk problems to the task. To the extent such experience is representative of the base rate of inherent risk problems in audit practice, it will increase the accuracy of experienced auditors' judgments. However, the use of information from personal experience is susceptible to representativeness and availability judgmental biases. Such biases have been observed among medical doctors making risk judgments (Christensen-Szalanski, et al., 1983). Additional research will be required to determine whether the differing initial anchors used by auditors result from personal experiences and whether such experiences lead to systematic judgmental biases.

Information Representation

We did not find support for previous suggestions that the use of numeric response representations for expressing risk will increase judgment accuracy relative to linguistic representations (Chesley, 1985). In fact, we found that both auditors and students expended much greater effort on numeric response scales with no corresponding increase in judgment accuracy. As a result, auditors may incur additional costs using numerical expressions of inherent risk, without deriving corresponding benefits. An alternative explanation for this result, however, is that the particular operationalizations chosen for the numeric and linguistic response scales produced the result. The numeric response scales used in both experiments were continuous, while the linguistic response scales were discrete, with five possible responses. This operationalization (i.e., numeric -- continuous, linguistic -- discrete) was chosen since previous research argues that numeric response representations allow for more finely partitioned judgments (Lichtenstein and Newman, 1967; Chesley, 1985). However, additional research will be required to determine to what extent the observed differences in cognitive effort result from scale representation (numeric versus linguistic) versus the number of scale points (continuous versus discrete). We are currently engaged in research exploring this issue.

For students, our data suggest that linguistic cue representation leads to greater consistency in judgments relative to numeric representations. One implication of this finding (consistent with Chesley, 1979) is that auditing students do not have well-developed skills in using numeric information for risk assessment. This suggests that auditing students and new staff auditors might benefit from additional training in using numeric information. Ashton (1984) offers an example of a training exercise that may be beneficial for auditing students. He suggests giving linguistic information expressions to students (e.g.,

"very likely") and asking them to state numeric equivalencies. These equivalencies are then shared and discussed. Such an exercise may prove useful in developing quantitative risk assessment skills among auditing students.

Experiment 2 suggests that cue representation primarily affects auditors' decision processes. The cue representation effects on auditors' relative cue weightings found in Experiment 1 disappeared in Experiment 2, possibly because of the use of individualized risk scales. An alternative explanation for this result, however, is sample differences between the two experiments. Auditors in Experiment 2 averaged 6.4 years of experience, while those in Experiment 1 averaged 3.5 years. More experienced auditors may be less susceptible to information representation effects. In either case, the effects of information representation decision aids in auditing appear to be complex and not necessarily positive. Boritz (1985) reports similar equivocal results from the use of audit decision aids.

Risk Scale Standardization

The standardized risk scales used in Experiment 1 resulted in higher risk assessments and lower judgment deviation for both auditors and students relative to the individual scaling used in Experiment 2. The mean risk assessments for both students and auditors averaged 0.83 for Experiment 1 and 0.65 for Experiment 2, while mean judgment deviation averaged 0.058 for Experiment 1 and 0.168 for Experiment 2. Accordingly, our results suggest that the use of standardized scaling will result in more conservative, higher consensus risk judgments. However, the joint effect of information representation and risk scale standardization on risk judgments suggests that movement towards standardized risk scales may not be unequivocally beneficial. Using standardized scales resulted in higher auditor consensus, but led to changes in relative cue weighting depending upon cue representation. Using standardized scaling therefore decreased the consistency of auditors' risk judgments across experimental

conditions, but increased their consensus. One explanation for this result is that the difference between auditors' individual perceptions of numeric/linguistic equivalencies and the numeric/linguistic scale equivalencies set in the standardized scale introduced additional variability into the assessment process, thereby decreasing the consistency of judgments across experimental conditions.

In contrast, using individually set scales decreased the consensus of risk judgments, but increased their consistency across experimental conditions. As a result, we hypothesize that the decision to standardize risk assessment scales appears to represent a tradeoff between two sources of risk assessment variance: (1) the error introduced by the difference between individuals' scale equivalencies and equivalencies set in a standardized scale (which decreases the consistency of individual risk judgments) and (2) the increased variability introduced by allowing different scalings across individuals (which decreases the consensus of risk judgments).

In general, we find that issues of information representation and scaling in inherent risk judgments are more complex than previously suggested in the auditing literature. Audit researchers often presume, either implicitly or explicitly, that using standardized scaling and numeric representations for risk judgments will unequivocally improve these judgments. However, the results of our experiments suggest that such changes result in complex tradeoffs between judgment accuracy and effort. Considerable additional research is needed to more fully explore the implications of these tradeoffs. Until audit research produces a better understanding of the processes of audit risk assessment, audit practitioners are advised to exercise caution in implementing decision aids intended to quantify and standardize audit risk assessment.

TABLE 1
Distribution of Participants to Experimental Cells
Experiment 1

Students

| <u>Cue Representation</u> | <u>Response Representation</u> | | <u>Totals</u> |
|---------------------------|--------------------------------|-------------------|---------------|
| | <u>Numeric</u> | <u>Linguistic</u> | |
| Numeric | 12 | 16 | 28 |
| Linguistic | 11 | 16 | 27 |
| Totals | 23 | 32 | 55 |

Auditors

| <u>Cue Representation</u> | <u>Response Representation</u> | | <u>Totals</u> |
|---------------------------|--------------------------------|-------------------|---------------|
| | <u>Numeric</u> | <u>Linguistic</u> | |
| Numeric | 11 | 11 | 22 |
| Linguistic | 8 | 10 | 18 |
| Totals | 19 | 21 | 40 |

TABLE 2
Deviation Scores by Cue Representation and Experience
Experiment 1

| <u>Cue Representation</u> | <u>Students</u> | <u>Auditors</u> |
|---------------------------|-----------------|-----------------|
| Numeric | 0.068 | 0.055 |
| Linguistic | 0.057 | 0.050 |
| Means | 0.063 | 0.053 |
| F(1,87) | 6.31 | 1.14 |
| p | 0.014 | 0.288 |

TABLE 3
Proportion of Variance Explained by Cue Representation and Experience
Experiment 1

| | <u>Management</u> <u>Incentives</u> | <u>Management In-</u> <u>fluence on Ac-</u> <u>counting Policies</u> | <u>Previous</u> <u>Errors</u> | <u>Accounting</u> <u>Complexity</u> | <u>Total</u> |
|---|--|--|----------------------------------|--|--------------|
| <u>Students</u> | | | | | |
| Numeric | 0.280 | 0.065 | 0.212 | 0.141 | 0.698 |
| Linguistic | 0.196 | 0.212 | 0.244 | 0.142 | 0.793 |
| Means | 0.239 | 0.137 | 0.228 | 0.142 | 0.745 |
| <u>Significance of cue representation</u> <u>effect within students:</u> | | | | | |
| F(1,87) | 1.99 | 27.67 | 0.68 | 0.00 | 5.91 |
| p | 0.162 | <0.001 | 0.413 | 0.958 | 0.017 |
| <u>Auditors</u> | | | | | |
| Numeric | 0.183 | 0.028 | 0.290 | 0.208 | 0.709 |
| Linguistic | 0.096 | 0.156 | 0.303 | 0.148 | 0.703 |
| Means | 0.144 | 0.085 | 0.296 | 0.181 | 0.706 |
| <u>Significance of cue representation</u> <u>effect within auditors:</u> | | | | | |
| F(1,87) | 2.44 | 18.14 | 0.01 | 0.50 | 0.06 |
| p | 0.122 | <0.001 | 0.911 | 0.481 | 0.805 |
| <u>Significance of student vs. auditor</u> <u>contrasts:</u> | | | | | |
| F(1,87) | 6.51 | 4.78 | 3.23 | 1.73 | 1.59 |
| p | 0.012 | 0.031 | 0.076 | 0.191 | 0.211 |

TABLE 4
Distribution of Participants to Experimental Cells
Experiment 2

Students

| <u>Cue Representation</u> | <u>Response Representation</u> | | <u>Totals</u> |
|---------------------------|--------------------------------|-------------------|---------------|
| | <u>Numeric</u> | <u>Linguistic</u> | |
| Numeric | 7 | 7 | 14 |
| Mixed 1 ^a | 8 | 7 | 15 |
| Mixed 2 ^b | 8 | 8 | 16 |
| Linguistic | 8 | 6 | 14 |
| Totals | 31 | 28 | 59 |

Auditors

| <u>Cue Representation</u> | <u>Response Representation</u> | | <u>Totals</u> |
|---------------------------|--------------------------------|-------------------|---------------|
| | <u>Numeric</u> | <u>Linguistic</u> | |
| Numeric | 8 | 6 | 14 |
| Mixed 1 ^a | 6 | 8 | 14 |
| Mixed 2 ^b | 8 | 6 | 14 |
| Linguistic | 5 | 7 | 12 |
| Totals | 27 | 27 | 54 |

^a Management incentives and management influence cues, numeric; previous audit errors and product complexity cues, linguistic.

^b Management incentives and management influence cues, linguistic; previous audit errors and product complexity cues, numeric.

TABLE 5
Consensus Scores by Cue Representation and Experience
Experiment 2

| <u>Cue Representation</u> | <u>Students</u> | <u>Auditors</u> |
|---|--------------------|-----------------|
| Numeric | 0.199 ¹ | 0.165 |
| Mixed 1 ^a | 0.165 ² | 0.196 |
| Mixed 2 ^b | 0.142 ³ | 0.162 |
| Linguistic | 0.111 ⁴ | 0.207 |
| Means | 0.154 | 0.182 |
| <u>Significance of Cue Representation Effects</u> | | |
| <i>F</i> (3, 97) | 3.09 | 0.52 |
| <i>p</i> | 0.031 | 0.669 |

¹ Numbers denote significant post-hoc comparison differences (Tukey HSD (3,97), $p \leq 0.05$).

^a Management incentives and management influence cues, numeric; previous audit errors and product complexity cues, linguistic.

^b Management incentives and management influence cues, linguistic; previous audit errors and product complexity cues, numeric.

TABLE 6
Proportion of Variance Explained by Cue Representation and Experience
Experiment 2

| | <u>Management Incentives</u> | <u>Management Influence</u> | <u>Previous Errors</u> | <u>Accounting Complexity</u> | <u>Total</u> |
|--|------------------------------|-----------------------------|------------------------|------------------------------|--------------|
| <u>Students</u> | | | | | |
| Numeric | 0.188 | 0.053 ⁴ | 0.280 | 0.181 ² | 0.702 |
| Mixed 1 ^a | 0.255 | 0.088 ³ | 0.345 | 0.100 ³ | 0.789 |
| Mixed 2 ^b | 0.135 | 0.129 ² | 0.247 | 0.275 ¹ | 0.787 |
| Linguistic | 0.219 | 0.207 ¹ | 0.254 | 0.128 ³ | 0.808 |
| Mean | 0.198 | 0.119 | 0.281 | 0.173 | 0.772 |
| Significance of cue representation effect within students: | | | | | |
| F(3,97) | 1.68 | 6.63 | 0.87 | 3.00 | 1.49 |
| p | 0.176 | <0.001 | 0.462 | 0.035 | 0.222 |
| <u>Auditors</u> | | | | | |
| Numeric | 0.181 | 0.089 | 0.248 | 0.227 | 0.745 |
| Mixed 1 ^a | 0.236 | 0.087 | 0.247 | 0.182 | 0.752 |
| Mixed 2 ^b | 0.197 | 0.126 | 0.233 | 0.257 | 0.813 |
| Linguistic | 0.161 | 0.096 | 0.259 | 0.180 | 0.695 |
| Mean | 0.195 | 0.099 | 0.246 | 0.213 | 0.753 |
| Significance of cue representation effect within auditors: | | | | | |
| F(3,97) | 0.76 | 0.40 | 0.08 | 0.14 | 1.40 |
| p | 0.519 | 0.751 | 0.973 | 0.933 | 0.249 |

^a Management incentives and management influence cues, numeric; previous audit errors and product complexity cues, linguistic.

^b Management incentives and management influence cues, linguistic; previous audit errors and product complexity cues, numeric.

¹ Numbers denote significant post-hoc comparison differences (Tukey HSD (3,97), $p \leq 0.05$).

TABLE 7
Numeric Equivalencies for Linguistic Expressions

| <u>Linguistic Value</u> | <u>Students</u> | <u>Auditors</u> | <u>F (1,97)</u> | <u>p</u> |
|-------------------------|-----------------|-----------------|-----------------|----------|
| Lowest | 0.390 | 0.131 | 49.30 | <0.001 |
| Moderately Low | 0.502 | 0.280 | 45.46 | <0.001 |
| Moderate | 0.670 | 0.490 | 40.58 | <0.001 |
| Moderately High | 0.859 | 0.724 | 29.88 | <0.001 |
| Highest ¹ | 1.000 | 1.000 | N/A | N/A |

¹Set as a constant at 1.0 for all participants.

TABLE 8
Cue Acquisitions and Time per Acquisition
by Cue Representation and Experience

| | <u>Total Cue</u> <u>Acquisitions</u> | <u>Time per</u> <u>Acquisition</u> |
|---|---|---------------------------------------|
| <u>Students</u> | | |
| Numeric | 85.21 | 3.07 |
| Mixed 1 ^a | 77.60 | 3.23 |
| Mixed 2 ^b | 76.50 | 3.58 |
| Linguistic | 73.57 | 4.06 |
| Mean | 78.15 | 3.48 |
| <u>Significance of cue representation</u> <u>effect within students:</u> | | |
| <i>F</i> (3,97) | 0.84 | 2.33 |
| <i>p</i> | 0.474 | 0.079 |
| <u>Auditors</u> | | |
| Numeric | 87.29 ¹ | 3.73 ³ |
| Mixed 1 ^a | 85.21 ¹ | 3.47 ³ |
| Mixed 2 ^b | 73.07 ² | 4.10 ² |
| Linguistic | 67.58 ³ | 4.68 ¹ |
| Mean | 78.69 | 3.97 |
| <u>Significance of cue representation</u> <u>effect within auditors:</u> | | |
| <i>F</i> (3,97) | 2.90 | 2.81 |
| <i>p</i> | 0.039 | 0.044 |
| <u>Significance of student vs.</u> <u>auditor contrasts:</u> | | |
| <i>F</i> (3,97) | 0.01 | 6.22 |
| <i>p</i> | 0.928 | 0.014 |

^a Management incentives and management influence cues, numeric; previous audit errors and product complexity cues, linguistic.

^b Management incentives and management influence cues, linguistic; previous audit errors and product complexity cues, numeric.

¹ Numbers denote significant post-hoc comparison differences (Tukey HSD (3,97), $p \leq 0.05$).

FIGURE 1
Cue Descriptions Given to Participants
Experiment 1

Management incentives: Management incentives refers to the extent to which management's compensation is based on bonus plans and stock options versus salary. In some cases, management's compensation has little relationship to net income (i.e., 5% of management's compensation is from bonus plans / stock options). In other cases, management's compensation is highly dependent upon net income (i.e., 50% of management's compensation is from bonus plans / stock options).

Management's influence on accounting: Management's influence on accounting is the extent to which upper-level management makes year-end changes to accounting estimates. In some cases, Fletcher's upper-level management has historically made only minor adjustments to the financial statements after preliminary trial balance figures were received from accounting (i.e., adjustments made = 1% of net income before taxes). In other cases, Fletcher's upper-level management has historically made significant adjustments to the preliminary trial balance figures (i.e., adjustments made = 10% of net income before taxes).

Previous audit errors: In some cases, total errors discovered in last year's audit of Fletcher were not material (i.e., total errors = 1% of net income). In other cases, total errors discovered in last year's audit of Fletcher were material (i.e., total errors = 20% of net income). [Adjustments for material errors were made before issuing last year's audit opinion.]

Product complexity: Fletcher is in two lines of business: (1) manufacturing microcomputers and (2) manufacturing and configuring custom mainframe systems. The microcomputer manufacturing process is relatively simple. Estimating costs related to this process is fairly routine. The mainframe manufacturing process is significantly more complex. At year-end, the mainframe manufacturing operation requires subjective estimates of revenue recognition and work-in-process inventory. These estimates are made by accounting personnel.

In some cases, relatively little of Fletcher's business involves complex manufacturing processes (i.e., 10% of revenues and cost of goods sold is from the custom mainframe line). In other cases, a substantial portion of Fletcher's business involves complex manufacturing processes (i.e., 70% of revenues and cost of goods sold is from the custom mainframe line).

In the cases that follow, please assess inherent risk for the Fletcher audit under different combinations of the above factors. In assessing inherent risk, please consider 1.0 inherent risk to be very high, 0.9 inherent risk to be high, 0.8 inherent risk to be moderate, 0.7 inherent risk to be low, and 0.6 inherent risk to be very low. The diagram below illustrates this relationship.

Inherent Risk Levels

| | | | | |
|-------------------------|-----|----------|------|--------------|
| Very Low | Low | Moderate | High | Very High |
| 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
| ----- ----- ----- ----- | | | | |

FIGURE 2
Numeric and Linguistic Cue Representations
of "Low" Cue Levels

Linguistic Representations for "low" level of cue

Assume the following facts about Fletcher manufacturing:

Management incentives: A small percentage of management's compensation is based upon bonus plans and stock options.

Management's influence on accounting: Historically, management has recommended only minor adjustments before taxes at year-end.

Previous audit errors: The total errors discovered in last year's audit of Fletcher were immaterial.

Product complexity: A small portion of Fletcher's business involves complex manufacturing processes.

Numeric Representations for "low" level of cue

Assume the following facts about Fletcher manufacturing:

Management incentives: 5% of management's compensation is based upon bonus plans and stock options.

Management's influence on accounting: Historically, management has recommended adjustments = 1% of net income before taxes at year-end.

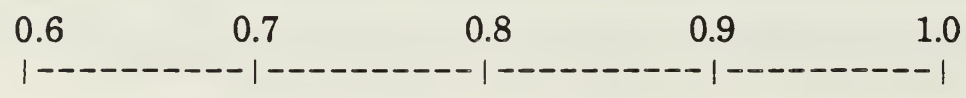
Previous audit errors: The total of errors discovered in last year's audit of Fletcher equalled 1% of net income before taxes.

Product complexity: 10% of Fletcher's business involves complex manufacturing processes.

FIGURE 3 Numeric and Linguistic Response Representations

Numeric response representation:

Please indicate your assessment of inherent risk by placing an "X" on the line below:



Linguistic response representation:

Please indicate your assessment of inherent risk by placing an "X" in the appropriate box below:

| | |
|-----------|--------------------------|
| Very Low | <input type="checkbox"/> |
| Low | <input type="checkbox"/> |
| Moderate | <input type="checkbox"/> |
| High | <input type="checkbox"/> |
| Very High | <input type="checkbox"/> |

Figure 4
Set Lowest Possible Inherent Risk

SET SCALE

Page 2 of 3

Creating a scale requires establishing equivalencies between words and numbers. The first step in establishing the scale is to set the range.

The maximum (HIGHEST) value of inherent risk is 1.00.

Please ENTER the minimum (LOWEST) value of inherent risk that you believe would ever be appropriate in an audit. (Type the value, then press ENTER. To change the value, CLICK the value box.)



A screenshot of a software interface. It shows a dark rectangular area with a grid pattern. On the right side of this area, the number "0.30" is displayed in a white font. The rest of the area is mostly black with some faint, illegible text.

Remember that the ratings you give later must fall within this range.

Next Page

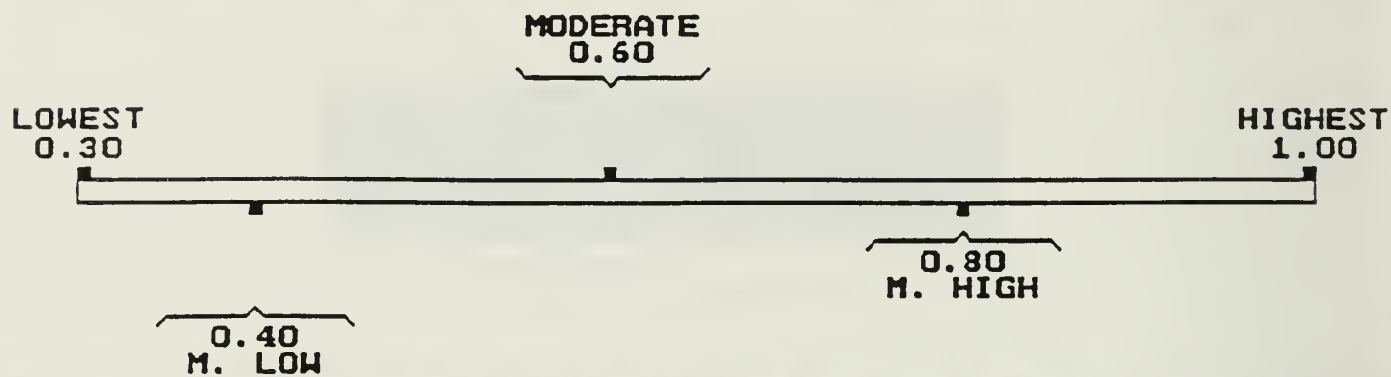
Figure 5
Set Inherent Risk Equivalencies

SET SCALE

Page 3 of 3

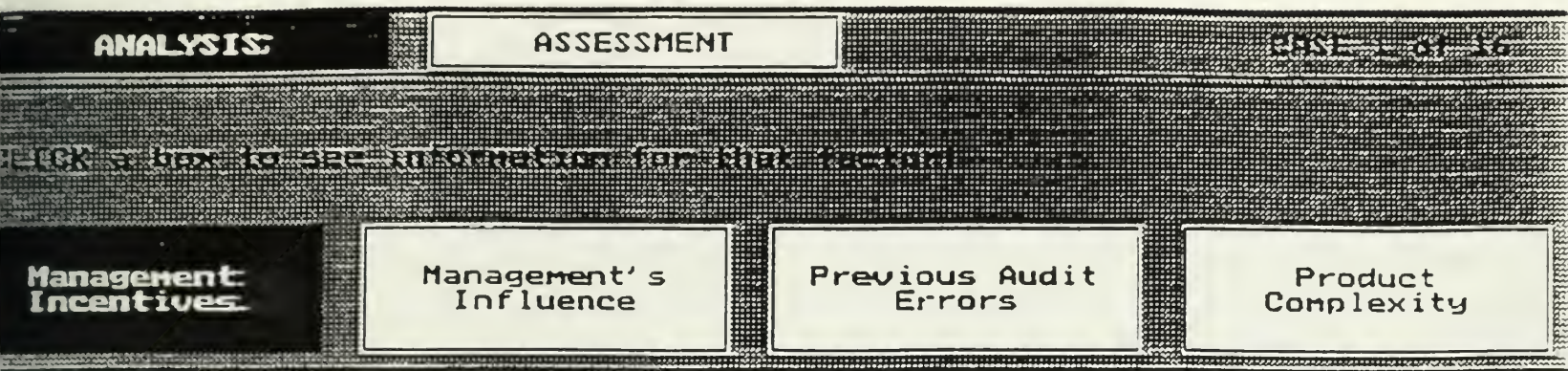
You will now enter some additional points on your scale.
Please ENTER numeric values for each of the following word descriptions
of inherent risk. (To change a value, CLICK the value box.)

| | |
|-------------------------------|------|
| MODERATELY HIGH INHERENT RISK | 0.80 |
| MODERATE INHERENT RISK | 0.60 |
| MODERATELY LOW INHERENT RISK | 0.40 |



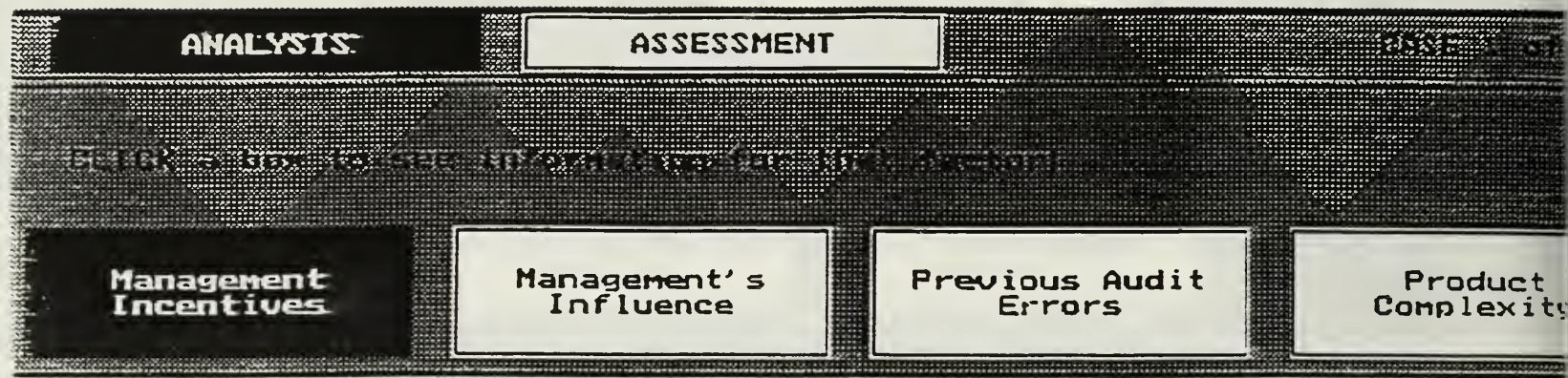
Next Page

Figure 6
Analysis Screen Display --
Numeric Representation of Management Incentives Cue



5% of management's compensation is from bonus plans and stock options.

Figure 7
Analysis Screen Display --
Linguistic Representation of Management Incentives Cue



A small percentage of management's compensation is from bonus plans and stock options.

Figure 8
Assessment Screen Display --
Numeric Response Scale

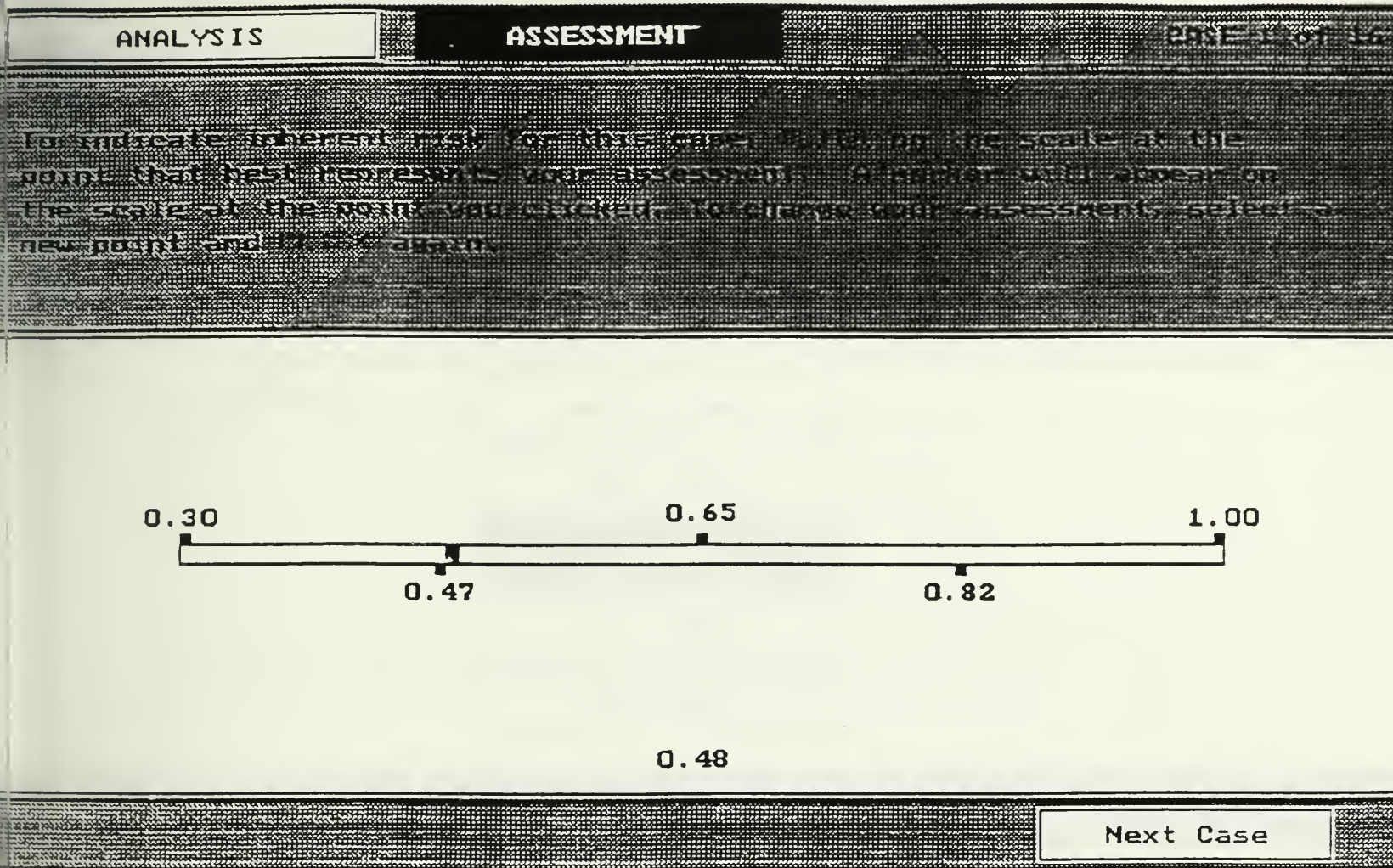


Figure 9
Assessment Screen Display --
Linguistic Response Scale

ANALYSIS **ASSESSMENT** CASES

To indicate when a case is assessed, click on the box that best represents your assessment. To change your assessment, click a box and click again.

Very High

Mod. High

Moderate

Mod. Low

Very Low

Next Case

FIGURE 10

Please estimate the percentage of audit engagements in which you believe each of the following situations occur:

- a. A high percentage of management's compensation (i.e., $\geq 50\%$) is based upon bonus plans and stock options _____%
- b. Management recommends significant adjustments to net income (i.e., $\geq 15\%$) at year-end _____%
- c. Significant errors (i.e., $\geq 15\%$ of net income) are discovered during the audit _____%
- d. A significant portion of a manufacturing client's business (i.e., $\geq 70\%$) requires subjective estimates of revenue recognition and work-in-process inventory _____%

REFERENCES

- Alderman, C. W. and R. H. Tabor, "The Case for Risk-Driven Audits," Journal of Accountancy (March 1989): 55-61.
- Anderson, U. and W. Wright, "Expertise and the Explanation Effect," Organizational Behavior and Human Decision Processes, (1988), Vol. 42: 250-269.
- Arens, A. A., and J. K. Loebbecke, Auditing: An Integreated Approach (4th ed.), Englewood Cliffs, Prentice-Hall, (1988).
- Ashton, A. H., "Does Consensus Imply Accuracy in Accounting Studies of Decision Making?" The Accounting Review (April 1985): 173-185.
- Ashton, R. H., "An Experimental Study of Internal Control Judgments," Journal of Accounting Research (Spring 1974), Vol. 12: 143-157.
- _____, "Human Information Processing in Accounting," Studies in Accounting Research #17 (American Accounting Association, 1982).
- _____, "Integrating Research and Teaching in Auditing: Fifteen Cases on Judgment and Decision Making," The Accounting Review (Jan. 1984): 78-97.
- _____, D. N. Kleinmuntz, J. B. Sullivan, and L. A. Tomassini, "Audit Decision Making," in Research Opportunities in Auditing: The Second Decade, A. R. Abdel-Khalik and I. Solomon, eds., American Accounting Association: Auditing Section (1988): 95-132.
- Bettman, J. R., An Information Processing Theory of Consumer Choice, Addison-Wesley, Reading, MA (1979).
- Boritz, J. E., "The Effect of Information Presentation Structures on Audit Planning and Review Judgments," Contemporary Accounting Research, Vol. 1, No. 2, (1985): 193-218.
- _____, B. G. Gaber, and W. M. Lemon, An Experimental Study of Review of Preliminary Audit Strategy by External Auditors, Toronto, The Canadian Academic Accounting Association, (1987a).
- _____, _____, and _____, "Managing Audit Risk," CA Magazine, (Jan. 1987): 36-41.
- _____, and A.K.P. Wensley, "Evidence, Uncertainty Moelling and Audit Risk," (1988): Working paper, University of Waterloo.
- Buckless, F. A., and S. P. Ravenscroft, "Contrast Coding: A Refinement of ANOVA in Behavioral Analysis," The Accounting Review, (Oct. 1990): 933-945.

- Chesley, G. R., "Procedures for the Communication in Auditors' Working Papers," in Burns, T. J. (Editor), Behavioral Experiments in Accounting II, College of Administrative Science, The Ohio State University, Columbus, OH, (1979).
- _____, "Interpretation of Uncertainty Expressions," Contemporary Accounting Research, 1985, Vol. 2, No. 2: 179-199.
- Christensen-Szalanski, J.J.J., D.E. Beck, C.M. Christensen-Szalanski, and T.D. Koepsell, "Effects of Expertise and Experience on Risk Judgments," Journal of Applied Psychology, Vol. 68, No. 2, (1983): 278-284.
- Colbert, J. L., "Inherent Risk: An Investigation of Auditors' Judgments," Accounting Organizations and Society, Vol. 13, No.2, (1988): 111-121.
- Cushing, B. A. and J. K. Loebbecke, "Analytical Approaches To Audit Risk: A Survey and Analysis," Auditing: A Journal of Practice & Theory, Vol.3, (1983): 23-41.
- Daniel, S. J., "An Empirical Examination of Auditor Inherent Risk Assessments," Working Paper, University of Hawaii (1988).
- Davis, J. and I. Solomon. "Expertise and Experience in Behavioral Accounting Research," Journal of Accounting Literature (1989).
- Elliott, R., "Unique Audit Methods: Peat Marwick International," Auditing: A Journal of Practice and Theory, (Spring 1983): 1-12.
- Erev, I. and B. Cohen, "Verbal versus Numerical Probabilities: Efficiency, Biases, and the Preference Paradox," Organizational Behavior and Human Decision Processes, (Feb. 1990): 1-18.
- _____, C. Gonzalez, and T. Wallsten, "Decision Efficiency as a Function of the Mode of Communicating Uncertainty," (1990), manuscript in preparation.
- Fennema, M.G., "Information Display and Task Characteristics in Multiattribute Choice," unpublished working paper, (1990), University of Illinois.
- Hays, W., Statistics, 3rd edition, New York, CBS College Publishing, (1981).
- Hylas, R. E. and R. H. Ashton, "Audit Detection of Financial Statement Errors," The Accounting Review (October 1982): 751-765.
- Johnson, E. J. and J. W. Payne, "Effort and Accuracy in Choice," Management Science, Vol. 31, No. 4, (April 1985): 395-414.
- _____, _____, D. A. Schkade, and J. R. Bettman, Monitoring Information Acquisitions and Decisions: Mouse Decision Laboratory

Software. Mouselab user's manual, Fuqua School of Business, Duke University, (1988).

Johnson, R. N., "Auditor Detected Errors and Related Client Traits--A Study of Inherent and Control Risks in a Sample of U.K. Audits," Journal of Business Finance and Accounting, Vol. 14, No. 1, (Spring 1987): 39-64.

Joyce, E. J. and G. C. Biddle, "Anchoring and Adjustment in Probabilistic Inference in Auditing," Journal of Accounting Research (Spring 1981): 120-145.

Keppel, G. Design and Analysis: A Researcher's Handbook (2nd ed.) Englewood Cliffs, NJ: Prentice-Hall, 1982.

Kinney, W. R., "Discussant's Response to An Analysis of the Audit Framework Focusing on Inherent Risk and the Role of Statistical Sampling in Compliance Testing," in Auditing Symposium VII, H. Stettler, ed., University of Kansas Press, 1984: 126-132.

_____, "Achieved Audit Risk and the Audit Outcome Space," Auditing: A Journal of Practice and Theory, 1989 Supplement: 67-84.

_____, and W.C. Uecker, "Mitigating the Consequences of Anchoring in Auditor Judgments," The Accounting Review (January 1982): 55-69.

Kleinmuntz, D. N., and D. A. Schkade, "Cognitive Processes and Information Displays in Computer-Supported Decision Making: Implications for Research," Working Paper, University of Illinois at Urbana-Champaign, February, 1990.

Kotovsky, K., J.R. Hayes, and H.A. Simon, "Why Are Some Problems Hard? Evidence from Tower of Hanoi," Cognitive Psychology, 17 (1985), 248-294.

Kreutzfeldt, R. W. and W. A. Wallace, "Error Characteristics in Audit Populations: Their Profile and Relationship to Enviornmental Factors," Auditing: A Journal of Practice & Theory, Vol. 6, No. 1, (Fall 1986): 20-43.

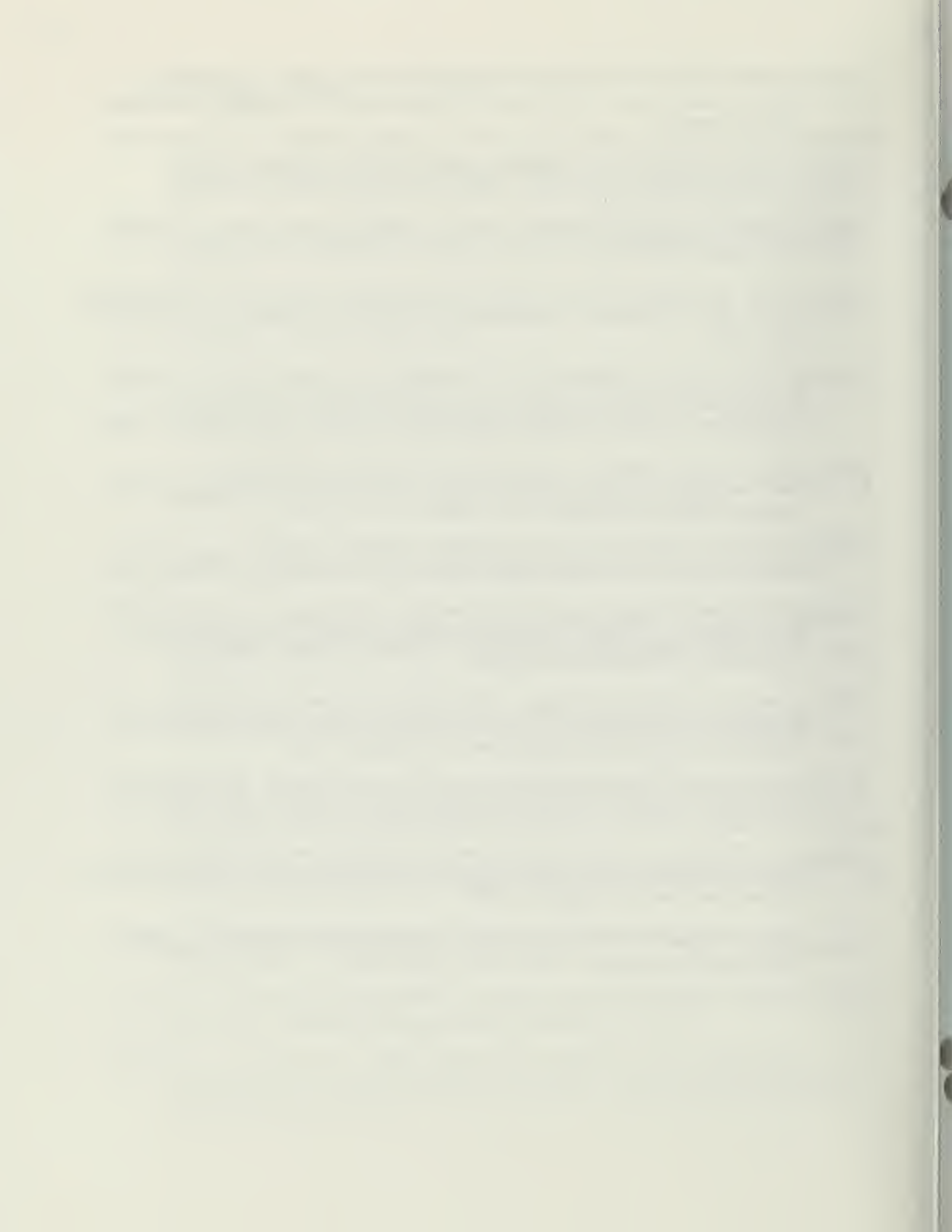
Leichti, J. L., "How to Evaluate Inherent Risk--And Improve Your Audits," The Practical Accountant, March, 1986: 60-64.

Libby, R. Accounting and Human Information Processing: Theory and Applications (Prentice-Hall, 1981).

_____, R. and D.M. Frederick, "Expertise and the Ability to Explain Audit Findings," Journal of Accounting Research (in press).

_____, R. and P. A. Libby, "Expert Measurement and Mechanical Combination in Control Reliance Decisions," The Accounting Review, (October, 1989): 729-747.

- Lichtenstein, S., and J. R. Newman, "Empirical Scaling of Common Verbal Phrases Associated with Numerical Probabilities," Psychonomic Science, (1967): 563-564.
- Neter, J. and W. Wasserman Applied Linear Statistical Models, 2nd ed. Homewood: Richard D. Irwin, 1985.
- Payne, J. W., "Contingent Decision Behavior," Psychological Bulletin, Vol. 92, No. 2, (1982): 382-402.
- Peters, J.M., "Assessing Inherent Risk During Audit Planning: An Operational Model of the Assessment Process," Working Paper, University of Oregon (July 1989).
- Schkade, D. A. and D. N. Kleinmuntz,, "The Effects of Information Display on Strategy Selection in Multiattribute Choice," Presentation at the Association of Consumer Research, New York, New York, October, 1990.
- Stone, D. N. and D. Schkade, "Numeric and Linguistic Information Representation in Multiattribute Choice," Organizational Behavior and Human Decision Processes, (in press).
- Waller, W. S., "Towards a Data-Based Model of Auditors' Risk Assessment," Working paper, presented at the 1990 USC Audit Judgment Symposium.
- Wallsten, T. S., "The Costs and Benefits of Vague Information," In R. Hogarth (Ed.) Insights in Decision Making: Theory and Applications. Chicago: University of Chicago Press, 1988.
- Willingham, J. J. and W. F. Wright , "Estimation of Errors in Financial Statements" Proceedings, 1984 University of Illinois Auditing Research Symposium, A. R. Abdel-Khalik and I. Solomon, eds.
- Wright, A. and R. H. Ashton, "Identifying Audit Adjustments with Attention-Directing Procedures," The Accounting Review, October, 1989: 710-728.
- Zimmer, A. "Verbal vs. Numerical Processing of Subjective Probabilities," in Decision Making Under Uncertainty, R. W. Scholz (editor), Elsevier Science Publishers (North-Holland), (1983).
- _____, "A Model for the Interpretation of Verbal Prediction," Journal of Man-Machine Studies, (1984), Vol. 20: 121-134.



UNIVERSITY OF ILLINOIS-URBANA



3 0112 039364002