THE ACCOUNTING REVIEW
Vol. 65, No. 1
January 1990
pp. 72–92

Experience Effects in Auditing: The Role of Task-Specific Knowledge

Sarah E. Bonner
University of Colorado at Boulder

ABSTRACT: Previous studies concerning experience effects in audit judgments have produced mixed results, possibly because they did not consider the knowledge necessary to complete the task and when it would normally be acquired. Further, many studies did not view the global judgment process as consisting of several components, e.g., cue selection. Task-specific knowledge may aid the performance of experienced auditors more in some components than in others. Not considering task-specific knowledge or viewing the judgment process as being comprised of components may have led to certain problems in generalizing the results of these studies to other auditing tasks. Those problems are addressed in the design of this study, which examines experience effects, specifically the role of task-specific knowledge, in the cue selection and cue weighting components of two audit tasks, analytical risk assessment and control risk assessment. Results indicate that task-specific knowledge aided the performance of experienced auditors in both the cue selection and cue weighting components only in analytical risk assessment.

MEASUREMENT of audit judgment performance is often difficult because there are no objective performance criteria for many audit tasks. As a result, the judgments of experienced auditors have been used as a substitute for other performance measures in determining firm policies and auditing standards. To determine the validity of this criterion, more evidence is needed

This paper is based on my dissertation at The University of Michigan. I appreciate the assistance of my committee members: Robert Libby, Anant Kshirsagar, Frank Yates, and David Wright. I particularly appreciate the comments and support of my chairman, Robert Libby. I would also like to thank two anonymous referees for comments on this version of the paper, as well as the following persons for comments on earlier versions: Matt Anderson, David Frederick, Ed Joyce, Barry Lewis, Bill Waller, and Mark Young. Financial support was provided by Deloitte Haskins & Sells and The University of Michigan.

Manuscript received October 1988.
Revisions received March 1989 and July 1989.
Accepted August 1989.

regarding why experienced auditors can perform tasks that inexperienced auditors cannot. Also, auditors with differing levels of experience are normally assigned to different tasks. Understanding the basis for these resource allocation decisions requires understanding of the nature of experience-related performance differences and the knowledge differences that underlie them (Libby and Frederick 1989).

Evidence regarding the effects of experience on audit judgments is somewhat mixed (see Wright 1988 for a review). As noted by Frederick and Libby (1986), these mixed results may have arisen because some studies did not consider the knowledge required to perform the experimental tasks, when that knowledge would be acquired (or the experience level of auditors who normally perform the task), and how this knowledge would be brought to bear upon the task. Further, many studies did not view the judgment process as being composed of several components, specifically cue selection, cue measurement, and cue weighting and combination (Lewis et al. 1983). Task-specific knowledge may aid the performance of experienced auditors in some components of judgments and not in others.

Not considering task-specific knowledge in general or its effects on various components of judgments has led to at least four problems in generalizing the results of these studies. First, some studies finding no experience effects used experimental tasks for which the designated experienced and inexperienced auditors both possessed the requisite knowledge (e.g., Ashton and Brown 1980; Hamilton and Wright 1982). Second, some studies may not have obtained experience effects because the experimental tasks did not contain the components in which knowledge acquired through experience would most aid performance, e.g., cue selection (Ashton and Brown 1980; Hamilton and Wright 1982). Third, studies finding a main effect for experience in one task may have demonstrated something other than task-specific knowledge differences, such as superior ability of experienced auditors at all tasks (e.g., Moriarity 1979). Finally, other studies with main effects for experience in multiple tasks may have confounded knowledge differences and task differences (Abdolmohammadi and Wright 1987).

To reduce the problems of generalizing the results described above, two experiments examine experience effects in, respectively, the cue selection and cue weighting components of two audit tasks, analytical procedure risk assessment and control risk assessment.[1] By addressing these issues, this study provides further understanding of the nature of experience effects in auditing, particularly the role of task-specific knowledge in general and its effect on performance in two components of these audit tasks. It also presents some preliminary evidence on how auditors make assessments of analytical procedure risk.

The next section reviews the literature on experience effects and describes

[1] The audit risk model provides a framework for planning audit tests, as follows (AICPA 1983). It includes components for control risk, or risk that the accounting control system will fail to prevent or detect a material error, if one occurs and analytical procedure risk, or risk that analytical procedures will fail to detect a material error, if one occurs and is not detected by the control system. Control risk is a function of the attributes of the control system. Analytical risk is dependent on how effective the analytical procedures will be at detecting errors.

the approach taken to study experience effects. Following this, the design and results of the two experiments are presented. A final section discusses the study's implications and limitations.

Problem Development

Experience Effects in Audit Judgments

As stated above, many previous studies on experience effects in auditing have not considered the role of task-specific knowledge, nor how task-specific knowledge can affect performance in various components of judgments. This has led to at least four problems in generalizing their results. First, some studies chose an experimental task for which the designated experienced and inexperienced auditors both possessed the knowledge necessary to perform the task, leading to findings of no experience effects. For example, Ashton and Brown (1980) divided subjects at one year, two years, and three or more years in their analysis of experience effects in evaluating payroll control system quality. There were no significant correlations between experience and multiple performance measures, such as judgment variance explained. One reason for high agreement among auditors across experience levels was that they put most weight on the separation-of-duties cues. The concept of separation of duties is one of the first taught in auditing courses and firm training. Thus, it is likely that auditors with one to three years of experience would all have had knowledge of the relative importance of those cues. Hamilton and Wright (1982) found similar results using experience levels of less than three years and greater than three years.

Second, some studies used tasks where there should be experience-related knowledge differences for the global judgment, but may have found no experience effects because the experimental tasks omitted components where knowledge would most aid performance. For example, Ashton and Brown (1980) preselected and premeasured cues for their subjects. As noted above, it is likely that inexperienced auditors would have already learned the relative importance of various controls over payroll and, therefore, have been able to perform as well as experienced auditors in judging control strength based only on weighting and combining cues. However, psychology literature (e.g., Johnson et al. 1981) predicts that inexperienced auditors would have difficulty in cue selection, so that there might be experience-related performance differences if all tasks components were to be performed. Similar situations may have occurred in Hamilton and Wright (1982) and Gaumnitz et al. (1982). Furthermore, in a recent study, partners and managers of several large firms provided mean ratings between 1.3 and 2.9 years for years of experience required to perform internal control tasks (Abdolmohammadi 1989). These ratings suggest that experience levels chosen by, for example, Hamilton and Wright (1982) are appropriate for internal control tasks which include all components.

Third, studies that found experience effects in one task may have demonstrated something other than experience-related knowledge differences. For example, Moriarity (1979) found better performance by experienced auditors at bankruptcy prediction; this may have simply reflected superior ability of experienced auditors at all tasks. In fact, in that study, experienced auditors did better than students at predicting bankruptcy with two presentation formats. Thus, it may be necessary to study experience effects by varying both experience levels and tasks.

A few studies have varied both tasks and experience levels. Butt (1988) compared experienced and inexperienced auditors' performance at judging the frequency of financial statement errors and vocabulary words. Experienced auditors were better than inexperienced auditors at judging the relative frequency of financial statement errors, but there were no experience effects for words. Marchant (1987) compared the ability of experienced and inexperienced auditors to use analogical reasoning in generating a potential explanation for an unexpected fluctuation encountered during analytical review. Because reasoning ability differences could have been confounded with knowledge differences, subjects were also asked to complete a general analogies test. His results indicated experience effects in the auditing task, but not in the general analogies task. Although these studies controlled for the effects of experience-related differences other than knowledge, they did not vary audit tasks, so that it is still not clear whether experienced auditors simply have superior ability at all *audit* tasks.

Two studies have varied audit tasks as well as experience levels. Abdolmohammadi and Wright (1987) used multiple audit tasks, that varied on "task complexity"[2] from "structured" to "unstructured," and varied the experience levels of auditors who performed those tasks. Although they considered when, in practice, auditors would normally perform the tasks in question, the mean experience level possessed by "inexperienced" and "experienced" auditors in each task was not constant across tasks. Their study, then, was an attempt to examine the main effect of experience separately in each of three tasks. They hypothesized and found a main effect in the "semi-structured" and "unstructured" tasks; because the judgment in the "structured" task was dependent on the judgment in the "semi-structured" task, the results for the "structured" task were not interpretable. The tasks, however, differed on dimensions other than the amount of structure, e.g., components. For example, the "semi-structured" and "unstructured" tasks required cue measurement, whereas the "structured" task did not. Furthermore, the experience levels that constituted experienced and inexperienced auditors may not have been appropriate in each task. Because there were multiple subject differences and task differences across tasks, it is difficult to determine whether main effects for experience were due to knowledge differences, other subject differences, task structure differences, other task differences (such as components), or some combination of the four.

Frederick and Libby (1986) compared experience effects across tasks which were the same except for the internal control weakness given as part of the background information; the experience levels for experienced and inexperienced auditors were also the same across tasks. They did not, then, have the potential confounding problems described above. They predicted and found that

[2] The terms "complexity" and "structure" are used to describe many different concepts. Abdolmohammadi and Wright (1987) specifically refer to the Simon (1960) model of decision processes, wherein structure refers to the extent of definition of the problem and cues, the number and specification of alternatives, and amount of judgment needed to make the decision, and complexity refers to amount of structure.

experienced auditors' knowledge of internal control weakness-financial statement error links led them to judge the probability of a conjunction of two errors to be higher than the probabilities of both constituent events (each error). Inexperienced auditors only judged the conjunction's probability to be higher than that of one of the constituent events. These judgment differences were found in both control weakness situations, demonstrating the role of task-specific knowledge in producing experience effects. Frederick and Libby addressed the issues of: (1) choosing appropriate experience levels to study experience effects for a specific task, (2) varying audit tasks as well as experience levels to learn about task-specific knowledge, and (3) holding task characteristics and experience levels constant across tasks so as not to confound task differences and knowledge differences. They did not exclude components from a task, as their subjects made only global judgments.

Approach for Studying Experience Effects

In this study, experience effects are investigated by varying experience levels and audit tasks, while holding task characteristics other than experience-related knowledge differences constant. Components in which knowledge acquired through experience is expected to aid performance (cue selection and cue weighting) are included in the tasks.

Two experience levels are used to determine the performance effects of experience-related knowledge differences. Two tasks with similar characteristics, but differing knowledge requirements (analytical procedure risk assessment and control risk assessment), are used to provide added controls for the effects of *subject* differences other than knowledge that might be related to experience, e.g., reasoning ability and personality (Shanteau 1984). As discussed above, if only one task is used, these other experience-related differences could produce a main effect for experience, and either disguise the lack of knowledge differences, or be confounded with the presence of knowledge differences. Here, however, since experienced and inexperienced auditors perform both tasks and since knowledge differences should be small in one task and large in the other, there should be an experience-task interaction. Other experience-related differences would create an interpretation problem only if they are expected to produce this same interaction. That is unlikely because factors such as general ability differences or personality differences would produce a main effect for experience across tasks. Thus, by comparing the experience-related performance differences across tasks that differ only on experience-related knowledge differences, most other explanations for experience effects are ruled out.

Analytical procedure risk assessment and control risk assessment were chosen because they meet the criteria specified above. That is, they have similar characteristics and there should be large experience-related knowledge differences about analytical risk, but small knowledge differences about control risk. Similar characteristics of the tasks include structure, components, and type of task. These tasks are both "semi-structured," as defined by Abdolmohammadi and Wright (1987). That is, they both have reasonably well-defined cues, a limited number of alternatives for output (e.g., low, medium, or high), and some judgment is needed.[3]

Further, the tasks have similar components. They both require cue selection, cue measurement, and cue weighting and combination. In the control risk task, an auditor selects the appropriate cues (specific controls for an account or transaction cycle), either from memory or using a memory aid such as an audit manual. He or she then collects auditee-specific measures on those cues and combines the measured cues to form a risk assessment. In the analytical risk task, an auditor considers what analytical procedures may be performed and selects the cues that are appropriate to analytical risk assessment (characteristics of the procedures). Again, these cues may be retrieved from memory or from an audit manual. He or she then determines (measures) how much of each characteristic the planned procedures possess and weights and combines the cues to make an assessment of analytical procedure risk.[4] Tasks with components as described above constitute "description" tasks (Smith 1988). Since the tasks here are similar on structure, components, and type of task, task differences and knowledge differences will not be confounded.

Finally, there should be large experience-related knowledge differences about analytical procedure risk assessment, but small knowledge differences about control risk assessment, specifically for the components of cue selection and cue weighting. In general, auditors acquire knowledge of relevant cues and how to weight them for judgment tasks by several means including collegiate auditing courses, audit firms' training programs, performance of the tasks in question, or by reviewing other auditors' performance of the tasks.

For the control risk task in this study, both experienced and inexperienced auditors[5] will have had college and audit firm training about the relevance and weighting of cues. Both groups of auditors will also have had experience evaluating controls. No predictions are made about experience-related performance differences in the control risk task alone, as those differences are not the focus of this study. It is possible that such differences exist, as experienced auditors may have had more firm training and certainly will have had more experience in reviewing others' control risk assessments. It is assumed, however, that because inexperienced auditors (as defined by this study) currently perform this task, their knowledge of the task is above some critical threshold and additional experience would not significantly improve their performance.

[3] These observations are based on a review of audit firm manuals and standard audit workpaper forms (Touche Ross 1978; Peat Marwick 1983; Ernst and Whinney 1986).

[4] These characterizations of the task are based on review of audit manuals and conversations with partners and managers. This does not imply that all auditors perform the tasks in this way. Anecdotally, in experiments conducted for Bonner (1988), subjects who had to measure cues in a judgment task seemed to first measure the cues, then combine them; that is, there were symbols such as "+" and "−" written on the instruments indicating that measurement was performed first, then combination.

[5] An experienced auditor is one who has training and extensive experience in both control risk and analytical risk assessments; such experience is held by managers (auditors with about five to seven years of experience). An inexperienced auditor is one who has extensive training and some experience in control risk assessment and minimal to no training or experience in analytical risk assessment. Staff auditors with two years of experience meet this criterion. These definitions are used to be consistent with the approach used here for studying experience effects.

For the analytical risk task, experienced and inexperienced auditors will have had similar collegiate training. Experienced auditors are expected to have had significantly more firm training than inexperienced auditors, as well as significantly more experience in both the assessment of analytical procedure risk and the review of others' assessments of analytical risk. Thus, experienced auditors are expected to have more complete knowledge of this task. Data collected on postexperimental questionnaires confirmed these expectations regarding training and experience for both tasks.

Experience Effects in Cue Selection and Cue Weighting

The components of the control risk and analytical procedure risk tasks to be studied are cue selection and cue weighting. Cue selection is included, as many psychology researchers agree that experience is important for superior performance in cue selection (e.g., Einhorn 1974; Dawes 1979). Yet, little empirical work has been conducted to support that idea. Cue weighting is studied as there are mixed results and opinions in both psychology and accounting as to the importance of experience therein (e.g., Slovic 1969; Dawes 1979).

Although no psychology research has directly addressed how experienced and inexperienced subjects differ in the retrieval and selection of cues, there are several related studies. The hypothesis generation literature (e.g., Johnson et al. 1981) concludes that inexperienced subjects have difficulty at the cue selection stage because they have incomplete knowledge bases. As Slovic et al. (1971, 18) note, "much of the inaccuracy of judgments can be traced to . . . the failure to recognize the relevant diagnostic signs." Dawes (1979, 394) states that a "model cannot replace the expert in deciding such things as 'what to look for'," and Einhorn (1974) included identification of relevant cues among the components of tasks he believed to be aided by experience. Johnson et al. (1984) also note that the ability to recognize relevant cues may be one of the hallmarks of superior performance by experienced subjects.

As discussed previously, experienced auditors should have much more knowledge than inexperienced auditors about the relevant analytical procedure risk cues via their training and performance of the task, and only slightly more knowledge about relevant control risk cues. Any current memory model (e.g., Anderson 1983) would predict that judges who have more complete knowledge (a simple content difference) would be better able to recognize relevant cues than those who have less complete knowledge and, thus, perform better at cue selection. Therefore, the following hypothesis is suggested:

H1: The difference between experienced and inexperienced auditors' cue selection performance will be larger in the analytical procedure risk task than in the control risk task.

Although extensive research has addressed the issue of cue weighting (see Libby 1981 for a review), very little research has investigated experience-related differences in cue weighting. Johnson et al. (1981) found that experienced clinicians agreed more about cue weights than did medical students. Slovic et al. (1972) documented such differences, but did not describe their nature.

Furthermore, the existing literature contains conflicting opinions on whether experience aids the judge in cue weighting. Some researchers hypothesize that experience is necessary for cue weighting. For example, Einhorn (1974) notes that agreement by experts is partly a matter of agreement on cue weighting and combination. Slovic (1969) notes that part of an experienced professional's success is due to his or her ability to weight cues differentially according to their relative importance.

There are others, though, who imply that experience is not necessary for superior performance in cue weighting. Shanteau (1984) suggests that, once given the relevant cues, inexperienced subjects may be able to act like experienced subjects. Dawes (1979) and Dawes and Corrigan (1974) suggest that experience is crucial only in the cue selection component. Some research also suggests that, when no criterion is available, equal weighting schemes may be optimal (e.g., Einhorn and Hogarth 1975); this would imply that knowledge gained through experience would not be necessary.

In accounting, Wright (1988) suggests that relative weighting of cues may depend on experience. Most researchers (e.g., Ashton 1974; Ashton and Brown 1980; Gaumnitz et al. 1982; Hamilton and Wright 1982; Abdolmohammadi and Wright 1987) have examined experience-related differences in cue weighting for internal control evaluations, and found few such differences. Nevertheless, results about the effects of experience on cue weighting are inconclusive. It is, therefore, difficult to hypothesize the effects of experience without considering the unique characteristics of the tasks to be studied.

In order to learn the relevant cues and their weights in the analytical risk task, the auditor needs experience performing analytical procedures, so that he or she may determine the diagnosticities of the various attributes (cues) of those procedures for detecting errors. To determine diagnosticities of cues and, consequently, their relative weights, he or she must have performed procedures in various situations over time. Although inexperienced auditors occasionally perform or assist with analytical procedures, they will not have experienced a wide variety of situations in which those procedures are performed. Furthermore, more experienced auditors usually hypothesize causes of these changes (Libby 1985), and evaluate those hypotheses based on evidence the inexperienced auditor may gather for them. As such, inexperienced auditors may have never received feedback on whether an error existed and, if so, which cues were most important in detecting the error and lowering analytical risk. The difference between experienced and inexperienced auditors' knowledge about cue weighting in analytical risk assessment should be fairly large. This suggests the following hypothesis:

H2: The difference between experienced and inexperienced auditors' cue weighting performance will be larger in analytical procedure risk assessment than in control risk assessment.

The next sections describe the two experiments designed to examine experience effects in one aspect of cue selection and cue weighting.

Overview of Experimental Methods

Each experiment used a 2 × 2 design with two tasks (control risk assessment and analytical procedure risk assessment), and two experience levels of auditors.

Analysis of experience effects in each experiment was done by comparing the experienced auditor-inexperienced auditor performance difference in one task to that performance difference in the second task. The purpose of Experiment 1 was to determine how experience affects cue selection. The purpose of Experiment 2 was to determine how experience affects cue weighting. Participants in both experiments were inexperienced and experienced auditors from two national accounting firms[6] who were attending, respectively, second-year staff and manager training schools (see fn. 5).

In each experiment, subjects were randomly assigned to perform the control risk or analytical risk task. In each condition in Experiment 1, each auditor was presented with a booklet including instructions, background information, the knowledge test, and a postexperimental questionnaire. Subjects in Experiment 2 received the same background information, the judgment task, and a postexperimental questionnaire. Postexperimental questionnaires collected data on collegiate training, firm training, experience in assessment, and experience in reviewing others' assessments for control risk or analytical risk.

Subjects in both experiments received brief verbal instructions.[7] The experiments were administered during the training programs and subjects were not allowed to use reference materials or confer with one another. Subjects required about 10 minutes to complete Experiment 1 and 30 minutes to complete Experiment 2.

Experiment 1—Cue Selection

Subjects

In Experiment 1, participants were 38 inexperienced auditors, 25 from Firm 1 and 13 from Firm 2, with an average of 2.10 years of experience, and 28 experienced auditors, 12 from Firm 1 and 16 from Firm 2, with an average of 6.72 years of experience.

Method

To test for the effects of experience on cue selection, experience level and audit task were combined in a 2×2 factorial design. Subjects in each audit task condition received background information for a small manufacturing company, including a brief description of the company's size, ownership, and business. Following this were instructions describing and defining the risk assessments that were to be made during audit planning and an indication as to which risk assessment the subject was to consider.

Subjects then received lists of 16 cues, obtained from professional sources, that were relevant (targets) or irrelevant (distractors) to one of the two risk assessments. They were asked to circle either "yes" or "no" for each cue, to indicate

[6] The firms are both in the medium range on the dimension of firm structure (Kinney 1986). They were chosen to be similar on that dimension so that firm structure could not account for any firm differences in results.

[7] In one administration of both experiments for a small number of experienced auditors from Firm 2, the researcher's verbal instructions were given by tape recording, not in person. The instructors who were present at that administration also reemphasized the instructions verbally.

whether they would or would not consider that factor to be relevant to the risk assessment in question. In the control risk condition, target (relevant) cues were controls over sales and receivables (Ernst and Whinney 1986; Robertson and Davis 1988, chap. 13), and distractor (irrelevant) cues were environmental considerations which would only affect the assessment of inherent risk for sales and receivables (Ernst and Whinney 1986; Dhar et al. 1988).[8] As an example, one target cue was "independent accounting for the serial order of billings." One distractor cue was "seasonality of sales." The Appendix contains the complete list of cues used in the control risk condition.

In the analytical procedure risk task, target cues were factors that an auditor would consider in assessing analytical risk (Ernst and Whinney 1986; Guy and Alderman 1987; AICPA 1988). Distractor cues in the analytical risk condition were sampling attributes that are used to limit test of details risk (AICPA 1981). A sample target was "number of independent sources of data used to develop expectations for sales/receivables"; a distractor was "method of projecting sample error to population." The Appendix contains all cues used in this condition.[9]

In both treatments, distractor cues were selected from the same category of variables as the target cues. Targets and distractors in the control risk task were client-specific variables, whereas targets and distractors in the analytical procedure risk task were auditor choice variables. As a result, distractors differed across tasks. Although having constant distractors across tasks would have been desirable, the relation of distractors to targets would have then differed across tasks. In the control risk task, both targets and distractors would have been client-specific variables; while in the analytical risk task, targets would have been auditor variables and distractors would have been client-specific variables. Because it was not clear how these differing relations of targets and distractors might interact with experience, the sacrifice in internal validity from using different distractors was thought to be less than the sacrifice from using constant distractors. Furthermore, the use of distractors from the same category as targets was thought to be a better test of the effects of knowledge gained through experience. Inexperienced auditors might be able to distinguish between items from different categories simply on surface features of those items, e.g., wording which would indicate that the items were either client-specific or auditor variables. Knowledge acquired through experience should allow auditors to distinguish within categories based on more substantive features of the items (Chi et al. 1982).[10]

[8] Although there are some factors which may be relevant to both control risk and inherent risk assessments, the cues used here were carefully chosen to be relevant to only one risk assessment.

[9] Pilot tests also verified the relevance of targets and the irrelevance of distractors. Participants in these pilot tests were experienced auditors at the two firms which participated in the study and at a third firm which did not participate.

[10] Subjects were not formally required to select the cues that they would use in making their judgments (cf. Abdel-khalik and El-Sheshai 1980), but instead classified them based on their relevance to the task. Knowledge of relevance is a necessary condition for proper cue selection. Though it appears unlikely, to the degree that subjects would not select cues judged to be relevant and would select those judged to be irrelevant, the results of this task would not generalize to cue selection. If this pattern emerged differentially across the two tasks, the conclusions would be improper.

Table 1
Experiment 1—Analysis of Variance: Firm × Experience Level × Task with Cue Selection Accuracy Scores as the Dependent Variable

Source	*Sum of Squares*	*df*	*Mean Square*	*F-statistic*	*Probability*
Firm	46.87	1	46.87	6.84	0.011
Experience Level	0.06	1	0.06	0.01	0.924
Task[1]	260.95	1	260.95	38.11	0.000
Firm × Level	15.55	1	15.55	2.27	0.137
Firm × Task	0.21	1	0.21	0.03	0.861
Level × Task	2.64	1	2.64	0.38	0.537
Firm × Level × Task	23.57	1	23.57	3.44	0.069
Error	397.18	58	6.85		

[1] Control Risk assessment and Analytical Procedure Risk assessment.

Results and Discussion

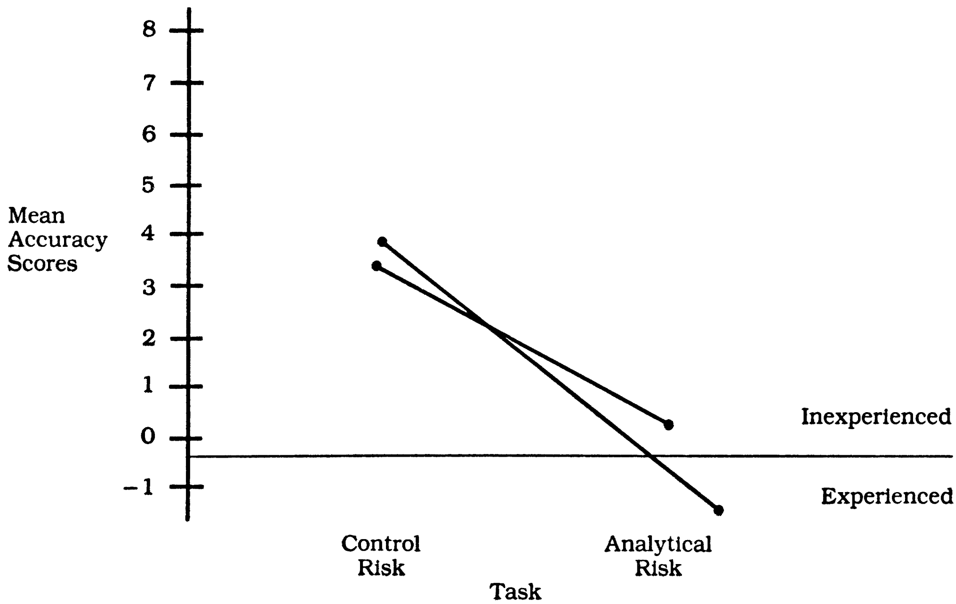
To test Hypothesis 1, a 2×2×2 ANOVA was estimated with firm, experience level, and task as between-subjects factors. The dependent variable for cue selection was an accuracy score, which was computed as the total number of targets for which "yes" was circled minus the number of distractors for which "yes" was circled, i.e., the number of targets answered correctly minus the number of distractors answered incorrectly. The maximum possible score, then, was eight, as there were eight targets and eight distractors (where the best possible performance was zero incorrect). That is, subjects were given credit for recognizing targets as being relevant. However, because distractors were from the same category of variables as targets and, thus, difficult to detect without appropriate knowledge, subjects were penalized for believing distractors to be relevant. This penalty is consistent with recognition performance measures used in psychology studies where the distractors were difficult to recognize (Gillund and Shiffrin 1984).[11]

Results of the analysis are presented in Table 1. As shown in Table 1, the existence of a marginally significant firm × level × task interaction required that the hypothesized level × task interaction be examined separately by firm. This ANOVA was run again to test for simple effects (level × task interactions) by firm. For Firm 1, the only effect was that accuracy scores were significantly different between tasks ($F=21.28$; $p=0.000$); this occurred because scores were higher in the control risk task than in the analytical procedure risk task, for both experienced and inexperienced auditors.

The analysis of simple effects for Firm 2 found the hypothesized interaction, i.e., the difference between experienced and inexperienced auditors' mean accuracy scores in analytical risk was significantly larger than the difference in control risk accuracy scores ($F=2.91$; $p=0.047$). Analyses of simple effects

[11] Further, note that this accuracy score equally weights Type I and Type II errors, as the lack of a definitive relationship between these errors and audit effectiveness makes it difficult to specify the costs attached to each type of error.

Figure 1
Experiment 1—Mean Accuracy Scores for Firm 1 for Control Risk and Analytical Risk Cue Selection

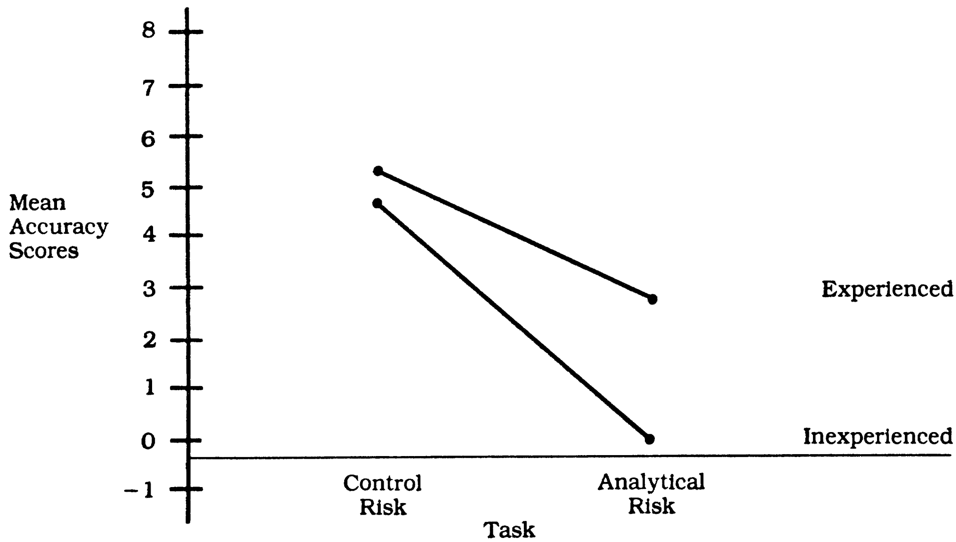


revealed that Firm 1's inexperienced and experienced auditors did not differ significantly in either control risk accuracy scores ($p=0.929$) or analytical risk scores ($p=0.203$). Firm 2's inexperienced and experienced auditors did not differ on control risk scores ($p=0.628$), but experienced auditors performed significantly better than inexperienced auditors on analytical risk scores ($p=0.008$), thus creating the experience level-task interaction for Firm 2. These firm differences are depicted in Figures 1 and 2.

Experienced auditors' accuracy scores were compared across firms, and simple effects were calculated for the control risk task and the analytical procedure risk task. Firm 2's experienced auditors performed significantly better than Firm 1's experienced auditors in analytical risk ($p=0.016$), but their performance in control risk was not significantly different ($p=0.325$). Firm 1 and Firm 2 inexperienced auditors did not differ on either analytical risk ($p=0.779$) or control risk ($p=0.130$) scores. The difference between Firm 1 and Firm 2, then, appears to have occurred only in the experienced auditors' performance in analytical risk. Furthermore, the difference in performance was due only to performance on the distractors. Firm 1's experienced auditors considered many more distractors to be relevant than did Firm 2's ($p=0.005$) but they performed equally well at recognizing targets as being relevant.

Hypothesis 1 was confirmed for Firm 2, but not for Firm 1. A review of the firms' audit manuals indicated that Firm 2 uses a form to assess analytical procedure risk, whereas Firm 1 does not. Firm 2 also has extensive instructions on

Figure 2
Experiment 1—Mean Accuracy Scores for Firm 2 for Control Risk and Analytical Risk Cue Selection



how to use this form, including which cues are relevant, to appropriately assess analytical risk, whereas Firm 1 has less extensive instructions in this area. Thus, results for Hypothesis 1 are consistent with the idea that training and experience in a task creates task-specific knowledge of relevant cues which can aid in cue selection.

Experiment 2—Cue Weighting

Subjects

Participants in Experiment 2 were 41 inexperienced auditors and 23 experienced auditors who did not take part in Experiment 1. The inexperienced auditors, 26 from Firm 1 and 15 from Firm 2, had an average of 1.91 years of experience; the experienced auditors, 8 from Firm 1 and 15 from Firm 2, had an average of 5.72 years of experience.

Method

To examine the effects of experience on cue weighting agreement, Experiment 2 used a 2×2 design. Experience level and task were between-subjects variables as in Experiment 1 (experienced auditor-inexperienced auditor and analytical risk-control risk). Each subject analyzed 16 cases for a single task (control risk or analytical procedure risk). The 16 cases represented a 1/2 replicate of a 2^5 orthogonal design.[12]

[12] The eight relevant cues used in Experiment 1 were combined into five cues for Experiment 2. The combination was made to reduce the number of cases a subject had to complete, and because some

Each case contained five relevant cues. In both the control risk and analytical risk conditions, cues indicated either a value of "high level" or "low level" of the attribute of the control system or analytical procedure. Based on the five cues in each case, subjects made the appropriate risk assessment on a nine-point scale. The scales were labeled from "high risk" to "low risk." The postexperimental questionnaires were the same as in Experiment 1. As examples, one of the control risk cases and an analytical procedure risk case are shown in the Appendix.

Results and Discussion

Cue weighting agreement. To assess the effect of experience on cue weighting and the effect of cue weighting on judgment agreement, the social judgment theory version of the lens model was used (e.g., Hammond et al. 1975). This model assesses agreement between two judges as being a function of their agreement on cue weighting and each individual's consistency and configurality. The model is:

$$r_a = G \times R_1 \times R_2 + C(1 - R_1)(1 - R_2),$$

where:

r_a = correlation between subject 1 and 2's judgments (here, risk judgments),
G = correlation between predictions based on models of subject 1 and 2's judgments, or the "matching index,"
R_i = consistency of subject i in applying his or her judgment policy (from a linear regression of the subject's judgments on the five cues), and
C = correlation between the residuals from subject 1 and 2's models (configurality or some other form of nonlinearity).

Most audit judgment studies have computed r_a, which includes the effects of cue weighting, consistency, and configurality (Wright 1988). Since G represents agreement on cue weighting alone, it provides more information about experience effects in this component of global judgments.

A measure of agreement (consensus) is being used here as there is no criterion for accuracy of cue weights. Although some researchers suggest that consensus may be a poor surrogate for accuracy (e.g., Wright 1988), others believe consensus is a good measure of "expertise" (Shanteau 1984). In fact, Ashton (1985) found high correlations between consensus measures and accuracy measures in two accounting studies. Furthermore, as Libby (1981, 31) notes, "consensus judgments provide the backbone for much of accounting practice," since many accounting and auditing tasks have no objective criterion. Also, training and decision aids are designed to promote consensus, and consensus is an often used defense in litigation.

G was calculated for each pair of experienced auditors and each pair of inexperienced auditors within each task by estimating a linear regression of each subject's judgments on the five cues, then correlating each pair of predicted values

---

of the cues were highly redundant. A 1/2 replicate was used here to prevent subject fatigue; however, interactions among cues were confounded with main effects. Since interactions among cues were not of interest here, the confounding of interactions did not affect either the dependent variables or the interaction between experience level and task, which was the effect of interest.

Table 2
Experiment 2—Analysis of Variance: Experience Level × Task with Matching Indexes (G, or Agreement on Cue Weighting) as the Dependent Variable

Source	*Sum of Squares*	*df*	*Mean Square*	*F-statistic*	*Probability*
Experience Level	3.60	1	3.60	13.84	0.000
Task	7.09	1	7.09	27.25	0.000
Level × Task	1.74	1	1.74	6.68	0.010
Error	134.40	517	0.26		

Cell Means

Experience Level	*Task*	*Mean*
Inexperienced	CR	0.468
Inexperienced	AR	0.201
Experienced	CR	0.735
Experienced	AR	0.743

CR = Control risk
AR = Analytical risk

from those regressions. These matching indexes were the dependent variables in a 2×2×2 ANOVA, where firm, experience level, and audit task were between-subjects factors. In this analysis, only within-firm *G*'s were used.

Since this ANOVA indicated that there was no firm × level × task interaction[13] ($p=0.874$) and since the hypothesis of interest was the level × task interaction, a 2×2 ANOVA was run with level and task as between-subjects factors and using all *G*'s, including both within-firm and across-firm *G*'s. The results are presented in Table 2. The level × task interaction occurred because the difference between experienced and inexperienced auditors' matching indexes was larger in the analytical risk task than in the control risk task ($p=0.010$).

In this analysis, then, Hypothesis 2 was confirmed, suggesting that experience provides better agreement for cue weighting in analytical risk assessment.[14] This result does not, however, lead to the conclusion that experience provides better agreement in the overall analytical risk assessment, as experience-related differences in consistency (*R*) or configurality (*C*) may offset experience-related cue weighting differences.

Effect of cue weighting agreement on judgment agreement. To assess the effect of cue weighting agreement on judgment agreement, r_a and *C* were calcu-

[13] The absence of a firm effect here is consistent with the results for targets and distractors in Experiment 1. Experienced auditors performed equally well at identifying targets as relevant. As only targets are used in this experiment, no firm difference is expected.

[14] When asked to rank the eight relevant cues from Experiment 1, i.e., provide a surrogate for subjective cue weights, experienced and inexperienced auditors did not differ on ranking agreement in either task (Bonner 1988). These results are feasible since subjects are unable to provide subjective weights which correspond to their objective weights (e.g., Summers et al. 1970).

Table 3
Experiment 2—Mean Lens Model Statistics

	r_a[1]	G	R	C
Experienced—Analytical Risk	.634	.743	.928	−.055
Inexperienced—Analytical Risk	.171	.201	.874	−.019
Experienced—Control Risk	.655	.735	.909	.284
Inexperienced—Control Risk	.442	.468	.837	.099

[1] r_a=correlations based on subjects' risk judgments.
G=correlations based on subjects' models.
R=consistency of subjects in applying their own models.
C=correlations based on model residuals.

lated for each pair of experienced auditors and inexperienced auditors, within each task. R was also calculated for each of these subjects. The mean for each of these statistics, as well as G, is presented for each treatment in Table 3. As shown in Table 3, G and R accounted for most of the magnitude of r_a, as the term $C(1-R_1)(1-R_2)$ was quite small on average. This result is consistent with early studies of judgments which had criterion variables (e.g., Hammond et al. 1964); again, there is some evidence in accounting that consensus and accuracy are related (Ashton 1985).

To further explore the effects of the various components of the model on r_a, 2×2 ANOVAs were estimated with experience level and task as between-subjects factors for r_a, R, and C as dependent variables. The results are as follows. There was a significant level×task interaction for r_a ($p=0.002$), as well as for C ($p=0.001$). There were no significant effects for R ($p=0.781$). The level×task interaction for C could not have created the level×task interaction for r_a, as the experienced auditor-inexperienced auditor difference was larger in control risk (.185) than in analytical risk (−.036) and, as mentioned above, the term $C(1-R_1)(1-R_2)$ was very small on average. Similarly, the level×task interaction in r_a could not have been caused by experience-related differences in R's, as there were none. As such the level×task interaction in r_a, where the experienced auditor-inexperienced auditor difference in analytical risk r_a's (.463) was much larger than that in control risk (.213), was mainly caused by the differences in cue weighting agreement.

The expected experience effect in cue weighting was found and, further, in a lens model judgment task, cue weighting was the factor which most accounted for experience effects in the global judgment. This finding is consistent with some psychology results (e.g., Johnson et al. 1981) and not others (e.g., Dawes 1979). It is unclear, then, whether task-specific knowledge differences consistently aid the performance of experienced auditors in cue weighting, or whether they can interact with certain task characteristics to create no differences in performance. Further, it is unclear whether, in all tasks, there are task-specific knowledge differences (with regard to cue weighting) between experienced and inexperienced auditors.

General Discussion

The purposes of this study were to investigate the role of task-specific knowledge in experience effects in audit judgments and how that task-specific knowledge affects the performance of experienced auditors in certain components of those judgments. The study also attempted to examine certain methodological questions raised by previous studies of experience effects in auditing. Results showed that experience aided auditors from only one of two firms in acquiring knowledge of relevant cues for cue selection in analytical procedure risk assessment. This may be attributable to differences in firm training as the firms were chosen to be similar on the dimension of firm structure. Experience aided auditors from both firms in agreement on cue weights; further, experience-related differences in cue weighting were the primary determinants of experience-related differences in overall judgment. The latter result was demonstrated by use of the social judgment theory version of the lens model, previously unused in accounting research. These results suggest that training and decision aids may be useful for both cue selection and cue weighting.

This study has several limitations. In Experiment 1, only one aspect of cue selection, knowledge of cue relevance, was studied. As discussed above, results from the knowledge task may not generalize to a true cue selection task. Generalizing results from the audit tasks examined to other audit tasks can also be difficult, especially if both task characteristics and subject characteristics affect performance. Further, not all components of the tasks were studied here. There may be components in which knowledge gained through experience does not aid performance.

Auditors from other firms may possess different characteristics than the subjects studied here, e.g., task-specific knowledge. Further, knowledge differences and judgment differences were indirectly linked; a full program of expertise research requires more investigation of their direct relationship, with careful consideration of firm differences and their effect on such subject characteristics as task-specific knowledge.

Appendix

Selected Experimental Materials

Experiment 1—Control Risk Cues

1. Introduction of new products this year (D)
2. Separation of credit approval from order entry and bad debt writeoffs (T)
3. Support for return credits with receiving reports or other appropriate documents (T)
4. Effect of general economic conditions on customers of Southern Mills (D)
5. Frequent aging of accounts and follow-up on delinquent accounts (T)
6. Composition of Southern's board of directors (D)
7. Extent of centralization of production (D)
8. Independent accounting for the serial order of billings (T)
9. Independence of detail accounts receivable recordkeeping from billing and shipping (T)

10. Seasonality of sales (D)
11. Southern's policies on control of product quality (D)
12. Proper approval on discounts and returns before issuance of a credit memo (T)
13. Customer dependence on Southern Mills as a source of supply (D)
14. Number of foreign customers (D)
15. Proper support for sales orders with approved customer lists and price and discount schedules (T)
16. Reconciliation of billings to original sales or shipping documents (T)

(T) = Target, or relevant cue
(D) = Distractor, or irrelevant cue

Experiment 1—Analytical Risk Cues

1. Expected number of errors in sales/receivables populations (D)
2. Number of independent sources of data used to develop expectations for sales/receivables (T)
3. Size of differences from expectations to be investigated (T)
4. Allowable risk of incorrect rejection (D)
5. Sophistication of method used to convert data into predicted sales and receivables (T)
6. Method of projecting sample error to population (D)
7. Size of sample (D)
8. Thoroughness of investigation concerning unusual sales transactions (T)
9. Level of detail of data used to develop expectations for sales and receivables (T)
10. Type of sample selection technique to be used (D)
11. Relationship of sample to audit objective (D)
12. Independence of data used to develop expectations for sales and receivables (T)
13. Method of evaluating sample results (D)
14. Allowable risk of incorrect acceptance (D)
15. Amount of work to be done to establish reliability of data used for expectations (T)
16. Planned reliance on client explanations for differences vs. our audit tests (T)

(T) = Target, or relevant cue
(D) = Distractor, or irrelevant cue

Experiment 2—Example of Control Risk Task

	Attribute of Control System	*Level of Attribute*
(1)	Segregation of sales and receivables duties	High level of segregation
(2)	Support and approval for credits for returns	High level of support and approval
(3)	Frequency of aging of accounts and follow-up on delinquent accounts	High level of aging frequency
(4)	Approval and support for sales orders	Low level of approval and support
(5)	Support for billings and independent accounting for sequence	Low level of support and independent accounting

Rating of control risk:

1	2	3	4	5	6	7	8	9
High risk Low assurance								Low risk High assurance

Experiment 2—Example of Analytical Risk Task

	Attribute of Procedures	*Level of Attribute*
(1)	Independence of data sources used to develop expectations for sales and receivables	High level of independence
(2)	Amount of detail of data used to develop expectations	High (large) amount of detail
(3)	Sophistication of method used to convert data into expectations	High level of sophistication
(4)	Amount of work done to establish reliability of data	Low (small) amount of work
(5)	Planned amount of investigation of differences from expectations	Low level of investigation of differences

Rating of analytical risk:

1	2	3	4	5	6	7	8	9
High risk Low assurance								Low risk High assurance

References

Abdel-khalik, A. R., and K. M. El-Sheshai. 1980. Information choice and utilization in an experiment on default prediction. *Journal of Accounting Research.* (Autumn): 325–342.

Abdolmohammadi, M. J. 1989. A taxonomy of audit task complexity for decision aid development. Working paper, Bentley College.

———, and A. Wright. 1987. An examination of the effects of experience and task complexity on audit judgments. *The Accounting Review.* (January): 1–13.

American Institute of Certified Public Accountants. 1981. *Statement on auditing standards no. 39: Audit sampling.* AICPA.

———. 1983. *Statement on auditing standards no. 47: Audit risk and materiality in conducting an audit.* AICPA.

———. 1988. *Statement on auditing standards no. 56: Analytical procedures.* AICPA.

Anderson, J. R. 1983. A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior.* (June): 261–295.

Ashton, A. H. 1985. Does consensus imply accuracy in accounting studies of decision making? *The Accounting Review.* (April): 173–185.

Ashton, R. H. 1974. An experimental study of internal control judgments. *Journal of Accounting Research.* (Spring): 143–157.

———, and P. R. Brown. 1980. Descriptive modeling of auditors' internal control judgments: Replication and extension. *Journal of Accounting Research.* (Spring): 269–277.

Bonner, S. E. 1988. Experience effects in the components of analytical risk assessment. Ph.D. diss., The University of Michigan.

Butt, J. L. 1988. Frequency judgments in an auditing-related task. *Journal of Accounting Research.* (Autumn): 315–330.

Chi, M. T. H., R. Glaser, and E. Rees. 1982. Expertise in problem solving. *Advances in the psychology of human intelligence.* Ed. R. J. Sternberg, 7–75. Lawrence Erlbaum Associates.

Dawes, R. M. 1979. The robust beauty of improper linear models in decision making. *American Psychologist.* (July): 571–582.

———, and B. Corrigan. 1974. Linear models in decision making. *Psychological Bulletin* (February): 95–106.

Dhar, V., B. Lewis, and J. Peters. 1988. A knowledge-based model of audit risk. *AI Magazine.* (Fall): 56–63.

Einhorn, H. J. 1974. Expert judgment: Some necessary conditions and an example. *Journal of Applied Psychology.* (October): 562–571.

———, and R. M. Hogarth. 1975. Unit weighting schemes for decision making. *Organizational Behavior and Human Performance.* (April): 171–192.

Ernst & Whinney. 1986. *Guide to the Ernst & Whinney audit approach.* Ernst & Whinney.

Frederick, D. M., and R. Libby. 1986. Expertise and auditors' judgments of conjunctive events. *Journal of Accounting Research.* (Autumn): 270–290.

Gaumnitz, B. R., T. R. Nunamaker, J. J. Surdick, and M. F. Thomas. 1982. Auditor consensus in internal control evaluation and audit program planning. *Journal of Accounting Research.* (Autumn, Part II): 745–755.

Gillund, G., and R. M. Shiffrin. 1984. A retrieval model for both recognition and recall. *Psychological Review.* (January): 1–67.

Guy, D. M., and C. W. Alderman. 1987. *Auditing.* Harcourt Brace Jovanovich.

Hamilton, R. E., and W. F. Wright. 1982. Internal control judgments and effects of experience: Replications and extensions. *Journal of Accounting Research.* (Autumn, Part II): 756–765.

Hammond, K. R., C. J. Hursch, and F. J. Todd. 1964. Analyzing the components of clinical inference. *Psychological Review.* (November): 438–456.

———, T. R. Stewart, B. Brehmer, and D. O. Steinmann. 1975. Social judgment theory. *Human judgment and decision processes.* Eds. M. F. Kaplan and S. Schwartz, 271–312. Academic Press, Inc.

Johnson, P. E., A. S. Duran, F. Hassebrock, J. Moller, M. Prietula, P. J. Feltovich, and D. B. Swanson. 1981. Expertise and error in diagnostic reasoning. *Cognitive Science.* (September): 235–283.

——, M. G. Johnson, and R. K. Little. 1984. Expertise in trial advocacy: Some considerations for inquiry into its nature and development. *Campbell Law Review.* (Fall): 119–143.

Kinney, W. R., Jr. 1986. Audit technology and preferences for auditing standards. *Journal of Accounting and Economics.* (March): 73–89.

Lewis, B., M. D. Shields, and S. M. Young. 1983. Evaluating human judgments and decision aids. *Journal of Accounting Research.* (Spring): 271–285.

Libby, R. 1981. *Accounting and human information processing: Theory and applications.* Prentice-Hall, Inc.

——. 1985. Availability and the generation of hypotheses in analytical review. *Journal of Accounting Research.* (Autumn): 648–667.

——, and D. M. Frederick. 1989. Expertise and the ability to explain audit findings. Working paper, University of Michigan.

Marchant, G. A. 1987. Analogical reasoning and error detection. Ph.D. diss., The University of Michigan.

Moriarity, S. 1979. Communicating financial information through multidimensional graphics. *Journal of Accounting Research.* (Spring): 205–224.

Peat, Marwick, Mitchell & Co. 1983. *KPMG Peat Marwick audit manual—U.S.* Peat, Marwick, Mitchell & Co.

Robertson, J. C., and F. G. Davis. 1988. *Auditing.* Business Publications, Inc.

Shanteau, J. 1984. Some unasked questions about the psychology of expert decision makers. *Proceedings of the 1984 IEEE conference on systems, man, and cybernetics.* Ed. M. E. El Hawary. IEEE.

Simon, H. 1960. *The New Science of Management.* Harper Row.

Slovic, P. 1969. Analyzing the expert judge: A descriptive study of a stockbroker's decision processes. *Journal of Applied Psychology.* (August): 255–263.

——, D. Fleissner, and W. S. Bauman. 1972. Analyzing the use of information in investment decision making: A methodological proposal. *Journal of Business.* (April): 283–301.

——, L. G. Rorer, and P. J. Hoffman. 1971. Analyzing use of diagnostic signs. *Investigative Radiology.* (January-February): 18–26.

Smith, G. F. 1988. Towards a heuristic theory of problem structuring. *Management Science.* (December): 1489–1506.

Summers, D. A., J. D. Taliaferro, and D. J. Fletcher. 1970. Subjective vs. objective description of judgment policy. *Psychonomic Science.* (24 February): 249–250.

Touche Ross & Co. 1978. *The Touche Ross audit process manual.* Touche Ross & Co.

Wright, W. F. 1988. Audit judgment consensus and experience. *Behavioral accounting research: A critical analysis.* Ed. K. R. Ferris, 305–328. Century VII Publishing Company.