Pergamon

Accounting, Organizations and Society, Vol. 22, No. 5, pp. 387-410, 1997
© 1997 Elsevier Science Ltd
All rights reserved. Printed in Great Britain
0361-3682/97 $17.00 + 0.00

PII: S0361-3682(96)00049-9

AUDIT CATEGORY KNOWLEDGE AS A PRECONDITION TO LEARNING FROM EXPERIENCE

S. E. BONNER
University of Southern California

R. LIBBY and M. W. NELSON
Cornell University

Abstract

Prior research indicates that inexperienced auditors lack knowledge of basic auditing categories (e.g. transaction cycles, audit objectives), instead developing this knowledge over time. As a consequence, learning from early experiences may be hampered because these experiences are not stored with respect to the category structures that are needed for important audit decisions. We performed an experiment which demonstrates that: (1) providing transaction cycle and audit objective category knowledge through instruction prior to experience facilitates one particular type of subsequent learning from experience (learning of category-level error frequencies), and (2) this learning advantage cannot be duplicated by providing listings and explanations of category memberships after experience. In addition, actual experience consequently has a greater influence on later audit decisions when category knowledge is acquired prior to experience. © 1997 Elsevier Science Ltd

Both accountants and psychologists have emphasized that category knowledge[1] has important direct effects on decision performance (e.g. hypothesis generation and evaluation; see Bédard & Chi, 1993; Libby & Luft, 1993). We suggest that having category knowledge in place prior to experiencing events may also improve decision performance indirectly by assisting in learning from those subsequent experiences. That is, category knowledge that organizes events may be a precondition to effective learning from later experienced events. This is a concern because research indicates that inexperienced auditors lack complete knowledge of basic auditing categories, instead developing this knowledge over time. As a consequence, learning from early experiences may be hampered because these experiences are not stored with respect to the category structures that are needed for important audit decisions.

Our experiment investigates whether providing inexperienced auditors with intensive instruction on basic auditing categories helps them learn from their post-instruction experiences and apply the knowledge acquired from those experiences when making later audit decisions. The first stage of the experiment provides novice subjects with intensive instruction on a small part of experienced auditors' category structure. Once accurate

The authors would like to thank KPMG Peat Marwick for supporting this project through a Research Opportunities in

[1]As used in this paper, "category knowledge" refers to classifications of items that auditors maintain in memory. For example, as shown in Table 2, transaction cycles and audit objectives are two types of categories that auditors can use to classify various financial statement errors they have encountered in the past.

learning of the category structure is assured, the second stage then tests the effects of this precondition on subsequent learning from experience in an abstract category frequency learning task. Category frequency learning was selected for this demonstration both because of its importance to audit decision making and because evidence that category frequency learning improved following category instruction would provide a clear indication that the instruction was effective in influencing subsequent learning from experience.

Our results indicate that instruction facilitated the acquisition of knowledge about transaction cycle and audit objective categories of errors (see Table 2 for the actual categories used in the experiment), and that having this basic category knowledge prior to experiencing errors improved learning of category-level error frequencies (i.e. increased the degree to which experienced category-level frequencies were reflected in subjects' frequency estimates). If subsequent learning of the appropriate category structure allowed auditors to reorganize and make use of their early experiences, the importance of this effect would be minimized. However, this learning advantage could not be duplicated by providing listings and explanations of category memberships after experiencing errors. In turn, pre-existing category knowledge increased the degree to which experienced category-level frequencies were reflected in audit decisions. The results have important implications for the timing of instruction, highlighting the usefulness of applying more effective methods to impart audit category knowledge that organizes important experiences before novices enter the field so they can better learn from their subsequent experiences (be they errors or other types of experiences).

The remainder of this paper is organized as follows. The next section examines existing research related to the issues at hand and develops hypotheses. The following section describes the method, procedures, and variables. The last two sections, respectively, present the results and provide concluding remarks.

RELATED LITERATURE AND HYPOTHESIS DEVELOPMENT

Libby and Luft (1993) and Libby (1995) provide a model of the antecedents and consequences of knowledge in which ability and experience affect the acquisition of knowledge, and ability and knowledge affect performance. This paper focuses on the linkages among experience, knowledge and performance. We first examine the effect of having one type of knowledge, category knowledge, in place prior to having the experiences which give rise to a second type of knowledge, category-level frequency knowledge. Specifically, we examine the effect of this pre-existing category knowledge on the relative accuracy of frequency knowledge gained from experience; we also examine the effect of category knowledge provided after experience on the accuracy of frequency knowledge. Previous research in accounting has not examined the effect of category knowledge (acquired either before or after experience) on the acquisition of other knowledge from experience. We then determine the extent to which effects observed on the experience-knowledge linkage extend to the knowledge-performance linkage.

Category-level error frequencies and audit decisions

Both audit firm guidelines and auditing standards support the usefulness of transaction cycle and audit objective categories for a variety of audit judgments and decisions. Previous research suggests that experienced auditors use category knowledge to access prior experiences when making judgments and decisions, and that inexperienced auditors do not. Audit judgments previously examined include the generation of hypotheses about, and the evaluation of, probabilities of financial statement errors (Libby & Frederick, 1990; Tubbs, 1992) and the evaluation of missing and ineffective controls (Weber, 1980; Frederick, 1991).

We examine another important role of category knowledge: as a structure that facilitates learning from experience. We focus on one type of audit knowledge that is learned from experience, knowledge of category-level error frequencies. Individual error frequencies in auditing have been acknowledged as important by prior archival studies (Coakley & Loebbecke, 1985; Hylas & Ashton, 1982; Wright & Ashton, 1989) and examined in prior experimental research (Butt, 1988; Libby, 1985; Libby & Frederick, 1990; Nelson, 1993; Waller, 1990). Category-level frequency knowledge is also relevant to many audit decisions (Nelson *et al.*, 1995; Bonner *et al.*, 1996).[2] For example, planning the allocation of audit effort to different transaction cycles should be influenced by (among other factors)[3] knowledge of the relative frequency of errors and irregularities in different cycle categories. If errors in the inventory cycle are more prevalent than those in other cycles, all other things equal, relatively more time would be allocated to auditing inventory.

Acquisition of error category knowledge in auditing

Research by Frederick *et al.* (1994) indicates that auditors develop their knowledge of membership of financial statement errors in cycle and objective categories over time. Audit managers in that study (with a median of five years' experience) categorized errors by cycle and objective violated based on underlying meaning, indicating that they had acquired knowledge of these categories. Auditing students sorted errors on the basis of similar wording (non-meaningful categories formed on the basis of surface attributes). Staff auditors (median of one year experience) exhibited characteristics of both the managers' and students' category knowledge, nearly matching the managers' categorization of errors by transaction cycle, but not by audit objective. Davis (1994), Libby and Frederick (1990) and Tubbs (1992) report consistent findings.

There are several reasons why categories of financial statement errors may not be learned well early in auditors' careers. First, there may be difficulty learning these categories from the instruction auditors receive. One reason for this is the manner in which category structures are specified in the authoritative literature. Frederick, *et al.* (1994) document that the category structures suggested in current auditing standards, and the texts and instruction materials which are based on those standards, are incomplete and/or inconsistent, particularly as they relate to audit objectives. Thus, inexperienced auditors might learn different sets of categories of errors in university instruction and early firm training, creating confusion. A second reason would be that students in auditing courses learn cycle and objective categories as they relate to a number of important concepts, including errors, internal controls, inherent risk factors, substantive tests, and so forth. Learning category memberships with respect to concepts other than errors may interfere with learning about the category membership of errors.

It also may be difficult to learn categories of financial statement errors through early experience in auditing, due to problems with both practice and feedback. New auditors tend to gain practice on the performance of tests of controls and substantive tests; this practice may

[2]Nelson *et al.* (1995) examine the effects of hierarchically organized knowledge structures on experienced auditors' ability to estimate conditional probabilities, and Bonner *et al.* (1996) examine the types of decision aids that can be used to mitigate those effects. Thus, these papers use experienced auditors who have developed enough category knowledge over time to allow them to learn frequency knowledge, and focus on relations between elements of the knowledge structures that influence conditional probability estimation. In contrast the current paper examines the importance to accurate frequency learning of students (proxying for novice auditors) having category knowledge in place prior to experiencing the errors from which they can acquire category frequency knowledge.

[3]Auditors should consider a number of other factors in making global time budget allocations, including the amount of time it takes to perform relevant procedures, the effectiveness of those procedures, and the liability exposure different types of potential errors present.

provide only indirect information about transaction cycle and audit objective categories of errors. Also, there is likely to be a lack of feedback in the audit environment (e.g., Ashton *et al.*, 1989; Ashton, 1991; Bonner & Pennington, 1991); lack of feedback can hamper category learning (Homa, 1984).[4] Finally, when developing categories through experience, novices tend to focus on surface (irrelevant) attributes of events rather than the deep (meaningful) attributes that experts use to categorize events (e.g. Adelson, 1984; Chase & Simon, 1973; Chi *et al.*, 1982). Because of these problems with the acquisition of knowledge about categories of errors through early instruction and experience, inexperienced auditors' early error experiences might not be stored in memory completely correctly with respect to the categories that they will eventually develop and employ in audit decisions, but rather with respect to incorrect categories that will eventually be replaced. This suggests that it might be useful to convey category knowledge about financial statement errors to novice auditors via more effective early instruction to aid subsequent learning from experience. In the next section, we discuss the issue of the effect of category knowledge on later learning from experience.

Effect of category knowledge on subsequent learning of category-level error frequencies from experience

Auditors experience the frequency of occurrence of financial statement errors through a variety of sources, including first-hand experience on audits, comments made by superiors on the audit team, discussions of other audits with associates, and case materials used in instruction (Butt, 1988; Libby, 1985).[5] Current evidence indicates that frequency information encountered through experience is stored indirectly in memory in the form of individual memory records ("memory traces") of the events experienced (see Nelson, 1994 for a review). At the time a frequency estimate is required, the numbers of memory traces associated with each type of event are accumulated and compared.

Early theory in psychology characterized this frequency learning as an automatic process which is insensitive to ability, motivation, and intention (see Hasher & Zacks, 1979, 1984 for reviews of the psychological theory and Libby, 1985 and Ashton, 1991 for discussions related to auditing). However, more recent research disputes the automaticity of frequency learning (see Nelson, 1994 for a review). Of particular interest is Barsalou and Ross (1986), who examined frequency learning relevant to two types of categories: "natural" categories (i.e. groupings that occur naturally in the real world and that people tend to group in the same way, e.g. birds and fruits) and "property" categories (e.g. round things, red things), which are based on dimensions that are known but not normally used for categorization. Barsalou and Ross found consistent sensitivity to natural category frequencies, but only found sensitivity to property category frequencies when subjects had been instructed in the categories related to the property dimension. Similarly, Alba *et al.* (1980) found sensitivity to the frequencies of natural categories, and Freund and Hasher (1989) found sensitivity to the frequencies of property categories primarily when subjects were instructed to categorize by property. The results of these studies imply that category-level frequencies are not retained with respect to every dimension on which an item could be categorized. Rather, the dimension must be one by which items are categorized prior to

[4]Auditors do obtain feedback through the review process, but this feedback may be delayed and/or neglect to specify the membership of errors in transaction cycle or audit objective categories.

[5]Frequencies can also be encountered in the form of summary statistics. This summary-form frequency knowledge tends to be learned less accurately and can be applied in fewer cognitive processes than can frequency knowledge acquired through experience (Butt, 1988; Nelson, 1994).

experiencing the items, implying that knowledge of categories needs to be in place prior to exposure. Therefore, category frequency learning effects are a particularly good indication that instruction has conveyed the category structure in a manner which influences how experiences are encoded in memory. Thus, results indicating category frequency learning effects are broader in implication, since improved encoding of experiences should also improve other category-related aspects of learning from experience (e.g. improved learning of the covariation of errors and control system attributes, improved ability to relate errors by audit objective which facilitates the identification of similarities in substantive testing approach, etc.).

However, important differences exist between the categories employed in prior psychology research and the audit categories of transaction cycles and audit objectives that might limit the generalizability of prior psychology studies to the audit setting. The categories used in prior psychological studies were existing and well-defined natural categories developed through experience; category-related instruction administered to subjects simply taught them to categorize on a known dimension not previously used for categorization. In contrast, audit instruction must convey previously unknown and ill-defined categories (e.g. audit objectives, transaction cycles). It is not clear that category knowledge conveyed via instruction in this manner will influence how future experiences are encoded in memory, and thus influence the accuracy with which frequencies are subsequently estimated. Supporting this concern, results in prior psychological studies of category-level frequency knowledge weaken as the categories become more conceptual and less concrete (e.g. Sherman *et al.*, 1992 find weaker results for categories like "introvert/extrovert" than they do for concrete objects like "men/women"). In sum, it is unclear that both knowledge of technical concepts and the categorical structure which organizes those concepts can be conveyed from scratch via instruction effectively enough to influence frequency estimation.

Our first hypothesis is related to these issues. We assert that subjects will better learn relative category frequencies from experience when they receive instruction that imparts knowledge of categories of errors prior to experiencing the financial statement errors. To test this assertion, we provided subjects with instruction concerning the relation between financial statement errors and either transaction cycle or audit objective categories, exposed subjects to a sequence of errors (with the number of times particular errors appeared in the sequence varied to manipulate the category-level error frequencies that subjects experienced), and elicited frequency estimates to measure frequency knowledge. Our hypothesis is:

> H1: Subjects receiving transaction cycle (audit objective) category instruction prior to experiencing frequencies will make frequency estimates which more closely reflect experienced error frequencies for transaction cycle (audit objective) categories than for audit objective (transaction cycle) categories.

Providing category knowledge after subjects experience category-level error frequencies

As noted previously, in current practice auditors learn transaction cycle and audit objective categories during their first few years of audit experience. Thus, inexperienced auditors cannot encode the error frequencies they experience with respect to correct categories at the time their early experiences occur. However, if they learn correct categories at some later point, these auditors may be able to estimate category-level frequencies by summing the frequencies of the individual errors that make up the category. Therefore, we also examine whether the effect suggested by hypothesis (H1) can be duplicated when subjects are presented with the definitions and listings and explanations of category memberships of errors after they experience the errors. This additional procedure examines the extent to which subjects' ability to estimate category-level frequencies is facilitated by having category knowledge prior to estimating those frequencies (but after experiencing frequencies), as opposed to having category knowledge prior to experiencing those frequencies (and thus available to aid the encoding of those experiences in memory as they occur).

Our results regarding this issue have implications both for the function that category knowledge serves with respect to frequency knowledge and for audit practice. Regarding the function of category knowledge, it is unclear whether possessing category knowledge facilitates:

1. the encoding of category-level information when errors are experienced, or
2. only the later estimation of category-level frequencies from prior experiences.

This additional test allows us to examine this issue. Also, if it were the case that category knowledge primarily facilitates estimation, the practical implication would be that auditors who acquire category knowledge through experience or who are provided with instruction or decision aids after experiencing errors will still be able to apply their early experiences to their frequency estimates. In other words, finding that what matters is only what auditors know and not when they know it would suggest more degrees of freedom concerning when category knowledge is learned.

In general, both early and more recent psychology research supports the importance of having an organizing framework prior to experience for learning from that experience (see Bonner & Walker, 1994 for a review). Because categories present an organizing framework which is important to learning category-level error frequencies from experience, we expect that receiving information about categories after experiencing error frequencies will not allow our subjects to estimate relative category-level frequencies for the previously untrained categories as well as they can for previously trained categories. That is, our second hypothesis (H2) asserts that the effect observed for H1 will not be eliminated by providing subjects with category information after experience. Our hypothesis is:

> H2: Providing subjects with category definitions and listings and explanations of category membership after they have experienced error frequencies does not eliminate the effect described by H1.

Effect of knowledge of category-level frequencies on audit decisions

A number of prior studies have demonstrated the impact of category knowledge on cognitive processes involved in audit decisions such as recall (e.g. Choo & Trotman, 1991; Frederick, 1991; Kida, 1984; Weber, 1980), but research is only beginning to examine the effects of category knowledge on later decisions (Nelson *et al.*, 1995). Also, previous research in auditing has shown a link between frequency knowledge and audit judgments. For example, Libby (1985) and Libby and Frederick (1990) demonstrate that financial statement errors judged to occur more frequently are more frequently generated as hypothesized explanations for ratio fluctuations in analytical review, and Waller (1990) provides evidence through verbal protocols that experienced auditors sometimes consider frequencies consciously when making audit judgments. However, these studies primarily examine the degree to which experienced auditors consider frequency information. Even when inexperienced auditors have been instructed to have the category structures necessary to discriminate category-level frequencies, they may not apply those frequencies in audit decision making. Further, they may be influenced by other factors such as salience of certain cycles or objectives from their previous coursework (e.g. accounting courses may tend to emphasize sales over other types of transactions).

Our third hypothesis asserts that the category-level frequencies subjects experience will have a greater influence on their audit decisions when they receive instruction concerning the relevant category structure prior to experiencing the financial statement errors. To test this hypothesis, we required subjects to perform the additional task of allocating hours of audit effort across transaction cycle and audit objective categories. Our hypothesis is:

H3: Subjects receiving transaction cycle (audit objective) category instruction prior to experiencing frequencies will make audit effort allocations which more closely reflect experienced error frequencies for transaction cycle (audit objective) categories than for audit objective (transaction cycle) categories.

METHOD

Subjects

This paper examines the importance to estimating frequencies of categories of events of having category knowledge in place prior to experiencing those events. As discussed previously, prior research indicates that auditors acquire this category knowledge gradually during their early years of audit experience. Consequently, it was necessary to obtain the participation of subjects who were not experienced enough in auditing to have developed this knowledge for themselves. While the best subject group might be newly hired auditors, we chose accounting students as our subject pool based on their availability, their knowledge of enough accounting to understand financial statement errors, and their lack of the audit experience necessary to have acquired audit category knowledge. To the extent that these subjects already had acquired audit category knowledge, the effects of our category instruction manipulation would be decreased (biasing against finding an effect). To the extent that newly hired auditors typically have acquired more audit category knowledge or other relevant knowledge than have our student subjects, a smaller effect of instruction would be anticipated in practice than would be indicated by our results. However, the poor performance of novice auditors in categorization tasks in prior research suggest that novice auditors in general possess low levels of the category knowledge we examine, similar to our subjects.

Subjects were 29 undergraduate accounting and 36 MBA students enrolled in intermediate accounting courses at two north-eastern universities. Subjects confirmed that they had not previously received instruction on transaction cycle or audit objective categories in their coursework.[6] The undergraduate volunteers for the experiment were compensated with course credit. The graduate student volunteers received a ticket to a lottery which paid two $100 prizes.[7]

Design and procedures

Overview. The experimental protocol was developed based on the frequency learning paradigm used by Butt (1989), measures of category knowledge used by Frederick *et al.* (1994), and error stimuli used by Nelson *et al.* (1995) and Bonner *et al.* (1996). Subjects completed all portions of the experiment on a Macintosh computer in a university laboratory. A computerized procedure was used to standardize the timing and presentation of instruction and learning trials in each phase of the experiment. Subjects were prohibited from using reference materials or conferring with others during all parts of the experiment.

Table 1 summarizes the experimental procedures. The experiment was conducted in two phases. In the first "category learning" phase, subjects completed a category knowledge pre-test, received instruction on either the transaction cycle or audit objective categories and reinforcement of that instruction, performed a distractor task, and completed a category knowledge post-test. After completing these tasks, they were instructed not to discuss the experiment with others and confirmed the

[6] One subject reported one course in auditing. Dropping this subject does not change the results. Also, eight of the 65 subjects indicated English as a second language, but dropping these subjects had no effect on the results. No other demographic variables differed across treatments or had an effect on the subjects' frequency knowledge and audit decisions.

[7] The difference in compensation reflected both differences in preferences of the subject groups and school human subject policies. No significant differences in subject group were found.

TABLE 1. Summary of experimental procedures

Phase	Sequence of procedures
Phase 1: Category learning	1. Pre-test category knowledge (sort task of nine financial statement errors)
	2. Instruction on either transaction cycles or audit objectives
	3. Reinforcement of instructed categories with 18 multiple-choice questions with feedback
	4. Reinforcement of instructed categories with recall of errors and categories with feedback
	5. Distractor task (eight demographic questions)
	6. Post-test category knowledge (sort task of nine financial statement errors)
Phase 2: Frequency learning (occurring one or two days after Phase 1)	1. Repeat post-test sort task from Phase 1
	2. Presentation of nine individual errors in varying frequencies (total of 49 individual presentations) — assignment of frequencies to categories and to errors within categories determined randomly for each subject
	3. Distractor task (four ability questions)
	4. Frequency knowledge tests for cycles and objectives — "no hint" (instructions included definitions of each category only)
	5. Audit decision task for cycles and objectives
	6. Repeat of frequency knowledge tests — "with hint" (adding category membership of each error and explanation to instructions followed by two ability questions)
NOTE: *Order of 4 and 5 determined randomly*	7. Debriefing questionnaire

time at which they would return for the second phase of the experiment. The first phase took an average of 33 minutes to complete. The second "frequency learning" phase of the experiment was conducted either one or two days later, depending on the availability of laboratory space. Subjects returned to the lab, repeated the knowledge post-test, observed individual presentations of financial statement errors designed to convey frequencies, and completed a distractor task, frequency knowledge tests, audit decision task, and a debriefing questionnaire. The second phase required an average of 44 minutes to complete.

Design. The overall design is a 2×2×3×2×3, with three between-subjects factors and two within-subjects factors. The primary between-subjects factor is "instruction type" (cycle or objective). The second between-subjects factor is whether the decision task used to test Hypothesis 3 (H3) appeared before or after the frequency knowledge test used to test H1 (the frequency knowledge test used to test H2, step 6 in phase 2, always appeared after these two tasks). This is a standard manipulation that was not central to our hypotheses; however, it allowed us to examine whether subjects would use frequency knowledge in their decision task without being prompted to do so. Finally, three subject groups were employed to obtain an adequate number of participants (undergraduates completing phase 2 after a one day interval and graduates completing phase 2 after a one and a two day interval).[8]

The two within-subjects variables were those relevant for testing our hypotheses. First, since both groups of subjects took two frequency knowledge tests (one for cycle and one for objective categories), but received only one type of instruction, the first within-subjects factor is "test consistency" (consistent or inconsistent test). Test consistency indicates which frequency test was consistent with and which test was inconsistent with the subject's instruction type. For example, a subject who received

[8]This variable was never significant in the analyses of variance.

instruction concerning the membership of errors in audit objective categories took a "consistent" frequency knowledge test when asked for the frequency of audit objective categories and an "inconsistent" test when asked for the frequency of transaction cycle categories. Second, in the frequency learning phase, all subjects experienced one sequence of 49 financial statement errors, with the number of times individual errors appeared in the sequence varied to convey category-level frequencies of 7, 14, and 28 for both the three transaction cycle categories and the three audit objective categories. As a consequence, regardless of the type of instruction each subject received, "category frequency" (7, 14, and 28) was varied within-subjects for both transaction cycle and audit objective categories.

Our hypotheses can be tested by examining the interaction between these two within-subjects variables (consistency of the test with a subject's instruction; actual experienced category frequencies [7, 14, and 28]). The expected form of the interaction is that actual relative category-level frequencies should have greater influence on subjects' responses to the consistent tests than the inconsistent tests. Subjects were randomly assigned to either the cycle or objective instruction groups and the decision- or frequency-first groups and proceeded as follows.

Phase 1 procedure: category learning. To determine if instruction created changes in subjects' knowledge of cycle or objective categories (and thus to provide a manipulation check), this knowledge was measured at the beginning and end of the "category learning" phase (i.e. both prior to and after instruction). Knowledge of category membership of errors was determined through a sorting task in which subjects were asked to sort the nine financial statement errors used in the experiment into three categories on the basis of either transaction cycle or audit objective, depending on their treatment group (see Table 2 for the errors).[9] To ensure that subjects understood this auditing terminology, definitions of the terms "transaction cycle" or "audit objective" were provided in the sort instructions. However, specific categories (e.g. sales and receipts, proper cutoff) were not listed, requiring subjects to rely on their own knowledge of category definitions and membership. To accomplish the sort task, subjects clicked on either a "1," "2," or "3" beside each error to indicate how they would group the errors.[10]

The nine financial statement errors used in the sort task and the remainder of the experiment are the same as those used in Nelson *et al.* (1995) and Bonner *et al.* (1996) studies of experienced auditors' conditional probability judgments. The errors selected were those for which there was the highest agreement on both cycle and objective category membership according to the results of Frederick *et al.* (1994). The transaction cycles are:

1. sales and receipts,
2. inventory/purchases, and
3. investments.

[9]Although this task may appear to be simple, previous research indicates that experienced auditors have some difficulty correctly categorizing errors in such a sort of task and, as noted previously, students have great difficulty (Frederick *et al.*, 1994). The latter result also is consistent with both our pilot tests and our pre-test sort results. Thus, a change in sort task performance from pre-test to post-test would indicate that knowledge of category membership of errors had been imparted through instruction.

[10]These directed sorts were similar to those used in the Frederick *et al.* (1994) directed sorts of 35 financial statement errors. The same type of sort task was also used in Nelson *et al.* (1995) and Bonner *et al.* (1996) to measure experienced auditors' pre-existing knowledge structures for errors. In these studies, however, auditors were asked to sort errors based on how they best went together (a free sort). These studies used a free sort task to determine the knowledge organizations which develop from auditors' experience, while we use our directed sorts to assess the effectiveness of our instruction manipulation.

TABLE 2. Financial statement errors

	Transaction cycle		
Audit objective	Sales	Inventory/purchases	Investments
Proper cut-off	Next period's sales were included in the current year's revenue and receivables	Raw materials were improperly shown as received after year-end	Purchases of treasury bills were recorded in the wrong fiscal period
Validity	Billings to legitimate customers were booked twice	More finished goods were recorded as received than were actually received	Fictitious investments were included in the account balance
Valuation	The bad debt expense and allowance were underestimated	Obsolete inventory was not written down to net realizable value	Marketable securities were not reduced to lower of cost or market

Based on Nelson *et al.* 1995.

The audit objectives are:

1. proper cutoff,
2. validity, and
3. valuation.[11]

After sorting the nine financial statement errors at pre-test, subjects received instruction on either transaction cycles or audit objectives. The transaction cycle (audit objective) instruction explained the importance of transaction cycles (audit objectives) to audits, then introduced the three cycles (objectives) used in the experiment. For each cycle (objective), subjects read a definition and a description of the types of errors that would be classified as belonging to that cycle (objective). Next, for each cycle (objective), subjects saw each of the three corresponding financial statement errors (see Table 2) followed by an explanation of why the error was a member of that category. The final screen then summarized the preceding instruction. For an example of this instruction, see Panel A of the Appendix.

To ensure that this instruction was understood fully, subjects answered 18 multiple-choice questions, nine of which asked subjects to select the cycle (objective) to which the nine financial statement errors corresponded and nine of which asked subjects to choose an error (from a group of 3) that related to a cycle (objective) category. The order of the questions within each group of nine was randomized for each subject. To ensure that any confusion was corrected, subjects were provided with feedback after each response. The feedback consisted of either the word "correct" or "incorrect" followed by a statement of the correct answer and an explanation of why the correct answer was correct.[12] Subjects repeated the multiple-choice test up to a total of four times until they completed the questions with one or none of the 18 questions answered incorrectly. To further ensure that subjects had learned the categories, they performed two recall tasks, similar to those used by Barsalou and Ross (1986). The first task required subjects to recall the three errors presented in each category, and the second to recall the category associated with each error. Subjects were presented with feedback after each of these recall tasks, but only performed each task once.

[11]Some of the errors were reworded so as to avoid having surface structure (wording of the errors) confounded with deep structure (transaction cycle or audit objective). For example, the word "sales" was only used to describe one of the sales cycle errors. The order of presentation of the nine errors in the sort task was randomized and held constant across subjects.

[12]The feedback is similar to the explanatory feedback provided by Bonner and Walker (1994) which promoted learning of a different type of audit knowledge.

TABLE 3. Individual and category-level frequencies of errors

	Transaction cycle			
Audit objective	A	B	C	Total-objective category
I	1	2	4	7
II	2	4	8	14
III	4	8	16	28
Total-cycle category	7	14	28	49

Based on Nelson *et al.* 1995.

After receiving instruction as described above, subjects completed a distractor task to clear their short-term memory. This distractor task involved answering eight demographic questions about age, undergraduate versus graduate status, university, major, number of accounting credits completed, number of auditing credits completed, amount of public accounting experience (if any), and whether English was a first or second language. Finally, subjects performed a post-instruction sort task following the same procedure as in the sort performed at the beginning of this phase. The purpose of this sort test was to confirm that the category instruction had influenced how subjects categorize financial statement errors.

Phase 2 procedure: frequency learning and testing. In phase 2 of the experiment, subjects first reperformed the sort task from phase 1 to determine whether they had retained the category knowledge acquired there. Then, they viewed individual presentations of each of the nine financial statement errors in varying frequencies determined in the same manner as in Nelson *et al.* (1995) and Bonner *et al.* (1996), for a total of 49 presentations. The frequency of presentation of individual errors was determined by randomly assigning the three transaction cycles and audit objectives to the rows and columns shown in Table 3 for each subject.[13] The subjects were instructed that they would be presented with a series of errors found during the audits of medium-sized manufacturers by a major accounting firm during the last quarter of 1992, and that their task was to remember the errors to the best of their ability. No mention was made of frequency learning. The subjects were then told that each error would remain on the computer screen for ten seconds; this time was the same as that used in Nelson *et al.* (1995) and Bonner *et al.* (1996) and similar to that used by Butt (1988).

After viewing the individual error presentations, subjects performed a distractor task to clear short-term memory. This task consisted of answering four items from the ability test used in Bonner and Walker (1994).[14] Next, subjects were asked to imagine that they were an auditor who was working for the firm from which the list of errors was obtained and to assume that they had been exposed to all of those audits of manufacturing companies, such that the errors constituted their experience. Then, subjects were presented with the unaided frequency knowledge test and audit decision task, the order of which was determined randomly as indicated earlier.

The unaided frequency knowledge test was used to examine H1, and required estimates of three category-level frequencies for both cycles

[13]The order of the 49 presentations was randomized between subjects with the constraints that: (1) each repetition of an error was separated by at least one other error, and (2) the three highest frequency errors were presented as often in the first half of the sequence as in the second half.

[14]Based on previous research (Goldstein *et al.*, 1983; Zacks *et al.*, 1982), we did not expect that ability would be related to the acquisition of frequency knowledge. Instead, we chose this distractor task because it was sufficiently complex to clear short-term memory. Consistent with our expectation, ability was not related to frequency judgments.

and objectives. For each test, a definition of cycle (objective) and the names and definitions of the three relevant cycle (objective) categories were listed. Neither the cycle nor objective category membership of each error nor an explanation of its membership was provided. Thus, the information presented was sufficient to classify the nine errors that had been presented if the subjects remembered all nine errors and understood the basic journal entries involved. Subjects were asked to mark on a sliding scale a number from one to 34 to indicate their frequency estimate. These endpoints were chosen to range six below and six above the minimum and maximum actual category frequencies to ensure that the answers were not obvious. The order of the two tests (cycle and objective) and the order of the three cycle or objective questions within each test were randomly selected for each subject. Panel B of the Appendix presents an example of a frequency knowledge test screen.

In the audit decision task, used to test H3, subjects were told to imagine that they were planning an audit of a similar medium-sized manufacturing company, that the error presentations they had just seen constituted their "experience" for the last quarter of 1992, and that their task was to allocate 100 hours of audit tests in a manner that would detect as many errors as possible. All subjects performed two tasks: allocating the 100 hours over transaction cycles and over audit objectives. The order of the allocation tasks (by cycle or objective) and the order of the three cycles or objectives within each task were determined randomly for each subject. As shown in panel C of the Appendix, subjects moved a pointer on a sliding scale with endpoints of zero and 100 to indicate the hours of testing they would budget for each of the three cycles and for each of the three objectives.[15]

Finally, to examine whether frequency knowledge gains could be duplicated by providing listings and explanations of category memberships after experience (as a test of H2), all subjects took the aided frequency knowledge test. This task provided subjects with all the category information necessary to estimate category-level frequencies and again elicited their estimates of those frequencies. Any differences here between the consistent and inconsistent tests in subjects' sensitivity to actual category-level frequencies should be due to the effect of knowing category memberships prior to experiencing error frequencies. An example of this information is provided in Panel D of the Appendix.

The aided frequency knowledge test contained two parts. In the first part, each subject was provided with definitions of categories and information about which of the errors were contained in each cycle (or objective) category and why. After answering two ability questions from Bonner and Walker (1994), the subject answered the same questions that were asked in the unaided frequency knowledge test. In the second part, this entire process (category information, ability questions, frequency test) was repeated with respect to the set of categories (either objective or cycle) not featured in the first part. As with the unaided frequency questions, the order of the two parts of the test (cycle and objective) and the order of the three cycle and objective questions within each part were randomly selected for each subject. The aided frequency knowledge test always appeared after both the initial frequency knowledge test and the audit decision task to avoid contaminating the results of these two measures.

RESULTS

Manipulation check

To ensure that our instruction on categories of errors was successful, we conducted a mani-

[15]This task was similar to the decision task performed by experienced auditors in Nelson *et al.* (1995), in that subjects allocated hours among various audit areas. However, the auditors in the Nelson *et al.* study made conditional audit allocations, i.e. they were asked: "given that you are auditing the sales cycle, how would you allocate hours among various audit objectives?" Given that this study focused on the effect of category instruction on frequency learning, we used the simpler task of allocating hours across cycle and objective categories.

pulation check using the pre-test and post-test sort data described above. This manipulation check was accomplished by comparing the fit of the pre-test and post-test category knowledge implicit in the subjects' two sorts to our predetermined categorizations. Effective instruction would be indicated by a closer fit of the post-test sorts than of the pre-test sorts to the predetermined categorizations.

To conduct the tests, each subjects' pre-test and post-test sorts were converted into 9×9 similarity matrices containing 1s if financial statement errors were grouped together by the subject and 0s if errors were not grouped together. 9×9 similarity matrices were also formed for the predetermined transaction cycle and audit objective categorizations presented in Table 2. The fit of each subject's category knowledge to the predetermined categorization was measured by correlating the vector formed from the lower left triangle of the subject's similarity matrix with the vector formed from the lower left triangle of the predetermined category similarity matrix for the appropriate basis of categorization (either transaction cycles or audit objectives). This measure of the similarity of two similarity matrices is commonly called a "cophenetic correlation" (Sneath & Sokal, 1973 p. 97), and has been used in prior research to examine the knowledge structures of auditor subjects (Bonner *et al.*, 1996; Frederick *et al.*, 1994; Nelson *et al.*, 1995).

The mean measures of fit for the pre-test and post-test sorts were compared for each type of instruction to test its effectiveness. At pre-test, the cycles group had an average correlation with the predetermined categories of 0·31 (with a 95% confidence interval of 0·13 to 0·49), while at post-test each subject attained a perfect score of 1·00. In the objectives group, the pre-test mean was 0·37 (with a 95% confidence interval of 0·21 to 0·54) and, again, at post-test, each subject obtained a perfect score. These results suggest that our instruction manipulation successfully imparted knowledge of category membership of the errors examined here. When the post-test sort was repeated at the beginning of phase 2, all but one subject again achieved a perfect score. This demonstrates that subjects retained the category knowledge they acquired in phase 1.

Hypothesis 1 — effect of category knowledge acquired prior to experiencing errors on category-level frequency knowledge

As noted earlier, the effects of the two within-subjects factors (consistency of test with instruction and experienced category-level frequencies) on the subjects' unaided frequency estimates relate to hypothesis 1, and one between-subjects factor tests the generality of findings across the two instruction types (cycle or objective). The two other between-subjects factors (subject type and order of frequency vs. decision tasks) were unrelated to this hypothesis, were balanced in our design, and thus are eliminated from our analysis.[16]

Cell means and standard deviations are presented in Table 4. Hypothesis 1 predicts that subjects receiving transaction cycle (audit objective) category instruction will make frequency estimates which more closely reflect subsequent experienced relative category-level error frequencies for transaction cycle (audit objective) categories than for audit objective (transaction cycle) categories. This difference in learning to discriminate frequencies would be indicated if subjects' frequency estimates for consistent test categories increase at a greater rate with actual frequencies than do their frequency estimates for inconsistent test categories.

The appropriate method for testing this directional hypothesis is with a priori contrasts based on the specific directional hypothesis.[17] Prior research in psychology and accounting has analyzed subjects' ability to discriminate

[16]When these factors are included in the analysis, they do not change interpretation of results.

[17]See Buckless and Ravenscroft (1990) for a discussion of accounting applications and Rosenthal and Rosnow (1985) for detailed descriptions.

TABLE 4. Frequency estimates (unaided): analysis of consistent frequency test vs. inconsistent test × linear trend in actual frequencies contrast

Actual frequency	Consistent test		Inconsistent test	
7	12·26(6·1)		12·80 (6·9)	
14	13·37 (6·5)		14·22 (6·5)	
28	19·18 (6·4)		15·54 (6·9)	
Analysis of variance				
Source		d.f.	F	p (two-tailed)
Linear trend in actual frequencies (LF)		1	26·56	0·000
LF × instruction		1	0·21	0·645
Error		63	(61·23)	
Instruction		1	2·43	0·124
Error		63	(102·55)	
Consistency of test		1	3·67	0·060
Consistency of test × instruction		1	2·51	0·118
Error		63	(15·50)	
LF × consistency of test		1	14·14	0·000
LF × consistency of test × instruction		1	0·62	0·433
Error		63	(24·50)	

Means (S.D.).
Values enclosed in parentheses are mean square errors.

relative frequencies by examining the significance of the linear trend in actual frequencies of subjects' frequency estimates (e.g. Hintzman *et al.*, 1982; Butt, 1986). The form of the above prediction suggests that the linear trend in actual frequencies of subjects' responses will be greater for the consistent test than the inconsistent test, which is examined with the consistent/inconsistent test × linear trend in actual frequencies contrast. This contrast was significant ($t(63)=3·76$, $p=0·0002$, one-tailed), indicating that subjects' frequency estimates more closely reflect experienced relative category-level error frequencies for the consistent categories than for the inconsistent categories, thus supporting H1.[18] The linear trend was significant for both the consistent ($t(63)=6·4$; $p=0·000$, one-tailed) and inconsistent categories ($t(63)=2·4$; $p=0·009$, one-tailed), indicating some ability to discriminate relative category-level frequencies in both conditions.[19] This suggests that, based on the definitions and explanations of each cycle or objective provided in the frequency test for the

[18]The consistent/inconsistent test × quadratic trend in actual frequencies contrast was not significant ($F(1,63)=2·16$, $p=0·147$, two-tailed).

[19]This test is equivalent to testing whether the slope of subjects' category frequency judgments was significantly different for the two tests. Plotted against actual category frequency, both slopes are two flat, such that an increase in slope for the consistent test represents an improvement in relative frequency judgment. Also note that our tests employed equal spacing, whereas the most appropriate contrast coefficients for linear trend with unequal spacing 7, 14, and 28 are -9·333, -2·333, and 11·666, and the appropriate error term for this contrast is linear trend in frequency, consistency of test × subjects within training groups (Winer, 1971, p. 540). Equivalent results are obtained if the data are analyzed using contrasts appropriate for unequal spacing. Equivalent results are obtained if the data are analyzed using a traditional ANOVA. The consistent/inconsistent test × actual frequency interaction is significant ($F(2, 126)= 7·75$; $p=0·007$), with simple effects indicating that actual category frequency is significant in both the consistent ($F(2,126)=27·04$; $p=0·000$) and inconsistent ($F(2,126)=3·41$; $p=0·036$) tests.

uninstructed categories, and any prior knowledge of the uninstructed categories, subjects were able to discriminate category relative frequencies to some degree, most likely by summing individual error frequencies.[20] However, they were significantly better able to do so for categories for which they received instruction, indicating the benefit of having category knowledge in place prior to experiencing the errors. Further, the three-way interaction of the contrast with type of instruction was not significant ($F(1,63)=0{\cdot}62$, $p=0{\cdot}433$), indicating no differences in the effectiveness of the two types of instruction (cycle vs. objective) for subsequent learning from experience.

If auditors use frequency knowledge simply to identify more vs. less risky areas for further examination, the ability to rank order the categories becomes critical. Thus, we also examined subjects' ability to rank order categories by frequency by computing the rank order correlation between their unaided category frequency judgments and the actual frequencies and comparing the mean correlations for the consistent and inconsistent test categories. The mean correlation for the consistent test categories (0·54) was significantly higher ($p=0{\cdot}000$, one-tailed) than the mean correlation for the inconsistent categories (0·18), indicating superior ability to rank order categories by frequency for categories with respect to which subjects received instruction prior to exposure to frequencies.

Hypothesis 2—effect of category knowledge acquired after experiencing error frequencies on category-level frequency knowledge

While the unaided frequency test included the definition and explanation of each cycle and objective, it did not include listings of the errors included in the categories and reasons for their inclusion. As noted earlier, after completing both the unaided frequency knowledge tests and audit decision task, subjects were presented with the information about category memberships, explanations for category memberships, and category definitions that was included in the phase 1 instruction for both the cycle and objective categories. Then, subjects took the frequency knowledge tests again. Recall that the purpose of these aided frequency tests was to test whether providing these instructions after experiencing the errors is a good substitute for instruction before experience.

Table 5 presents cell means and standard deviations for the aided frequency knowledge test. This table indicates that the consistent/inconsistent test × linear trend in actual frequencies contrast was still highly significant ($t(63)=3{\cdot}56$, $p=0{\cdot}000$, one-tailed), indicating that actual category-level relative frequencies still had a greater effect when subjects had category knowledge in place before experiencing the errors, supporting H2. The linear trend was significant for both the consistent ($t(63)=6{\cdot}80$; $p=0{\cdot}000$, one-tailed) and inconsistent tests ($t(63)=3{\cdot}32$; $p=0{\cdot}001$, one-tailed), again indicating some ability to discriminate category frequencies in both conditions.[21] This table indicates a slightly smaller effect for initial instruction in the aided frequency knowledge test than in the unaided frequency knowledge test. However, when both frequency tests (aided and unaided) were included in the analysis together, the aided/unaided × consistent/inconsistent test × linear trend in actual frequencies contrast was not significant ($p=0{\cdot}5$), indicating no significant

[20]Subjects were better able to discriminate between the medium and high levels of frequency than between the low and medium levels of frequency. In the last section of the paper, we note a possible explanation for this effect and compare the performance of our subjects to those in prior psychology and accounting studies.

[21]Equivalent results are obtained if the data are analyzed using contrasts appropriate for unequal spacing. Equivalent results are also obtained if the data are analyzed using a traditional ANOVA. The consistent/inconsistent test × actual frequency interaction is significant ($F(2,126)=6{\cdot}69$; $p=0{\cdot}002$) tests, with simple effects indicating that actual category frequency is significant in both the consistent ($F(2,126)=29{\cdot}72$; $p=0{\cdot}000$) and inconsistent ($F(2,126)=6{\cdot}69$; $p=0{\cdot}002$) tests.

TABLE 5. Frequency estimates (aided): analysis of consistent frequency test vs inconsistent test × linear trend in actual frequencies contrast

Actual frequency	Consistent test	Inconsistent test
7	12·71 (5·4)	14·09 (5·3)
14	14·32 (6·0)	14·86 (5·5)
28	19·34 (6·6)	17·02 (6·4)

Analysis of Variance

Source	d.f.	F	p (two-tailed)
Linear trend in actual frequencies (LF)	1	36·68	0·000
LF × Instruction	1	0·01	0·939
Error	63	(41·18)	
Instruction	1	0·47	0·494
Error	63	(102·63)	
Consistency of test	1	0·16	0·695
Consistency of test × instruction	1	1·53	0·221
Error	63	(10·11)	
LF × consistency of test	1	12·68	0·001
LF × consistency of test × instruction	1	0·00	0·971
Error	63	(19·11)	

Means (S.D.).
Values enclosed in parentheses are mean square errors.

difference in the effect of instruction between the aided and unaided conditions.

As in our test of H1, we also examined subjects' ability to rank order categories by frequency by computing the rank order correlation between their aided category frequency judgments and the actual frequencies and comparing the mean correlations for the consistent and inconsistent test categories. The mean correlation for the consistent test categories (0·48) was significantly higher (p=0·012, one-tailed) than the mean correlation for the inconsistent categories (0·25), indicating superior ability to rank order categories by frequency for categories with respect to which subjects received instruction prior to exposure to frequencies, even when subjects were provided.

In sum, these results indicate that the instruction effect is still highly significant for aided frequency judgments. This indicates that prior instruction in categories can provide category knowledge which assists in the later learning of relative category-level frequencies from experience, and that this benefit of having pre-existing knowledge cannot be easily compensated for after the fact.

Hypothesis 3 — effect of acquired category-level frequency knowledge on audit decisions

H1 hypothesizes that instruction affects category-level frequency learning, and H3 hypothesizes that instruction consequently affects audit decisions through the effect of the learned frequency knowledge on those decisions. To test H3, we first performed an ANOVA similar to that employed to test H1 using the allocation decisions as the dependent variable and instruction, experienced frequencies, and test consistency as the key independent variables. This ANOVA tests the direct effect of instruction on audit decisions. Then, we added subjects' unaided frequency estimates as a covariate in an ANCOVA to determine if the effect of the treatment variables (instruction and experienced frequencies) was at least partially explained by the category-level frequency knowledge acquired from experience.

TABLE 6. Audit decisions (hours allocations): analysis of consistent decision task vs. inconsistent task × linear trend in actual frequencies contrast

Actual frequency	Consistent test	Inconsistent test
7	28·42 (9·9)	31·26 (10·2)
14	30·39 (11·6)	32·55 (12·5)
28	41·18 (12·2)	36·18 (12·6)
Audit decision first		
7	36·44 (10·5)	37·59 (10·5)
14	28·33 (9·1)	30·59 (9·3)
28	35·22 (6·5)	31·81 (11·2)

The means and standard deviations for the audit decisions are presented in Table 6. In the first part of the analysis, if relative category-level frequencies are more closely reflected in audit effort allocations for the category type (cycle or objective) on which subjects received instruction (the consistent task), subjects' decisions regarding hours to allocate for testing will produce a significant consistent/inconsistent decision task × linear trend in actual frequencies contrast. A significant consistent/inconsistent task × linear trend in actual frequencies contrast was present for the audit hours decisions ($t(53)=1{\cdot}89$, $p=0{\cdot}032$, one-tailed). The linear trend was significant for the consistent task ($t(53)=2{\cdot}78$; $p=0{\cdot}004$, one-tailed) but not the inconsistent task ($t(53)=0{\cdot}02$; $p=0{\cdot}44$, one-tailed).[22] As in the analysis of the frequency tests, the interaction occurred because subjects instructed in cycle (objective) categories were better able to allocate audit hours on a basis which discriminates between low and high frequency levels for cycle (objective) categories than for objective (cycle) categories. However, allocations of audit hours appear to reflect actual frequencies to a lesser extent than do the frequency estimates analyzed previously; e.g. the audit hours allocation is lower in the frequency 14 category than in the frequency 7 category.

In the second part of the analysis, we added subjects' unaided frequency estimates as a covariate in the above analysis, then examined the changes in the significance levels of the actual frequency variable and the consistent/inconsistent test × linear trend in actual frequencies contrast (see Heiman-Hoffman, 1992, and Moser, 1992, for similar approaches). If the effect of pre-existing category knowledge (as reflected by the interaction of decision task consistency and the linear trend in actual frequencies) on the audit decisions is mediated by subjects' frequency learning, we would expect the covariate (frequency estimates) to be significant and the actual category frequency variable and the consistent/inconsistent task × linear trend in actual frequencies contrast to become less significant. In the ANCOVA, the covariate, frequency estimate, was significant ($F(1,52)=10{\cdot}15$, $p=0{\cdot}001$, one-tailed), the independent variable, actual category frequency, became nonsignificant ($F(1,52)=0{\cdot}08$, $p=0{\cdot}61$, one-tailed), and the consistent/inconsistent test × linear trend in actual frequencies contrast became nonsignificant ($F(1,52)=1{\cdot}28$, $p=0{\cdot}87$, one-tailed). This suggests that subjects' frequency knowledge at least partially explained the interactive effect of category knowledge and actual frequency on their decisions, supporting H3.

[22]The order and subject type variables included in these analyses change the degrees of freedom in the relevant tests. Also, the traditional test of the consistent/inconsistent test × actual frequency interaction cannot be computed because the three allocations must sum to 100. When the middle allocation is removed to eliminate the linear dependence, the actual frequency by consistent/inconsistent test interaction is the same as the linear trend by consistent/inconsistent test interaction.

Finally, unlike the tests related to frequency knowledge, an additional interaction was in evidence which provides further insight into the audit decision data. In particular, the order (frequency test or audit decision first) × linear trend in actual frequencies contrast was significant (p=0·001; two-tailed). When subjects performed the frequency test first, their mean audit decisions were strictly increasing in actual frequencies as we had expected. However, when the audit decision task was performed first, they actually allocated the most time to the lowest frequency event (7), followed by the most frequent (28), followed by the middle level (14). We see no obvious cause for this pattern. However, it is important to note two points. First, this interaction does not involve a three-way interaction with the consistent/inconsistent test × linear trend in actual frequencies contrast, and thus does not qualify the results for H3 presented above. Second, it seems clear that taking the frequency test first reminds subjects of the relevance of frequency data which they are then more likely to use in their audit decisions. Recall that, prior to the frequency tests, subjects were told only that their task was to learn the errors included on the list of 49 instances. Given the nature of these instructions, and the lack of use of frequencies in the decision first condition, many of the subjects in the decision-first condition may have still been unaware that they were supposed to have learned frequencies, let alone that the frequencies should be applied in decisions.

DISCUSSION

Implications for auditing theory and practice

Category knowledge plays a variety of roles in audit decisions. We focused on the effect of category knowledge that organizes experiences on subsequent learning from those experiences, and the impact of the knowledge learned from experience on audit decisions. We also examined whether the observed learning effects could be duplicated by providing subjects with information about categories after experience. Specifically, we investigated whether focused, intensive instruction in transaction cycle and audit objective categories of errors facilitates learning of category knowledge, learning of relative category-level error frequencies from subsequent experience, and application of these frequencies to a simplified audit planning decision. These issues are important to examine because understanding the particular aspects of knowledge that auditors need for further acquisition and application of decision-relevant knowledge is necessary before educators and firms can determine the best ways to organize auditors' instruction and experience.

Results were consistent with expectations. The results of our sort task confirmed the effectiveness of our focused instruction in creating category knowledge. The results for our frequency knowledge test support an important role for instruction in determining the accuracy of frequency learning, consistent with the findings of Barsalou and Ross (1986) and opposing the extreme notion that such processes are fully automatic even when category structures are not well known. Further, comparisons of aided to unaided frequency estimates indicate that instruction which occurs after exposure to frequencies may be less effective than instruction which occurs before exposure. Although important caveats apply (see the next section), this result suggests that the effect of instruction is stronger when it occurs before frequencies are encoded, rather than after, which is consistent with some theories of frequency learning (e.g. Begg *et al.*, 1986; Greene, 1989) but inconsistent with others (e.g. Alba *et al.*, 1980; Hasher & Zacks, 1979, 1984). These findings also are consistent with earlier findings that suggest auditors learn and perform better on tasks where instruction about an organizing framework precedes experience (Bonner & Pennington, 1991; Bonner & Walker, 1994).

Of additional interest is the significant order effect, where subjects performing the frequency knowledge test first were much more likely to incorporate their frequency knowledge in their audit decisions. This result may suggest that, in addition to providing category instruction before experience to aid in gaining frequency knowledge, prompting novice auditors to use this frequency knowledge when making decisions may also be helpful. Prior research by Butler (1985) and Heiman (1990) supports the importance of decision aids which prompt auditors to consider relevant knowledge.

Limitations and future research

Several limitations of this study suggest directions for future research. The first limitation relates to our use of student subjects. Intermediate accounting students who had not taken auditing are an appropriate group of subjects for examining the effects of instruction regarding auditing-related categories on frequency learning and estimation because this group possessed the basic knowledge of accounting necessary to understand financial statement errors, but not auditing categories. This suggests one circumstance where constructive use can be made of student subjects: whenever subjects are required to lack an element of knowledge that develops with experience, students who lack that knowledge can be used to examine the effect of instruction. We are less comfortable drawing implications from the decision results, since other knowledge besides frequency knowledge may develop with experience and influence these sorts of decisions. Although it is difficult to say to what degree the decision results would generalize to more experienced auditors, we can note that the experienced auditors who participated in the Nelson *et al.* (1995) study were able to apply experienced frequencies to conditional audit allocation decisions, which are somewhat similar to the audit planning decisions required of students here.

A related limitation concerns interpretation of the results of H2. Recall that the learning advantage provided by category instruction prior to experience could not be duplicated by the information we provided after experience. We interpret this result as suggesting that it is important to have category knowledge in place prior to experiencing frequencies. However, this result could also be generated by subjects learning category memberships of errors better from the instruction provided during phase 1 than in the information provided during phase 2, which made it more difficult for them to sum individual error frequencies in phase 2. Given that the category memberships were directly presented to subjects in phase 2 without extraneous information, we consider this explanation unlikely. Another explanation for the result could be carryover effects from the unaided frequency knowledge tests and/or audit decision task (since the aided frequency knowledge test always occurred after the unaided frequency knowledge test and audit decision task). Since several tasks separated the aided frequency estimates from previous estimates and decisions, we also consider this explanation unlikely.

Another potential limitation is that, as noted earlier in the paper, while instruction facilitated learning of relative category-level error frequencies from subsequent experience, subjects' absolute frequency estimates were still inaccurate. In particular, their absolute estimates were too flat; estimates of low frequency events were too high and estimates of high frequency events were too low. A comparison with levels of performance in prior psychology and accounting studies provides a basis for evaluating these performance levels. In psychology, Freund and Hasher (1989) presented findings with respect to an experiment where subjects were instructed to categorize simple items (e.g. fruits) based on a particular property they possessed, e.g. color. Actual category frequencies ranged from zero to eight in their experiment. Similar to our results, subjects overestimated actual frequencies for the two lowest-frequency categories and underestimated frequencies for the three highest-frequency categories. Alba *et al.* (1980) and Barsalou and Ross (1986) present similar findings. In accounting, although there have been no prior studies of category-level frequencies with complex accounting stimuli, we can estimate both student and experienced auditor performance, given perfect aggregation of individual frequencies into categories, by forming *ad hoc* categories using Butt's (1988) data. To make this estimate, we combined triples of her subjects' individual error frequencies into category frequencies (10, 18, and 26) which most closely matched our three levels of category frequencies (7, 14, and 28). Between the 10 and 26 frequency levels, her student subjects' *ad hoc* category frequency judgments increased 30·2% as quickly as actual frequency. Between the 7 and 28 frequency levels, our instructed subjects' category frequency judgments increased 33·0% as quickly as actual frequency. Butt's (1988) auditor subjects' *ad hoc* category frequency judgments increased 44% as quickly as actual frequency. Given that our addition of Butt's data assures perfect knowledge of categories and perfect aggregation of individual frequencies into category-level frequencies, the performance of our subjects appears reasonable and may be approaching the maximum that can be expected in such an experimental task.

A related point is that the primary effect of instruction occurred in helping subjects discriminate between the medium (14) and high (28) levels of category frequency. We chose category-level frequencies of 7, 14 and 28 based on Weber's law, which holds that "just-noticeable differences" increase at a constant ratio (Stevens, 1975). According to Weber's law, the constant 2:1 ratio between levels of category frequency we selected should have resulted in a constant level of ability to discriminate frequencies between adjacent levels, counter to what we observed. The only explanation we can offer for this is that, for some reason, subjects pay more attention to raw levels of differences rather than ratios of differences in our auditing context. This suggests future research designed to determine the amount of difference in actual frequency necessary for auditors to perceive that a difference exists.

Another limitation is that we examine only one of the many potential factors which may determine the effectiveness with which auditors learn from instruction and experience. Our results indicate that a category structure which is present prior to experience improves future judgments and decisions that access memories of experiences with respect to that structure. However, other factors also might determine the effectiveness with which an event is stored and available for use in subsequent judgments. For example, our simple, one-sentence error presentations are likely to be less salient than errors experienced in practice, and thus processed less well. On the other hand, errors in practice have more unique features than do the error presentations used in our experiment While these unique features would render more discriminable the memory traces of errors experienced in practice, they also may render memory traces less likely to be categorized in the same group. Also, errors in practice are encountered in widely-spaced intervals, while error presentations in the experiment occurred in closely-spaced intervals, which would tend to favor frequency discrimination in our experiment over practice. Future research should seek to understand these and other determinants of effective learning to assist audit firms in planning the timing and content of auditors' instruction and experience.

A final question relates to whether decision aids or instruction would be more appropriate for improving category-level frequency estimates and the decisions based thereon, which are currently made by each individual auditor based on his or her experiences. For example, one possible decision aid would provide auditors with category-level error frequencies, rather than teaching category structure and having the auditors accumulate frequencies themselves. However, auditors largely ignore such summarized base rate data in audit judgments (see Nelson, 1994, for a review); further, accumulating category-level frequencies that would be appropriate for each engagement might be prohibitively expensive for a firm. Alternatively, a decision aid that lists the errors that occur in each transaction cycle might take the place of category cycle instruction by enabling auditors to recall and sum individual error frequencies which have been learned through experience. However, this approach also has weaknesses. Such a decision aid would be equivalent to the instruction we provided after experience in this study, which indicates that category-level error frequencies still would be better estimated when category instruction had preceded experience. Further, using such a decision aid in place of category instruction would not enhance audit decisions that rely on knowledge of categories for other processes such as recall (see, e.g. Frederick, 1991). Finally, decision aids that allow auditor inputs can be vulnerable to misapplication (Kachelmeier & Messier, 1990). Future research should seek to determine the costs and benefits of decision aids and instruction with respect to each audit decision in which they might be used before selecting among alternative approaches.

BIBLIOGRAPHY

Adelson, B. (1985). Comparing natural and abstract categories: a case study from computer science. *Cognitive Science*, 417-430.

Alba, J. W., Chromiak, W., Hasher, L., & Attig, M. S. (1980). Automatic encoding of category size information. *Journal of Experimental Psychology: Human Learning and Memory*, 370-378.

Ashton, A. H. (1991). Experience and error frequency knowledge as potential determinants of audit expertise. *The Accounting Review*, 218-239.

Ashton, R. H., Kleinmuntz, D. N., Sullivan, J., & Tomassini, L. A. (1989). Audit decision making. In A. R. Abdel-Khalik & I. Solomon (Eds.), *Research opportunities in auditing: the second decade*. Sarasota: American Accounting Association.

Barsalou, L. W., & Ross, B. R. (1986). The roles of automatic and strategic processing in sensitivity to superordinate and property frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 116-134.

Bédard, J., & Chi, M. (1993). Expertise in auditing. *Auditing: A Journal of Practice and Theory* (Suppl.), 21-45.

Begg, I., Maxwell, D., Mitterer, J. O., & Harris, G. (1986). Estimates of frequency: attribute or attribution? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 496-508.

Bonner, S. E., Libby, R., & Nelson, M. W. (1996). Using decision aids to improve auditors' conditional probability judgments. *The Accounting Review*, 221-240.

Bonner, S. E., & Pennington, N. (1991). Cognitive processes and knowledge as determinants of audit expertise. *Journal of Accounting Literature*, 1-50.

Bonner, S. E., & Walker, P. L. (1994). The effects of instruction and experience on the acquisition of auditing knowledge. *The Accounting Review*, 157-178.

Buckless, F. A., & Ravenscroft, S. P. (1990). Contrast coding: a refinement of ANOVA in behavioral analysis. *The Accounting Review*, 933-945.

Butler, S. A. (1985). Application of a decision aid in the judgmental evaluation of substantive tests of details samples. *Journal of Accounting Research*, 513-526.

Butt, J. L. (1986). *Frequency judgments in auditing*. Unpublished Ph.D. dissertation, University of Michigan.

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

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Choo, F., & Trotman, K. T. (1991). The relationship between knowledge structure and judgments for experienced and inexperienced auditors. *The Accounting Review*, 464-485.

Coakley, J. R., & Loebbecke, J. K. (1985). The expectation of accounting errors in medium-sized manufacturing firms. *Advances in Accounting*, 199-246.

Davis, C. E. (1994). *Experience and the organization of auditors' knowledge of financial statement errors*. Working paper, Baylor University.

Frederick, D. M. (1991). Auditors' representation and retrieval of internal control knowledge. *The Accounting Review*, 240-258.

Frederick, D. M., Heiman-Hoffman, V., & Libby, R. (1994). The structure of auditors' knowledge of financial statement errors. *Auditing: A Journal of Practice and Theory*, 1-21.

Freund, J. S., & Hasher, L. (1989). Judgments of category size: now you have them, now you don't. *American Journal of Psychology*, 333-352.

Goldstein, D., Hasher, L., & Stein, D. K. (1983). The processing of occurrence rate and item information by children of different ages and abilities. *American Journal of Psychology*, 229-241.

Greene, R. L. (1989). On the relationship between categorical frequency estimation and cued recall. *Memory and Cognition*, 235-239.

Hasher, L., & Zacks, R. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, 356-388.

Hasher, L., & Zacks, R. (1984). Automatic processing of fundamental information: the case of frequency of occurrence. *American Psychologist*, 1372-1388.

Heiman, V. B. (1990). Auditors' assessments of the likelihood of error explanations in analytical review. *The Accounting Review*, 875-890.

Heiman-Hoffman, V. B. (1992). *Experience effects in analytical review: the relation between knowledge and judgment.* Working paper, University of Pittsburgh.

Hintzman, D. L., Nozawa, G., & Irmscher, M. (1982). Frequency as a nonpropositional attribute of memory. *Journal of Verbal Learning and Verbal Behavior*, 127-141.

Homa, D. (1984). On the nature of categories. In G. H. Bower (Ed.), *The psychology of learning and motivation* Vol. 18. New York: Academic Press.

Hylas, R., & Ashton, R. H. (1982). Audit detection of financial statement errors. *The Accounting Review*, 751-765.

Kachelmeier, S. J., & Messier, W. F. Jr. (1990). An investigation of the influence of a nonstatistical decision aid on auditor sample size decisions. *The Accounting Review*, 209-226.

Kida, T. (1984). The impact of hypothesis-testing strategies on auditors' use of judgment data. *Journal of Accounting Research*, 332-340.

Libby, R. (1985). Availability and the generation of hypotheses in analytical review. *Journal of Accounting Research*, 648-667.

Libby, R. (1995). The role of knowledge and memory in audit judgment. In R. H. Ashton & A. H. Ashton (Eds.), *Judgment and decision-making research in accounting and auditing.* Cambridge: Cambridge University Press.

Libby, R., & Frederick, D. M. (1990). Experience and the ability to explain audit findings. *Journal of Accounting Research*, 348-367.

Libby, R., & Luft, J. (1993). Determinants of judgment performance in accounting settings: ability, knowledge, motivation, and environment, *Accounting, Organizations and Society*, 425-450.

Moser, D. V. (1992) Does memory affect judgment? self-generated versus recall memory measures. *Journal of Personality and Social Psychology*, 555-563.

Nelson, M. W. (1993). The effects of error frequency and accounting knowledge on error diagnosis in analytical review. *The Accounting Review*, 803-824.

Nelson, M. W. (1994). The learning and use of frequency information in auditing. *Journal of Accounting Literature*, 185-211.

Nelson, M. W., Libby, R., & Bonner, S. E. (1995). Knowledge structure and the estimation of conditional probabilities in audit planning. *The Accounting Review*, 27-47.

Rosenthal, R., & Rosnow, R. (1985). *Contrast analysis: focused comparisons in the analysis of variance.* Cambridge, England: Cambridge University Press.

Sherman, S. J., McMullen, M. N., & Gavanski, I. (1992). Natural sample spaces and the inversion of conditional judgments. *Journal of Experimental Social Psychology*, 401-421.

Sneath, P. H., & Sokal, R. R. (1973). *Numerical Taxonomy.* San Francisco: W. H. Freeman.

Stevens, S. S. (1975). *Psychophysics: introduction to its perceptual, neural, and social aspects.* New York: John Wiley and Sons.

Tubbs, R. M. (October 1992). The effect of experience on the auditor's organization and amount of knowledge. *The Accounting Review*, 783-801.

Waller, W. S. (1990). *Auditors' probability assessments: a constructive perspective.* Unpublished monograph, University of Arizona.

Weber, R. (1980). Some characteristics of the free recall of computer controls by EDP auditors. *Journal of Accounting Research*, 214-240.

Winer, B. J. (1971). *Statistical principals in experimental design.* 2nd ed. New York: McGraw Hill.

Wright, A., & Ashton, R. H. (1989). Identifying audit adjustments with attention directing procedures. *The Accounting Review*, 710-728.

Zacks, R. T., Hasher, L., & Sanft, H. (1982). Automatic encoding of event frequency: further findings. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 106-116.

APPENDIX

Example screens from experiment

Panel A: Example of instruction screen for audit objectives

Instructions

Audits are structured to examine specific audit objectives for each account to find errors that would result in misstated financial statements. Audit objectives provide a helpful framework because errors that violate the same audit objective are related in that they have similar causes and normally can be detected by the same general types of audit tests. Because each test applied to a specific account can provide evidence concerning more than one type of processing error that violates the same objective, testing objectives allows auditors to expend less effort than they would expend if they tested separately for all the types of processing errors that violate the same objective. All audit tests are designed to examine a specific audit objective.

Given the importance of audit objectives, you will be asked to learn the basic characteristics of three audit objectives. The objectives are:

Proper cut-off
Validity
Valuation

Each objective will be described in turn. In addition, several examples of financial statement errors that violate each objective will be provided. Your goal is to learn the meaning of each objective and the examples of financial statement errors that are related to each objective. You will be tested later to determine whether this goal has been achieved.

Objective 1: Proper cut-off. The audit objective of "Proper Cutoff" relates to ensuring that only the current period's transactions are reflected in the current period's financial statements. Errors which violate the Proper cut-off objective occur when either next period's transactions are recorded in this period or this period's transactions are recorded in the next period. Tests of this objective typically involve examining the documents supporting transactions made at the beginning and end of the accounting period. Examples of errors that relate to the Proper cut-off objective are:

> "Next period's sales were included in the current year's revenue and receivables.

This error violates the Proper cut-off objective because sales were included in the wrong year. As a consequence, this period's sales and accounts receivable and next period's sales and accounts receivable will be misstated.

> "Raw materials were improperly shown as received after year-end."

This error violates the Proper cut-off objective because raw materials were included in the wrong year. As a consequence, the inventory and either the accounts payable or cash accounts will be misstated at the end of this year and at the beginning of next year.

> "Purchases of treasury bills were recorded in the wrong fiscal period."

This error violates the Proper cut-off objective because purchases of treasury bills were included in the wrong period. As a consequence, the investments and cash (or payable) accounts will be misstated at the end of this fiscal period and at the beginning of the next fiscal period.

Again, the three examples of errors that violate the Proper cut-off objective are:

> "Next period's sales were included in the current year's revenue and receivables."

> "Raw materials were improperly shown as received after year-end."

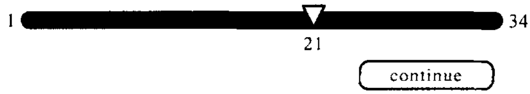
> "Purchases of treasury bills were recorded in the wrong fiscal period."

As you can see from these examples, the Proper cut-off objective is violated when either next period's transactions are recorded in this period or this period's transactions are recorded in the next period.

Panel B: Example of frequency test screen for audit objective valuation

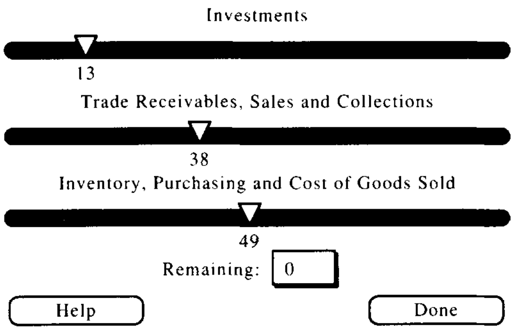
The audit objective of "Valuation" relates to ensuring that accounts shown on the financial statements are valued at the correct dollar amount. Errors which violate the Valuation objective occur when calculations are not arithmetically correct or when declines in the value of assets are not recognized. Tests of this objective typically involve reperforming calculations or determining that any declines in asset values are recognized.

How man times did you see Valuation errors on the list of errors to which you were exposed? (Remember, each repetition of a single error adds to the total number of Valuation errors to which you were exposed). Please respond by moving the slider to the appropriate number on a scale of 1-34.



Panel C: Example of audit decision screen for transaction cycles

Allocate hours to each transaction cycle by clicking on the slider that corresponds to the cycle and dragging the slider to the appropriate number of hours. Note that you can only allocate a total of 100 hours. If you decide to change your allocation, you can take some hours away from one transaction cycle to enable you to give more to another.



Panel D: Example of frequency test "hint" for audit objectives

Following are the three errors with which you were presented previously that violate the Proper cut-off objective.

1. "Next period's sales were included in the current year's revenue and receivables."

   This error violates the Proper cut-off objective because sales were included in the wrong year. As a consequence, this period's sales and accounts receivable and next period's sales and accounts receivable will be misstated.

2. "Raw materials were improperly shown as received after year-end."

   This error violates the Proper cut-off objective because raw materials were included in the wrong year. As a consequence, the inventory and either the accounts payable or cash accounts will be misstated at the end of this year and at the beginning of next year.

3. "Purchases of treasury bills were recorded in the wrong fiscal period."

   This error violates the Proper cut-off objectives because purchases of treasury bills were included in the wrong period. As a consequence, the investments and cash (or payable) accounts will be misstated at the end of this fiscal period and at the beginning of the next fiscal period.

Please think about why each of the three examples are errors that violate the Proper cut-off objective. When finished, please click on "continue".