RISK REDUCTION AND THE AUDIT REVIEW PROCESS

Vincent E. Owhoso
Assistant Professor
Bentley College

William F. Messier Jr.
Deloitte & Touche LLP Professor
Georgia State University.

John G. Lynch, Jr.
Hanes Corporation Foundation Professor of Business Administration
Duke University

March 1999
Revised: January 2000

Please address all correspondence to:

Professor William F. Messier, Jr.
Deloitte & Touche Professor
School of Accountancy
J. Mack Robinson College of Business
Georgia State University
P. O. Box 4050
Atlanta, GA 30303-4050
Office: 404-651-4465
Fax: 404-651-1033
Email: accwfm@langate.gsu.edu

The first author would like to thank the KPMG Foundation for their support of his dissertation research. We are grateful to Bernie Milano, the industry specialists who assisted with the design of the case materials, and the seniors, managers, and partners who participated in the study. We thank Mike Bamber, Don Jones, Richard Leftwich, Brian Mayhew, Christine Moorman, Steve Salterio, Mark Taylor, and two anonymous reviewers for their helpful comments on the paper. Earlier versions of the paper were presented at the 1999 International Symposium on Audit Research, 1999 National AAA meeting, and workshops at Georgia State University, University of Nebraska, and the University of Texas. Participants in those forums also provided helpful comments. Special thanks to Ken Trotman for his detailed and insightful comments.

RISK REDUCTION AND THE AUDIT REVIEW PROCESS

1. Introduction

The audit review process is a critical part of the audit and a major means of controlling the quality of the audit work (AICPA [1999], Bamber and Bylinski [1982], Solomon [1987]). Rich et al. [1997b], in their review of the literature, point out that more research is needed on the *effectiveness* of the audit review process. We examine the effectiveness of the review process to reduce risk by seeding errors[1] in two cases (banking and health care) and have banking and health care industry audit specialists review the work papers in senior-manager teams. Our subjects are designated industry specialists by their firm and they possess significant practice experience in their respective industries. We define risk reduction as the identification of the seeded errors by the reviewers and we use an independent set of industry experts (partners) to rate the seeded errors by type.

This paper extends prior research on audit review process in three significant ways (see section 2 for more detail). First, we have seniors and managers complete a case individually and as members of a team. This allows us to examine the *individual* contribution of the seniors and managers to the teams. Second, by comparing nominal and real teams, we show how the effectiveness of the audit team is less than the sum of the effectiveness of the individual members. Third, we answer the call by Solomon et al. [1999] to test whether industry knowledge leads to better *performance* by examining whether industry knowledge results in an increased ability to detect errors in a hierarchical review.

We find that managers add value to the audit *team* only in detecting conceptual (v. mechanical) errors *within* their industry specialization. They add much less in detecting mechanical errors even when working within their industry specialization (seniors detect more mechanical errors) and to detecting conceptual errors out of their industry specialization. We also find that real teams underperform their potential (as measured by nominal team performance) when working in their industry specialization for

[1] We used the terms errors and misstatements interchangeably throughout the paper.

mechanical errors, but not conceptual errors. Out of specialization, there is process loss for both mechanical and conceptual errors. Our results suggest that the process loss is due to output interference and not motivational factors. We conclude that sequential review is an effective structure for specialized industry teams, but that an independent (simultaneous) review structure would be more effective when audit teams are *not* organized by industry specialization. The fact that both the senior and manager contribute to risk reduction supports the effectiveness of current review practices in which teams work within industry specialization.

The remainder of the paper proceeds as follows, In section 2, we present the background for the study, including a brief review of the literature and the hypotheses tested. The methods used to test the hypotheses are described Section 3. Section 4 presents the results of the main study. Section 5 discusses the results, limitations, and areas for future research. Appendix A presents the results of the error rating task.

2. Background and Hypothesis Development

2.1 PRIOR RESEARCH

The audit review process is characterized as an hierarchical, sequential, and iterative process where evidence is gathered and evaluated by a subordinate and then reviewed by the next superior in the hierarchy (Rich et al. [1997a], Solomon [1987]). The objective of work paper review by a superior team member is to provide the support for the auditor's report and to aid the auditor in the conduct and supervision of the audit (AICPA [1999, AU 339.02]).

Rich et al. [1997a, b] and Solomon [1987] review prior research into the audit review process. This research has examined: individual judgments by staff, seniors, or managers (e.g., Ramsay [1994], Trotman [1985], Trotman and Yetton [1985]), composite judgments (nominal team) prepared from the individual judgments of a staff and a senior, two seniors, or a senior and manager (e.g., Bamber and Ramsay [1997], Harding and Trotman [1999], Trotman [1985]), interacting group judgments composed of two seniors (e.g., Trotman [1985], Trotman and Yetton [1985]), judgments of a hierarchical group with

interaction between a senior and manager (e.g., Trotman [1985]), and judgments of a hierarchical group without interaction between a senior and manager (e.g., Trotman and Yetton [1985]).

Our work extends prior research in the following ways. First, we compare the judgments of a nominal team composed of a senior and manager with the judgments of a senior and manager working in a real team. This comparison allows us to determine how the senior and manager contribute to the team's performance. Second, we use a more realistic representation of the hierarchical, sequential, iterative nature of the review process by having a senior review the staff, and then have the manager review the senior. Review notes are prepared at each stage of review. Third, we develop the dependent variable based in the auditors' review notes. As Rich et al. [1997b, p. 117] point out, review notes provide "a natural and externally valid measure of reviewer decision/action selection." Prior studies (e.g., Bamber and Ramsay [1997], Harding and Trotman [1999], Ramsay [1994]) used a true/false test to classify the auditors' detection of errors. This type of task can be classified as a recognition test while detection of errors based on review notes can be classified as a performance test (Rich et al. [1997b]). Fourth, we use different definitions for mechanical and conceptual errors than Ramsay [1994] and Bamber and Ramsay [1997], and we had the seeded errors classified by an independent, experienced group of partners specializing in each industry. Last, we extend the recent work by Solomon et al. [1999] on the effect of industry specialization to the review process by testing whether industry knowledge leads to better team *performance*.[2] Prior work on industry specialization examined only the link between experience and knowledge.

2.2 TEAM PROCESSING

2.2.1 Incremental Performance of Nominal Teams v. Individual Members. When individuals are combined into a review team, one would expect team performance to outstrip that of the individual

[2] Similar to Solomon et al. [1999], we assume that there is a link between experience and knowledge (Libby [1995]). We also assume that this knowledge leads to better performance (Libby and Luft [1993]). We did not directly test the auditors' knowledge, but relied on the public accounting firm's classification of the auditors as industry specialists as an indicator of their knowledge (Gramling and Stone [1999]).

members (Hill [1982], Solomon [1987], Rich et al. [1997a], Trotman et al. [1983], Trotman and Yetton [1985]). The literature indicates that the superior performance of teams over individuals is context specific. For example, when there is an overlap of knowledge, the group may not outperform the most knowledgeable member of the team. In an audit setting, if a senior's review of a staff member's work detects almost all errors then the manager's review of the senior's work would be unnecessary, and results in an *inefficient* audit. On the other hand, if the manager's review of the senior's review work detects additional errors, then the manager's review results in a more *effective* audit. Ramsay [1994] stated that because of the nature of the senior's review versus the manager's review, each is likely to focus on different types of errors. He argues that the senior is responsible for the detailed review of the work papers prepared by the staff with an emphasis on technical accuracy and completeness. The manager is normally responsible for performing a general review that focuses on the adequacy of the overall audit work, including the adequacy of the senior's detailed review. The foci of the manager's review is generally the more complex and subjective issues related to the area being audited. Ramsay [1994] and Bamber and Ramsay [1997] find that seniors detect more mechanical errors when reviewing work papers while managers detect more conceptual errors.[3] However, their results are based on seniors and managers completing their reviews as individuals. We create a *nominal* team from the *paired* seniors and managers working individually[4] (see section 3.5) as follows:

$$\text{Probability of nominal team detection} = P(\text{Senior Detects}) + P(\text{Manager Detects}) - P(\text{Senior Detects and Manager Detects}) \quad (1)$$

The following hypotheses related to team processing are then tested:

H1: The *incremental* performance (risk reduction) of the nominal team over the senior working individually will be greater for conceptual errors than for mechanical errors.

[3] Though this was not a primary goal of our research, the design used in this study allowed us to directly test the findings of Ramsay [1994] and Bamber and Ramsay [1997]. We expect to replicate those basic findings; i.e., we expected that seniors working individually will detect more mechanical errors than managers working individually, and the reverse to be true for conceptual errors. However, as we show later, their findings are moderated by industry specialization.

[4] Bamber and Ramsay [1997] also create nominal teams. However, their teams are not matched, but are created *ex post* by randomly pairing 50 teams of seniors and managers from each treatment condition.

H2: The *incremental* performance (risk reduction) of the nominal team over the manager working individually will be greater for mechanical errors than for conceptual errors.

"Incremental" performance of the team over either senior or manager alone is simply the probability that the second person would detect the error minus the probability that both detect it. Either member's contribution to the team is greater when their individual contributions are nonoverlapping.

2.2.2 Real Team Performance Relative to Team Potential. A second benchmark compares real team performance to the "potential" performance of the group, as measured by nominal teams. A real team can do better than a nominal team if each member of team has only part of the expertise necessary to detect certain errors, as in the group problem solving literature on "hidden profiles" (Stasser and Stewart 1992; Stewart and Stasser, 1995). However, with real, interacting groups, the general result in the problem solving literature is for real groups to underperform relative to such a nominal team benchmark (Taylor et al. [1958]). In an audit review setting, Bamber [1983] and Messier and Tubbs [1994] show that reviewing a subordinate's work can affect the reviewer's judgments. We hypothesize that when the manager in the real team reviews the senior's work papers that the review will inhibit the manager's ability to detect errors.[5] We test the following hypothesis:

H3: Nominal teams will outperform real teams in reducing risk.

Hearing what one person says tends to inhibit the ability of other members to produce non-overlapping insights that they would have been able to produce if each group member worked independently. Diehl and Stroebe [1987] argue that the shortfall of (simultaneous, not sequential) real v. nominal groups is reliable and is attributable to "production blocking." It is unclear whether this production blocking is due to motivations, memory interference as in "output interference" (Hoch [1984]), or the fact that one has to wait to speak in real interacting groups. Waiting to speak is not an issue in the sequential review structure we examine. However, motivation and output interference are

[5] Frederick's [1991] "part-cue listing" results would also support a similar prediction. In his study, a partial list of errors inhibited auditors' ability to recall other errors.

plausible reasons to expect real teams to fall short of the nominal team benchmark when the group members do not interact.

We are able to diagnose the relative contributions of motivational and output interference on memory by the pattern across our experimental conditions of support for H3. The manager may believe that s/he should concentrate on conceptual errors when operating in sequential (real) teams, but not when auditing a case as an individual. This is a motivational factor that would lead one to predict a variant of H3:

> H3A: Nominal teams will outperform real teams in reducing risk for mechanical but not for conceptual errors, independent of whether teams are working in or out of specialization.

Alternatively, output interference in memory rather than lack of motivation may cause managers to add less to real team output than they do for nominal teams. The manager in a real team reads the senior's work papers, providing a set of "cues" not present when the same manager works as an individual. The memory literature in psychology shows that cues facilitate retrieval by allowing access to new categories in memory that would not have been generated spontaneously without those cues. But once a category is activated, cueing more of its members only inhibits one's ability to generate unmentioned members. One keeps on cycling through the recently cued paths, making it harder to reach new parts of the network. Thus, whether seeing n cues hurts or helps is a function of the balance of (1) a facilitating effect when the senior's work cues a category the manager would have otherwise forgotten and (2) an inhibiting effect when the senior's work simply cues more members of a category the manager would have retrieved anyway. When the latter effect dominates, "output interference" is observed. Alba and Chattopadhyay [1985] showed that output interference increases with the number of intra-list cues provided, but only when expertise is low and the respondent has only a few, coarse categories in the knowledge domain. Because experts have finer subcategories, it is more likely that a set of n cues will cue different categories for them than for their less expert counterparts. In the context of a sequential review process, the effect of reading the senior's work is a function both of the number of cues (errors

detected) therein and whether they are overlapping with categories the manager would have generated without the senior's work. The implication of this is that viewing a senior's work is least likely to inhibit a manager's contribution to a real team when the senior detects few errors of a given type and when, holding constant the number of errors detected by the senior, the manager has a finer category structure. This situation holds in our experiment for a manager's ability to detect conceptual errors working within specialization This implies a variant of H3:

> H3B: Nominal teams will outperform real teams in reducing risk except for conceptual errors when the team is operating within its area of specialization.

Like Hypothesis 3, Hypotheses 1and 2 are likely to be affected by whether the individuals or teams are working in or out of their industry specialization. This issue is addressed in the next section.

2.3 ERROR DETECTION BY INDUSTRY SPECIALISTS

Empirical evidence supports the expectation that auditors who acquire training and/or task-specific experiences in specialized areas are likely to exhibit superior performance in those specific areas (see Gramling and Stone [1999] for a review). Solomon et al. [1999, p. 194] point out that most prior studies that have examined the effect of industry-related experience-judgment performance have *not* been limited to designated industry specialists.[6] We use auditors who have been designated as specialists in banking and health care by their firm and we examine whether this industry specialization leads to better performance. We expect, based on the above analysis, that auditors working as individuals will detect more misstatements working in their specialization than when working outside of their specialization. It is an open question, though, whether the incremental contribution of the manager to the senior's work will be greater in or out of specialization. We noted in discussing Equation 1 that the manager's incremental contribution to the team is greater when the manager detects errors that the senior fails to detect. The probability that the senior will fail to detect an error is higher out of specialization than in specialization, but it is also true that the manager may be less likely to detect errors missed by the senior when out of

[6] Two notable exceptions are Solomon et al. [1999] and Taylor [1998].

specialization. The net effect of specialization may be a matter of the calibration of the relative strength of its effects on these two component parts of team performance.

3. Method

3.1 OVERVIEW AND EXPERIMENTAL DESIGN

We test the research hypotheses by administering a set of two cases to differentially experienced auditors (seniors and managers) from two industries (banking and health care). The two industries were chosen because they are specialized industries with their own peculiar and unique services and products, accounting rules, and regulatory requirements (Solomon et al [1999]). We seeded each case with errors that were classified as mechanical or conceptual, and the key dependent variables to be reported are the percentages of the seeded mechanical and conceptual errors detected by the auditor or team in each case. We conducted a separate study with audit partners from the banking and health care industries to verify the classification of errors as mechanical or conceptual. The details of that study are presented in Appendix A. One Big 5 firm provided subjects for both the rating task and the main experiment.

Seniors and managers were assigned to 2-person (banking or health care) teams, and the senior-manager team is the unit of analysis in the designs reported below. Each team worked on both cases – one in their area of specialization and one outside of their area of specialization. One case was completed "Alone" by the two members as individuals and one as members of a "Team." The "Alone" case was completed independently by the senior and manager, without referring to each other's work. For the "Team" case, the manager reviewed the work papers that the senior had completed earlier. For half of the senior-manager teams, the Alone case was performed in specialization and the Team case out of specialization and for the other half the Alone case was performed out of specialization and the Team case in specialization. We also counterbalanced the order of the two cases. Figure 1 shows the experimental design.

[Insert Figure 1 here]

The overall design is 2x2x2x2x2x2 mixed factorial design with three between-subjects factors and three repeated measures factors. The three between-subject factors were Industry Specialization (Banking or Health Care), Task-Pair (Individual In-Specialization, Team Out-of-Specialization v. Individual Out-of-Specialization, Team In-Specialization), and Order of Cases (Alone, Team v. Team, Alone). The three repeated measures factors were Review Hierarchy (Senior v. Manager), Team Processing (Alone v. Team), and Error Type (Mechanical v. Conceptual). The Error Type and Order of Cases factors are not shown in Figure 1.[7] Nine review teams were randomly nested in each of the 8 between-subjects conditions for a total of 72 senior-manager teams.

Though Figure 1 depicts the overall design, we will report a system of planned contrasts on parts of the overall experimental design to test our hypotheses. When this approach is followed, the repeated Team Processing factor drops out of the design and the "Task Pair" factor becomes a manipulation of In v. Out of Specialization. We will elaborate the analysis design for each hypothesis test in the results section.

3.2 CASE MATERIALS AND EXPERIMENTAL PROCEDURES

We presented two cases to the subjects involving the review of work papers in accounts receivable and related accounts for a hypothetical hospital, and the audit of loans receivable and loan loss reserves for a hypothetical bank.[8] The accounts receivable or loans receivable subsystems were chosen because they require specialized knowledge, including regulatory requirements, general economic conditions, industry conditions, credit policies, customer characteristics, future prospects, and other business risks. We consulted auditing and related text books, along with academic, practice, industry-specific, and trade journals. In addition, meetings were held with industry specialists in health care and banking audits at the Big 5 accounting firm that provided the subjects, and accounting executives of a large hospital and a local

[7] The "Order of Cases" factor was included to prevent any order effects due to always receiving one case first. This factor was not significant ($p > .05$) in any of the analyses and is dropped from further discussion.

[8] Copies of the case materials can be obtained from the first author.

bank. These sources provided the resources and information needed for developing the cases. The participating Big 5 firm provided work paper information from a hospital and a bank engagement. The figures in the original work papers were modified or scaled by a factor in order to mask the original amounts and to allow seeding of the errors.

Each set of case materials included a cover letter which described the purpose of the study and consent letter required by the University of Florida. The second page presented the manipulation of the review process and additional instructions for completing the cases. The subjects were told to spend about half of the allocated time on each case. Each case was placed in a separate envelope, and labeled booklet 1 and booklet 2 to control the order that participants completed the cases.

Each case consisted of three parts. The first part provided the instructions for completing the case and background information about the hospital or bank. This background information included an overview of the company's accounts receivable or loan receivable, audit plans, and the firm's financial data (i.e., the balance sheet and income statement for two years). The senior (manager) subjects were instructed to assume the role of an in-charge (engagement manager) for the current audit and to read the background information about the hospital or bank. The subjects were instructed to assume that any work paper not included was properly prepared and references to other work papers were correct. They were also told that their work may be reviewed by an audit manager for the senior subjects and by a partner for the manager subjects. This statement was made in order to simulate an actual audit. After reading the company's background information and the financial statements, the subjects provided a preliminary judgment about the likelihood of material misstatement in the financial statements and the likelihood that the subordinate detected such misstatement. The subjects then wrote down the time they began the review of the work papers.

The second part of the case materials included the work papers. The subjects reviewed the work papers and used a pre-printed review note form to record their review notes. The subjects were able to refer back to prior pages of the case materials (similar to a normal review process) and could prepare any

type or form of review notes.[9] After reviewing the work papers the subjects were asked to provide a final assessment on the likelihood of any remaining material misstatement and to rate whether the subordinate detected the misstatements. They also recorded the time the review of the materials was completed.

The third part of the case material requested demographic information about gender, age, rank, total audit experience, specialized industry experience, other experiences in current or other specializations, ability to conduct audits in health care and banking, the realistic nature of the case, their reliance on the subordinate's expertise and specialization, and level of effort expended based on the reviewer's perception of the subordinate's specialized knowledge.

3.3 SEEDED ERRORS AND CASE EVALUATION

We identified the seeded errors by consulting various auditing and related text books along with academic, practice, industry-specific, and trade journals. Meetings were also held with industry specialists in health care and banking audits from the Big 5 firm, and account receivable and billing executives of a large hospital and a local bank. Based on these meetings and the work paper information provided by the CPA firm, a list of 24 errors was developed: 13 health care errors and 11 banking errors. Table 1 contains a description of the seeded errors.[10] We seeded the errors in the work papers to create a rich situation for investigating the effectiveness of the review process to detect material misstatements. The errors varied in their importance and conceptuality. Many of these errors are industry-specific and only auditors with specialized industry knowledge are likely to detect them. The seeded errors were classified by an independent group of audit partners (see Appendix A) as to their importance and whether they were mechanical or conceptual errors.[11]

---

[9] If there was a potential disadvantage in allowing the reviewers to review the work papers and write review notes in any format they chose, this disadvantage is mitigated by the fact that a pre-determined number of material errors were seeded in the cases. Thus, an effective review should identify the material errors and the format in which a reviewer identifies the seeded errors should not become a critical issue.

[10] As discussed later, one error type from each industry was not included in the final data analysis.

[11] Ramsay [1994] defined mechanical errors as "objective, verifiable, and concrete" and conceptual errors as "subjective, unverifiable, and imprecise." He coded the errors as mechanical or conceptual using some AICPA

[Insert Table 1 here]

Each case was reviewed by either a health care and banking industry specialist at the Big 5 firm for relevance, appropriateness, and realism. After some minor revisions, the cases were judged to be realistic, and the errors appropriate and relevant. To provide an additional check on the generalizability of the case materials, each case was sent to an audit partner specializing in the respective industries at another Big 5 firm. These partners proposed minor changes that were communicated to the original industry specialists and incorporated into the case materials.[12]

3.4 SUBJECTS

The subjects who participated in the main study included 144 auditors (72 audit seniors and 72 managers) from 42 offices of the Big 5 firm. One half of the seniors and managers specialized in hospital audits while the other half specialized in bank audits. Table 2, Panel A contains information on the subjects' mean months of total audit experience and specialized audit experience.

[Insert Table 2 here]

A partner from the executive office of the firm provided a health care audit specialist and a financial services audit specialist to assist with the selection of subjects for the study. After discussing with the firm specialists[13] concerning the subject requirements for the study, only banking and health care

---

guidance and then had two accounting professors and a doctoral student confirm his coding. We used the following statements in our survey of the industry specialists (see Appendix A): "Mechanical errors are errors that require little or no subjective judgment on the part of the auditor. Judgmental errors, on the other hand, require significant subjective judgment on the part of the auditor."

[12] The case materials were pilot tested using six doctoral students, two college professors and two audit seniors in a Big 5 firm. Four of the doctoral students had experience as audit seniors. One of the doctoral students had experience in banking audits while another had experience in hospital audits. The audit seniors had experience in either banking or health care. The two college professors had prior experience in bank audits. Based on the pilot testing, a minor change was made to the rating scales utilized for the post-study questionnaire.

[13] These industry specialists are different from the specialists that helped with the case development. These specialists also provided contacts at various offices of the firm for obtaining the subjects.

auditors participated in the study.[14] Criteria for selecting auditors for participation in this study were that they were designated as specializing in banking or health care audits and their principal work assignment was banks or hospitals (Gramling and Stone [1999]). The demographic data in Table 2 indicates that the subject pool was highly experienced in their respective industries.

3.5 CONSTRUCTED REVIEW TEAMS

The review teams were constructed from the pool of senior and manager subjects. A constructed team has features that are identical to a real audit team in which a subordinate is reviewed in his or her absence (no interaction between the senior and manager occurred).[15] One purpose of using constructed teams was to preserve the hierarchical structure of the review process in an experimental team setting.

Seventy-two dyad teams were constructed with 36 teams in each industry. Each dyad team consists of a randomly matched senior and manager from the same industry specialization. For example, if a senior was in the treatment condition that required completing the hospital case individually and the banking case for review, only the banking case was reviewed by a manager. A copy of the senior's bank case materials was made, and the manager level instructions and post-study questionnaire were substituted for the relevant pages of the bank case materials. The manager was told that the staff member that was reviewed by the senior had one year of bank audit experience and that the senior had three years of banking experience. The manager would then complete the health care case individually and review the bank case as part of the team. This method allowed us to match and track the completed work of a senior and manager in the same team.

[14] Insurance, real estate, mutual funds and investment auditors are included in the firm's financial services group. Similarly, medical and pharmaceutical companies are included in the firm's health care group. Given the nature of the cases, such subjects were not included in this study.

[15] The economic resources required to have a manager perform a face to face review with a senior were enormous and the firm was not willing to commit that level of resources for the experiment.

3.6 ADMINISTRATION

The experimental instrument was administered first to the seniors and then their instruments were randomly matched with a manager specializing in the same industry. Thirty-eight banking and 30 health care seniors completed the instrument at group sessions that were arranged by the firm. An additional 25 banking and 27 health care senior responses were received from other offices by mail because it was not possible to organize onsite meetings. Seventeen banking and 14 health care audit senior responses were unusable because they were completed by seniors without requisite industry experience.[16] The instructions for the study were the same for both the onsite and mail administration, except that a sentence in the cover letter for the mail administration requested the subjects to return the completed materials directly to the researchers. The time allowed to complete the cases at the senior level was 75 minutes. There was no statistical difference in the performance of the onsite and mail administration groups in the self-reported amount of time spent in completing the cases.

One hundred eighty instruments were mailed to a randomly selected group of 90 banking and 90 health care audit managers. For the banking subjects, contact persons at the office of the firm distributed the materials and coordinated their return. For the health care subjects, the firm provided a list of names and addresses, and the instruments were mailed directly to the subjects. Forty-three completed instruments were received from banking managers and 41 were received from health care managers. Three banking and 2 health care instruments were unusable because they were completed by audit seniors. Two health care cases were completed by non health care specialists and one manager did not provide industry demographic data.[17] Consistent with the practice in an actual audit setting where the

[16] The 31 unusable banking and healthcare senior instruments came from two regional offices in which the primary health care and banking contacts delegated another person to distribute the instruments to seniors. It appears that this person distributed instruments to any senior and staff auditors in the region. Thus, these instruments are unusable because they were not completed by industry specialists.

[17] The design used in this experiment calls for an equal number of subjects in each cell and matched pairs of seniors and managers. Because we received only 36 usable health care manager instruments we deleted 13 subjects for the following reasons: (1) 2 healthcare and 3 banking seniors because the matched manager did not return the instrument and (2) 4 banking teams (4 seniors and 4 managers) to equalize cell sizes.

detailed audit work consumes more time than review of the completed work, the time allocated for the managers to complete the review was 60 minutes. Table 2, Panel B presents the self-reported time to complete the cases.

3.7 DEPENDENT VARIABLE

We define risk reduction as any critical misstatement or consequential omission in the client's financial statements or in the audit work papers[18] prepared by a subordinate that are detected by another auditor (usually a higher hierarchy auditor) who is reviewing the financial statements or work papers. Thus, the dependent variable reported in this paper is percentage of errors detected. For each of the two cases, we calculate the percentage of mechanical and conceptual errors (0-100%) detected by that individual. The bases for these percentages varied. For the banking case, there were 5 mechanical errors and 3 conceptual errors. For the hospital case, there were 4 mechanical errors and 3 conceptual errors. So, for example, if an auditor detected 3 of the seeded mechanical errors and 2 conceptual errors in the banking case, the risk reduction score would be 100*(3/5) = 60% for mechanical errors and 100*(2/3) = 67% for conceptual errors.

3.8 CODING OF THE REVIEW NOTES

The review notes were coded by a former audit manager specializing in banking from the firm that provided the subjects and an adjunct professor of accounting at a local university. The professor had a doctoral degree in accounting, and teaching and research interests in financial and international accounting. Neither of the coders had work experience in the health care industry, and they were not aware of the research hypotheses prior to coding the review notes.

The coders were provided with copies of the test instruments that listed the errors seeded in the case, each subject's review notes, and a set of case materials. The coders were instructed that the subjects were not expected to write the exact words and that as long as the subject's review notes raised significant

[18] This definition does not include "technical errors," such as working paper presentation or format errors that only affect the appearance of the work papers and not the risk of failing to detect material misstatements.

questions or context issues about a specific error type, the subject is said to have detected that error. Identified items that might represent other errors that were not seeded were not included in determining the level of risk of reduction.

The agreement between the two coders on the classification of the errors was 96 percent. The areas of disagreement involved bank error number 11 and health care error number 10. There was no attempt to reconcile the coders on these two items; rather, the two errors were excluded from the analyses because the partners participating in the error rating task (see Table 1) disagreed on whether these error were material errors or just housekeeping errors.

4. Results

Prior to presenting the hypotheses tests, we provide descriptive data on the detection rates for individual seeded errors and provide tests of the individual data.

4.1 DESCRIPTIVE DATA ON DETECTION OF SEEDED ERROR TYPES

Table 3 shows the detection rates of the seeded errors for each treatment condition and by industry specialization. The overall detection rate for the banking specialists working in banking is 44.2% (Panel A) while the rate for the health care specialists working in health care is 46.5% (Panel B). These rates are lower than prior studies (Ramsay [1994], Bamber and Ramsay [1997], Harding and Trotman [1999]) which reported overall detection rates of between 61% to 84%. Our rates may be lower because we used the review notes to code detected errors whereas prior studies used a true-false test. Because the true-false test presents the subjects with the seeded errors those rates are likely to overstate the "true" detection rate. The detection rates for the banking and health care specialists working outside their specialization are considerably lower (15.3% and 10.9%, respectively).

[Insert Table 3 here]

4.2 RISK REDUCTION - INDIVIDUAL DATA

It is necessary to demonstrate the expected patterns of *individual* performance before proceeding to our formal hypothesis tests of team performance. We expect to replicate the findings of Ramsay [1994]

and Bamber and Ramsay [1997] that seniors working individually are better able to detect mechanical errors while managers are better able to detect conceptual errors. Additionally, we expect that when auditors work as individuals that they will perform better in their industry specialization than when they work out of specialization (Johnson and Jamal [1988], Solomon et al. [1999]).

We test these relationships by analyzing data from the Alone half of the overall design in Figure 1 in a 2x2x2x2x2 mixed-factor ANOVA (not presented) with Industry Specialization, Order of Cases, and In v. Out of Specialization as between-subjects factors, and Hierarchy and Error Type as repeated measures factors. As predicted, there was a significant Hierarchy*Error Type interaction [$F(1,64) = 41.95$, $p < .0001$] and there was a significant main effect of In v. Out of Specialization [$F(1,64) = 127.60$, $p < .0001$].

Both effects, however, were moderated by a significant three-way interaction of Hierarchy*Error Type*In/Out of Specialization [$F(1,64) = 18.58$, $p < .0001$]. The data for the seniors and managers are shown in Figure 2. We analyze this interaction by examining mechanical and conceptual errors separately. For mechanical errors, the simple interaction of Hierarchy*In/Out of Specialization is significant [$F(1,64) = 4.36$, $p < .05$]. Follow-up tests showed that seniors are better (53.3%) than managers (44.7%) at detecting mechanical errors in specialization [$F(1,32) = 5.15$, $p < .05$]. However, out of specialization, there is no significant difference between seniors' (11.4%) and managers' (12.2%) ability to detect mechanical errors. Similar simple effects tests for conceptual errors revealed a significant simple interaction of Hierarchy*In/Out of Specialization [$F(1,64) = 13.95$, $p < .001$]. The simple-simple effect of Hierarchy shows that managers are better (51.9%) than seniors (13.9%) at detecting conceptual errors in specialization, [$F(1,32) = 38.62$, $p < .001$]. Out of specialization, managers are still better (19.4%) than seniors (9.3%) [$F(1,32) = 5.76$, $p< .05$], but by a narrower margin. Thus, the degree of superiority of managers over seniors in detecting conceptual errors is moderated by their degree of industry-specific expertise. The three-way interaction of Hierarchy*Error Type*In/Out of Specialization

is not moderated by the background factors such as Industry Specialization.[19] Thus, the findings of prior studies replicate In but not Out of Specialization (e.g., Ramsay [1994], Johnson and Jamal [1988]).

[Insert Figure 2 here]

4.3 INCREMENTAL PERFORMANCE OF NOMINAL TEAMS OVER INDIVIDUALS

The first two hypotheses were tested by analyzing the data in a 2x2x2x3x2 mixed-factor ANOVA with Industry Specialization, Order of Cases, and In v. Out of Specialization as between subjects factors, and Hierarchy (senior, manager, nominal team) and Error Type as repeated measures factors. Planned contrasts of the seniors v. the nominal teams tell the incremental value of the manager while the planned contrasts of the managers v. the nominal teams tell the incremental value of the senior. The results are shown in Table 4 and Figure 2.

[Insert Table 4 here]

4.3.1. Incremental Value Of The Manager. Incremental value of the manager can be seen in each panel of Figure 2 as the difference between the 3rd (Nominal Team) and 1st (Senior Alone) bars. The key factor in Table 4 is the three-way interaction of Hierarchy*Error Type*In/Out of Specialization [$F(2,128) = 9.30$, $p < .000$]. We assessed the incremental contribution of the manager by the contrast of seniors v. nominal teams. Overall, nominal teams detected more errors than seniors alone [$F(1,64) = 32.96$, $p < .001$]. However, this was moderated by a significant three-way interaction of the senior v. nominal Team*Error Type*In/Out of Specialization [$F(1,64) = 9.10$, $p < .005$]. H1 predicted that the advantage of the nominal team over the senior should be greater for conceptual than for mechanical errors. We found that this hypothesized effect was observed In Specialization [$F(1,32) = 11.24$, $p < .005$], but not Out of Specialization [$F(1,32) = 1.31$, $p > .25$]. In Specialization, the advantage of the nominal team over seniors (i.e., the incremental value of the manager) was significantly greater for conceptual (55.6%-13.9%

[19] There was also a theoretically irrelevant main effect of industry. Health care specialists outperformed banking specialists, *averaging across performance on the two case*s. This effect is not theoretically significant, as we have no way of verifying that our hospital and banking cases were equally difficult for auditors working in specialization.

= 41.7%) than for mechanical errors (73.2% – 53.3% = 19.9%). Out of Specialization, however, the incremental value of the manager did not differ for conceptual (28.7% - 9.3% = 19.4%) v. mechanical (21.4% - 11.4% = 10%) errors. Thus, H1 is moderated by whether the auditors are working In v. Out of Specialization. Figure 2 lends some insight by showing how this advantage of nominal teams over seniors alone is affected by the percentage of errors detected by managers but not by seniors in their nominal teams. The manager detects 41.7% more conceptual errors in specialization; values are below 20% for mechanical errors or when working out of specialization. The manager adds little for mechanical errors in specialization because the probability is low that the senior will miss these errors. The manager adds little for both mechanical and conceptual errors out of specialization because, though the senior misses a high percentage of errors, there is a low probability that a manager will detect an error given that it was missed by the senior.

4.3.2. Incremental Value of The Senior. Incremental value of the senior can be seen in each panel of Figure 2 as the difference between the 3rd (Nominal Team) and 2nd (Manager Alone) bars. Analyses of the incremental value of the senior revealed that teams were clearly superior to managers [$F(1,64) = 148.6$, $p < .0001$]. Moreover, as predicted by H2, the advantage of the nominal teams over managers was greater for mechanical than for conceptual errors [$F(1,64) = 24.03$, $p < .0001$]. However, both of these effects were moderated by a significant three-way interaction of manager v. nominal Team*In/Out of Specialization*ErrorType interaction [$F(1,64) = 24.39$, $p < .000$]. In Specialization, the advantage of the nominal team over managers was significantly greater for mechanical (73.2 - 44.7 = 28.5%) than for conceptual errors (55.6 – 51.9 = 3.7%), [$F(1,32) = 55.5$, $p < .0001$]. Out of specialization, no such effect was found [$F(1,32) = 0$, $p>.50$]. Because of errors uniquely detected by the senior, the nominal team outstripped the manager by 9.2% of the mechanical errors and 9.3% of the conceptual errors. The errors caught by seniors out of specialization almost never overlapped the (few) mechanical errors caught by the managers. For this reason, seniors' incremental contribution to the nominal team is actually greater out of specialization than for conceptual errors in specialization. In the latter case, they performed better as

individuals, but the errors they caught were almost entirely redundant with those caught by their managers.

4.4 NOMINAL VERSUS REAL TEAMS[20]

H3 predicts that nominal teams will outperform real teams. Nominal team performance represents the team's potential if it caught any error recognized by one or both independent team members. The shortfall in performance of real teams is attributed to process loss when team members work non-independently. Either motivation is reduced (H3A) or "output interference" is created by seeing the senior's work (H3B). The extent of process loss can be seen in each panel of Figure 2 as the difference between the 3rd (Nominal Team) and 4th (Real Team) bars.

To compare the nominal teams with the real teams, a 2x2x2x2x2 mixed ANOVA (see Table 5) with Industry Specialization, Task-Pair, and Order of Cases as between subjects variables, and Nominal/Real Team and Error Type as repeated measures factors. In this design, the interaction of Task Pair*Nominal/Real Team reflects the main effect of working In v. Out of Specialization (see Figure 1).

[Insert Table 5 here]

If nominal teams outstrip real ones due to motivation and if managers see their role in a team review as handling the conceptual issues, we should see shortfalls for real teams in detecting mechanical errors but not for conceptual errors. In a real team, the manager assumes that the mechanical errors are "the senior's job, not mine" and so is less vigilant for such errors in a team than he or she would be when working as an individual. Moreover, the Nominal/Real Team*Error Type interaction should be independent of whether the case is in or out of specialization. That is, there should be a two-way interaction (Nominal/Real Team * Error Type), but no three way interaction. If the pattern is due to output interference, we should see a three-way interaction of Nominal/Real Team*Error Type*In/Out of

[20] We performed separate ANOVAs on the nominal team and real team data. The results show that nominal and real teams perform better in specialization, the nominal team data supports H1 and H2 in specialization, and the real team data supports H1 in specialization (H2 cannot be tested because there is no independent measures of which errors detected by the senior would have been detected by the manager).

Specialization. In specialization, for conceptual errors, there should be no simple effect of real v. nominal team. There should be a simple effect (advantage) of nominal teams over real ones for conceptual errors out of specialization and for mechanical errors both in and out of specialization.

Our results supported the latter hypothesis (H3B). As a main effect, nominal teams outperform the real teams [$F(1,64) = 8.26$, $p < .01$].[21] Contrary to H3A, there was no two-way Nominal/Real * Error Type interaction [$F(1,64) = 0.39$]. Consistent with the output interference explanation and H3B, there was a significant three-way interaction of Nominal/Real * Error Type * In/Out of Specialization [$F(1,64) = 3.90$, $p < .053$]. In Specialization, there is no advantage of nominal teams (M = 55.6%) over real teams (M=58.3%) when managers review seniors' work on conceptual errors. But, for mechanical errors, nominal teams outperform real ones (M = 73.2% v. 62.2%,). Out of specialization, nominal teams outperform real teams for both mechanical (M = 21.4% v. 13.6%) and conceptual errors (M = 28.7% v. 13.9%).

To follow up this significant three-way interaction, we tested for the simple interaction of Nominal/Real Team*In v. Out of Specialization separately for mechanical and for conceptual errors. For mechanical errors, the simple interaction was not significant [$F(1,64) = 0.26$, $p < .61$]. For mechanical errors, nominal teams were uniformly superior to real teams, as evidenced by a simple main effect of Nominal/Real, [$F(1, 64) = 8.39$, $p < .005$]. For conceptual errors, the simple interaction was significant [$F(1,64) = 4.56$, $p < .037$], demonstrating that there was process loss from team review out of specialization (M = 28.7% v. 13.9% for Nominal v. Real teams, [$F(1,64) = 6.46$, $p < .014$], but not in specialization (M = 55.6% v. 59.3%, [$F(1,64) = 0.23$, $p > .63$]). Thus, the results support H3B, predicting that the advantage of the nominal team over the real team will be significant except for conceptual errors,

[21] There were other significant effects that do not bear on our central focus here on the degree to which nominal teams outperform real ones. As in prior analyses, teams detect far more errors working In than Out of Specialization [$F(1,64) = 256.8$, $p < .0001$]. As in prior analyses, the advantage of working in specialization is greater for health care than for banking teams [$F(1,64) = 9.39$, $p < .004$]. As before, this effect was moderated by a marginal interaction with Error Type [$F(1,64) = 3.81$, $p < 056$]. It was stronger for conceptual than for mechanical errors, due to the relatively stronger performance of managers in health care in catching conceptual errors in specialization.

in specialization.

One other analysis supported our conclusion that "process loss" was not due to lowered effort by managers reviewing a senior's work compared to when auditing the original work papers as individuals. We analyzed the data on time spent on each case (Table 2, Panel B) as a function of the same design variables as in the Real v. Nominal Team analysis, except that we had no mechanical v. conceptual factor. Results showed a significant Hierarchy * Alone/Team interaction [$F(1,64) = 6.25$, $p < .015$]. Seniors spend the same time when working Alone or in Teams (M = 33.6 and 31.6 minutes, respectively), but managers spend significantly more time when working in teams (M = 32.5 minutes) than when working alone (M = 29.7), [$F(1,64) = 4.12$, $p < .05$]. A motivational account would have predicted that managers would spend less time, not more, in real teams because they could assume that seniors had detected most of the mechanical errors. The two-way interaction just described was itself moderated by its interaction with In/Out of Specialization [$F1,64) = 7.93$, $p < .01$]; managers spent more time in real teams than when working alone when working out of specialization (real team M = 36.7 v. 29.8 when working alone), but not when working in specialization (real team M = 28.3 v. 29.7 when working alone), [$F(1,64) = 11.98$, $p < .001$]. We saw in the error detection results that managers showed process loss for both mechanical and conceptual errors when working out of specialization; it is inconsistent with a motivation account to find that managers were spending significantly less time when they were working alone.

5. Conclusions and Implications

This research examined risk reduction at three levels: individual, nominal team, and real team. First, our results for the individual processing support prior research, but using a richer context. Working individually, audit seniors detected more mechanical errors than managers did; managers detected more conceptual errors than seniors (Ramsay [1994] and Bamber and Ramsay [1997]). However, this result only holds when auditors are working in their industry specialization (cf., Johnson and Jamal [1988], Solomon et al. [1999]). Out of specialization, both managers and seniors perform poorly.

Second we compared nominal teams to the individuals to examine how the nominal team performance differs from individual performance; i.e., how would team performance have suffered in the absence of each individual, given redundancy in the errors detected by each? Our results show that the nominal teams are clearly superior to either the seniors or managers working individually and in specialization. In specialization, the advantage of the nominal team over the senior working alone is greater for mechanical than for conceptual errors, and the advantage of the nominal team over the manager working alone is greater for conceptual than for mechanical errors. Out of specialization, there is no such asymmetry. Since a prerequisite for effective team performance is that the individual members each contribute nonoverlapping skills, our results show that hierarchical review aligns seniors and managers with different skills, but only in specialization. Out of specialization, one might speculate that a nominal team of two (less expensive) seniors would have performed as well as our senior-manager teams.

Third, we compare real teams to nominal teams. Our findings show that nominal teams performed better than real teams. Managers detect fewer incremental errors when they review the seniors' work along with original work papers than when they examine the work papers independently of the senior. The results also suggest that the process loss is due to output interference, because managers were not adversely affected by exposure to the senior's work for detecting conceptual errors in specialization. Out of specialization, we observed equal process loss for mechanical and conceptual errors. This is consistent with memory literature that implies that cues from the senior's work should not be harmful if the manager's expertise provides a rich categorical structure and the senior's work is not redundantly cuing categories that the manager would have retrieved without that work.

Our results have implications for both practice and research. The basic notion is that the current review process is effective because of the complementary nature of the error types detected by seniors and managers when they are working within specialization. When audit teams are working out of specialization, we do not find that kind of complementarity. Moreover, out of specialization, real teams clearly fall short of their (nominal team) potential. Collectively, these results (1) support the wisdom of

assigning teams in specialization, (2) suggest that the complementarities in types of errors detected (mechanical v. conceptual) make it worthwhile assigning high-priced managers rather than a second lower-priced senior to review the first senior's work, and (3) support the current sequential review practice in which managers use senior's work as an input, rather than "reviewing" by performing an independent audit, then comparing notes.

Our results also suggest that, working out of specialization, a somewhat different organization is called for. First, the absence of complementarities in errors detected suggests that the senior's work could as effectively be reviewed by another senior. Second, the pervasiveness of process loss suggests that it might be more effective for the "reviewer" to perform his or her own independent audit of the workpapers before comparing notes with the work of the first senior. Our real v. nominal team comparisons suggest that, out of specialization, such a structure may require less time from the manager and but yield more incremental contribution of the manager to team performance when compared to the traditional sequential review. However, caution should be exercised since the review process itself did not detect all of the errors in this task. In our experiment, the seniors were reviewed in a traditional manner. Audit firms should continue experimenting with different methods of reducing the risk of misstatements such as face-to-face review (Rich et al. [1997a, b]).

Industry specialization played a strong role in all of our results. Our findings should lend credence to the current approach by public accounting firms of aligning their practices along industry lines. Prior work had documented the importance of specialized knowledge for individual performance but we demonstrate more subtle benefits of specialization for teams. Not only did managers and seniors perform better as individuals in specialization than out, but their patterns of performance displayed more team synergy. In specialization, seniors detect mechanical errors that would not be detected by the manager while managers detect conceptual errors that would not be detected by the senior. Thus, the industry specialization approach leads to an effective deployment of the efforts of seniors and managers.

APPENDIX A

MECHANICAL AND CONCEPTUAL ERROR RATING TASK

The main objective of the error rating task was to gather independent, expert information on which errors were mechanical or conceptual (Ramsay [1994]). There is no measurement criteria in the auditing literature for determining which errors are mechanical or conceptual in nature, or for determining the potential multidimensional nature of errors. By having industry experts (audit partners) rate the errors on these dimensions, we were able to develop an independent measure to classify the subject's responses as conceptual or mechanical.

After developing the list of errors for the error rating task, the rating instrument was sent to the respective industry specialists at the Big 5 firm for review. A copy of the instrument was also sent to two audit partners of another Big 5 firm who specialized in each industries. Minor changes were made to the instrument based on the specialists' review. The final error rating task was pilot tested with two senior audit manager specialists from the participating Big 5 firm who spent approximately 15 minutes each in completing the error rating task.

ADMINISTRATION AND SUBJECTS

The error rating instrument was mailed to each audit partner at his or her office with instructions on how to complete the instrument and they were instructed to return it directly to the researchers. If more than one partner was available in a participating office, one partner served as the main contact person who distributed the instruments to other partners and collected the completed instruments for return to the researchers. A total of 100 instruments were distributed (50 health care and 50 banking). Twenty health care (40 percent) and thirty-one banking (62 percent) partner responses were returned. The mean (median) years of audit experience of the partners was 19.47 (17.00), while the years of industry experience in health care and banking was 16.8 and 17.9, respectively.

ERROR CATEGORIZATION METRIC

The audit partners' responses were categorized into their respective error types (mechanical, conceptual, and mixed) using the following approach: If 90 percent or more of the industry partners rated an error to be mechanical or conceptual, the error was categorized as such. If less than 90 percent of the partners rated an error as mechanical or conceptual, the error was categorized as mixed. We did not include the mixed errors in the analyses reported in the paper because it was unclear from the literature what patterns should be hypothesized for them. Empirically, they showed results very similar to those for conceptual errors.

RESULTS

The results of the error rating task in terms of classifying the errors as mechanical or conceptual are presented in Table 1. As noted in Panel A, 5 errors were classified as mechanical, 3 errors were classified as conceptual, and 2 errors were classified as mixed. Banking error number 11 was deleted from the analysis because a number of the partners indicated that, depending on the loan, it was appropriate to reduce the loan principal instead of recognizing interest income. Panel B shows that 4 errors were classified as mechanical, 3 errors were classified as conceptual, and 5 errors were classified as mixed. Health care error number 10 was deleted because a number of partners indicated that most hospitals reduce net patients service revenue and credit liability any way in this type of situation. Therefore, if the liability was accrued, the effect is the same. These classifications were used to code the subjects' responses in the main experiment.

Figure 1
Experimental Design

Panel A: Health Care Industry Specialized Teams

	Working Alone		Working as Team	
	manager	senior	manager	senior
Alone In Specialization Team Out of Specialization Teams 1-18*	Hospital Case by In Specialization		Banking Case by Out of Specialization	
	Working Alone		Working as Team	
	manager	senior	manager	senior
Alone Out of Specialization Team In Specialization Teams 19-36	Banking Case by Out of Specialization		Hospital Case by In Specialization	

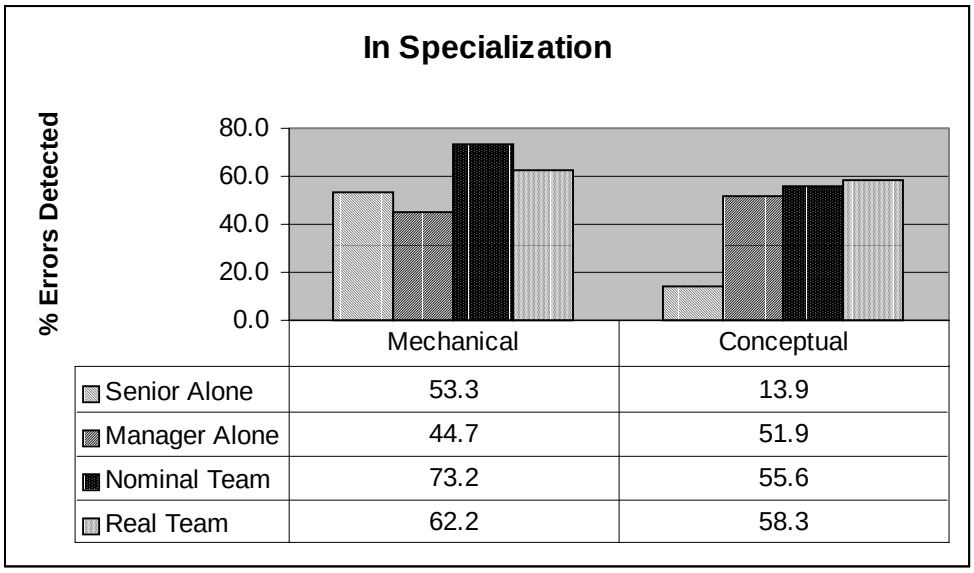
Panel B: Banking Industry Specialized Teams

	Working Alone		Working as Team	
	manager	senior	manager	senior
Alone In Specialization Team Out of Specialization Teams 37- 54	Banking Case by In Specialization		Hospital Case by Out of Specialization	
	Working Alone		Working as Team	
	manager	senior	manager	senior
Alone Out of Specialization Team In Specialization Teams 55-72	Hospital Case by Out of Specialization		Banking Case by In Specialization	

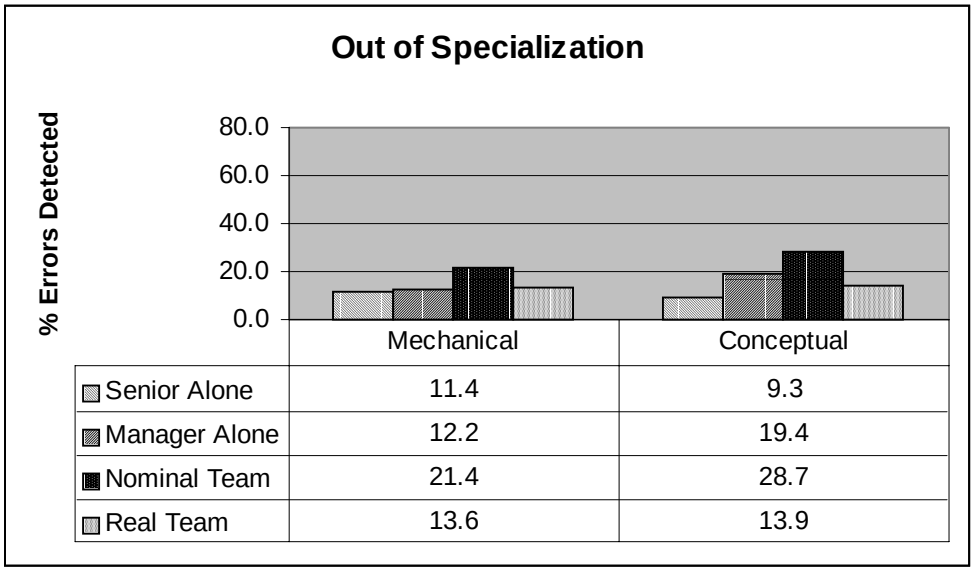
Note that each group of 18 teams in this figure was actually split into two sets of 9 teams each, varying the order in which Alone and Team cases were completed. This is not shown in the Figure.

Figure 2

Percent of Mechanical and Conceptual Errors Detected by Seniors, Managers, Nominal, and Real Teams Working In v. Out of Specialization



	Mechanical	Conceptual
Senior Alone	53.3	13.9
Manager Alone	44.7	51.9
Nominal Team	73.2	55.6
Real Team	62.2	58.3



	Mechanical	Conceptual
Senior Alone	11.4	9.3
Manager Alone	12.2	19.4
Nominal Team	21.4	28.7
Real Team	13.6	13.9

Table 1
Description and Categorization of Industry Specific Seeded Errors

Panel A: Description of Seeded Banking Errors

Error Description	Error Category*	Importance Ratings**
1. Current period revenue includes interest income from non accrual loans.	M(90)	47.2
2. Current period revenues reduced by charged-off loans.	MX(77)	29.8
3. Allowance for uncollectible loans not fully recorded.	C(100)	99.8
4. Revenues recorded but not earned due to misclassification of loans.	M(100)	46.0
5. Current year recoveries not recorded or understated.	M(90)	21.4
6. Interest payments received on non performing loans not treated on cash basis.	C(90)	34.8
7. Re-estimated allowance for loan loss reserve are more than reported allowance due to mis-classification of loans.	C(90)	53.8
8. Payments received on written-off loans are used to reduce loan principal instead of re-instituting the loss and recognizing revenue.	MX(63)	17.2
9. Loans are mis-classified.	M(90)	33.6
10. Allowance for uncollectible loans in the balance sheet does not match the balance from test of details.	M(93)	39.0
11. Payments received on non accrual/non performing loans are used to reduce loan principal instead of recognizing interest revenue.	deleted	

Panel B: Description of Seeded Health care Errors

Error Description	Error Category	Importance Ratings
1. HMO accounts included in Medicare contractual adjustment report.	M(90)	49.1
2. Allowance for uncollected accounts has not been adjusted for current period recoveries.	MX(60)	34.1
3. Fund Balance does not roll.	M(93)	53.2
4. Failure to record current period contingent liability for prior period over billing of Medicare.	C(90)	72.8
5. Current period Medicare receivables are not properly contractualized.	MX(55)	91.0
6. Current period HMO receivables are not properly contractualized.	C(94)	79.6
7. Failure to separate Medicare and Medicaid patient's charges.	M(90)	60.6
8. Billing Medicare for patient charges that are more than allowable DRG charges.	MX(63)	54.8
9. Billing Medicare for non-covered and full outpatient charges.	MX(63)	58.9
10. Current period revenue reduced for prior year Medicare adjustment instead of recognizing a liability.	deleted	
11. Allowance for uncollectible accounts in balance sheet does not match the balance from test of details.	M(90)	38.6
12. Credits, co-payments, and deductions larger than normal.	C(90)	22.1
13. Abnormal addition or writeoff of accounts receivable at year end.	MX(42)	31.9

M = mechanical, C = conceptual, MX = mixed.

* The number shown in parentheses for the mixed error types is the percentage of partners who rated the error type as mechanical.

** The importance ratings were provided by the experts using the following instructions:

"Please rate the importance of the following errors in the health care (banking) industry in the following manner. Assign the number 100 to the error that is the most important, and values between 1 and 99 to the remainder depending on how important they are compared to the error assigned 100. In rating the importance of each error, consider how *material, critical, and consequential,* it would be on a typical banking (health care) audit."

Table 2
Demographic and Completion Time Data

Panel A: General and Industry Experience

Industry Specialization	General Audit Experience		Industry Audit Experience	
	Senior	Manager	Senior	Manager
Banking (n=72)	37	85	34	68
Health care (n=72)	54	90	40	84

Panel B: Length of Time Spent (in minutes) to Complete the Cases

Industry Specialization	Banking	Health care	Overall
Banking:			
Senior	32.13	25.33	57.46
Manager	30.07	26.62	56.69
Team*	26.87	33.93	60.08
Health care:			
Senior	34.15	38.72	73.07
Manager	32.94	29.30	62.24
Team*	39.49	29.62	69.11

* Team represents the amount of time spent by the review team manager reviewing the work of the senior.

Table 3
Individual and Team Detection Rates by Error Type and Industry Specialization

Panel A: Detection Rate of Errors Seeded in Banking

		Banking Specialists					Health care Specialists				
Error Number	Error Type	Overall	Seniors Individually	Managers Individually	Nominal Team	Real Team	Overall	Seniors Individually	Managers Individually	Nominal Team	Real Team
1	M	0.472	0.500	0.444	0.833	0.611	0.194	0.222	0.167	0.278	0.278
2	MX	0.583	0.389	0.889	0.833	1.000	0.139	0.056	0.222	0.278	0.056
3	C	0.319	0.139	0.500	0.528	0.839	0.153	0.083	0.222	0.306	0.167
4	M	0.528	0.556	0.500	0.833	0.667	0.139	0.167	0.111	0.278	0.222
5	M	0.486	0.528	0.444	0.750	0.611	0.069	0.028	0.111	0.139	0.111
6	C	0.264	0.083	0.444	0.361	0.444	0.222	0.167	0.278	0.444	0.222
7	C	0.264	0.194	0.333	0.472	0.500	0.111	0.056	0.167	0.222	0.056
8	MX	0.792	0.694	0.889	0.806	1.000	0.292	0.250	0.333	0.417	0.500
9	M	0.458	0.472	0.444	0.806	0.556	0.056	0.056	0.056	0.111	0.056
10	M	0.250	0.333	0.389	0.444	0.500	0.153	0.139	0.222	0.194	0.056
Overall		0.442	0.389	0.528	0.667	0.673	0.153	0.122	0.189	0.267	0.172

Panel B: Detection Rate of Errors Seeded in Health care

Error Number	Error Type	Health care Specialists					Banking Specialists				
		Overall	Seniors Individually	Managers Individually	Nominal Team	Real Team	Overall	Seniors Individually	Managers Individually	Nominal Team	Real Team
1	M	0.500	0.556	0.444	0.667	0.667	0.153	0.194	0.111	0.306	0.278
2	MX	0.514	0.417	0.611	0.639	0.611	0.167	0.056	0.278	0.333	0.278
3	M	0.653	0.694	0.611	0.917	0.778	0.111	0.000	0.222	0.222	0.167
4	C	0.556	0.333	0.778	0.889	0.778	0.153	0.083	0.222	0.306	0.222
5	MX	0.500	0.389	0.611	0.722	0.722	0.069	0.028	0.111	0.083	0.000
6	C	0.361	0.111	0.611	0.611	0.833	0.181	0.083	0.278	0.306	0.056
7	M	0.542	0.583	0.500	0.806	0.889	0.042	0.028	0.056	0.083	0.056
8	MX	0.472	0.333	0.611	0.722	0.889	0.000	0.000	0.000	0.000	0.111
9	MX	0.375	0.194	0.556	0.583	0.833	0.097	0.111	0.111	0.139	0.056
11	M	0.458	0.583	0.333	0.750	0.500	0.069	0.083	0.056	0.139	0.056
12	C	0.250	0.056	0.444	0.500	0.500	0.069	0.083	0.056	0.139	0.167
13	MX	0.431	0.250	0.611	0.861	0.833	0.194	0.167	0.222	0.333	0.222
Overall		0.465	0.375	0.560	0.722	0.736	0.109	0.076	0.143	0.199	0.139

TABLE 4
ANOVA Results for Nominal Teams Versus Individuals (Senior v. Manager)

MODEL: RR = INDUSTRY, IN/OUT OF SPEC., ORDER, HIERARCHY, ERRTYPE				
Independent Variables	df	SS	F-value	P-value
Between Factors:				
INDUSTRY (Banking v. Health Care)	1	7183.96	7.40	.008
IN/OUT OF SPEC.	1	155205.56	159.88	.000
ORDER (Alone 1st, Team 2nd v. Team 1st, Alone 2nd)	1	239.51	0.25	.621
INDUSTRY x IN/OUT OF SPEC.	1	3530.61	3.64	.061
INDUSTRY x ORDER	1	28.86	0.03	.864
IN/OUT OF SPEC. x ORDER	1	670.84	0.69	.409
INDUSTRY x IN/OUT OF SPEC. x ORDER	1	25.52	0.03	.872
Error	64	970.74		
Within Factors:				
HIERARCHY (Seniors v. Managers v. Nominal Teams)	2	7503.26	27.54	.000
HIERARCHY x INDUSTRY	2	193.79	0.71	.492
HIERARCHY x IN/OUT OF SPEC.	2	865.09	3.18	.045
HIERARCHY x ORDER	2	35.50	0.13	.878
HIERARCHY x INDUSTRY x IN/OUT OF SPEC.	2	92.77	0.34	.712
HIERARCHY x INDUSTRY x ORDER	2	159.65	0.59	.558
HIERARCHY x IN/OUT OF SPEC. x ORDER	2	2.24	0.01	.992
HIERARCHY x INDUSTRY x IN/OUT OF SPEC. x ORDER	2	69.96	0.26	.774
Error (HIERARCHY)	128	272.46		
ERRTYPE (Mechanical v. Conceptual)	1	48.66	0.08	.772
ERRTYPE x INDUSTRY	1	286.27	0.50	.482
ERRTYPE x IN/OUT OF SPEC.	1	1881.95	3.29	.075
ERRTYPE x ORDER	1	27.17	0.05	.828
ERRTYPE x INDUSTRY x IN/OUT OF SPEC.	1	1150.52	2.01	.161
ERRTYPE x INDUSTRY x ORDER	1	774.23	1.35	.249
ERRTYPE x IN/OUT OF SPEC. x ORDER	1	249.54	0.44	.512
ERRTYPE x INDUSTRY x IN/OUT OF SPEC. x ORDER	1	325.52	0.57	.454
Error (ERRTYPE)	64	572.82		
HIERARCHY x ERRTYPE	2	1382.19	6.90	.001
HIERARCHY x ERRTYPE x INDUSTRY	2	97.27	0.49	.616
HIERARCHY x ERRTYPE x IN/OUT OF SPEC.	2	1861.16	9.30	.000
HIERARCHY x ERRTYPE x ORDER	2	38.43	0.19	.825
HIERARCHY x ERRTYPE x INDUSTRY x IN/OUT OF SPEC.	2	360.59	1.80	.169
HIERARCHY x ERRTYPE x INDUSTRY x ORDER	2	442.06	2.21	.114
HIERARCHY x ERRTYPE x IN/OUT OF SPEC. x ORDER	2	62.43	0.31	.733
HIERARCHY x ERRTYPE x INDUSTRY x IN/OUT OF SPEC. x ORDER	2	140.80	0.70	.500
Error (HIERARCHY x ERRTYPE.)	128	200.21		

Table 5
ANOVA Results for Nominal v. Real Teams

MODEL: RR = INDUSTRY, TASKPAIR., ORDER, Nominal/Real Team, ERRTYPE				
Independent Variables	df	SS	F-value	P-value
Between Factors:				
INDUSTRY (Banking v Health Care)	1	4739.59	10.19	.002
TASKPAIR (Alone In-Team Out v. Alone Out-Team In) = Nominal/Real * In/Out of Specialization)	1	932.88	2.01	.162
ORDER (Alone 1st, Team 2nd v. Team 1st, Alone 2nd)	1	381.95	0.82	.368
INDUSTRY x TASKPAIR	1	29.18	0.06	.803
INDUSTRY x ORDER	1	51.40	0.11	.740
TASKPAIR x ORDER	1	505.80	1.09	.301
INDUSTRY x TASKPAIR x ORDER	1	132.03	0.28	.596
Error	64	29777.62		
Within Factors:				
Nominal/Real Team (H3)	1	4265.29	8.26	.006
Nominal/Real Team x INDUSTRY	1	136.58	0.26	.608
Nominal/Real Team x TASKPAIR = In/Out of Specialization Main Effect	1	132684.04	256.85	.000
Nominal/Real Team x ORDER	1	7.03	0.01	.907
Nominal/Real Team x INDUSTRY x TASKPAIR = INDUSTRY * In/Out of Specialization	1	4848.39	9.39	.003
Nominal/Real Team x INDUSTRY x ORDER	1	33.57	0.06	.780
Nominal/Real Team x TASKPAIR x ORDER	1	5.10	0.01	.921
Nominal/Real Team x INDUSTRY x TASKPAIR x ORDER	1	.78	0.00	.969
Error (Nominal/Real Team)	64	516.59		
ERRTYPE (Mechanical v. Conceptual)	1	873.85	1.75	.190
ERRTYPE x INDUSTRY	1	862.28	1.73	.193
ERRTYPE x TASKPAIR = ERRTYPE * Nominal/Real * In/Out of Specialization (H3B)	1	1944.45	3.90	.053
ERRTYPE x ORDER	1	.01	0.00	.996
ERRTYPE x INDUSTRY x TASKPAIR	1	87.04	0.17	.678
ERRTYPE x INDUSTRY x ORDER	1	123.16	0.25	.621
ERRTYPE x TASKPAIR x ORDER	1	429.41	0.86	.357
ERRTYPE x INDUSTRY x TASKPAIR x ORDER	1	.09	0.00	.989
Error (ERRTYPE)	64	499.20		
Nominal/Real Team x ERRTYPE (H3A)	1	202.79	0.39	.532
Nominal/Real Team x ERRTYPE x INDUSTRY	1	250.00	0.49	.488
Nominal/Real Team x ERRTYPE x TASKPAIR = ERRTYPE * In/Out of Specialization	1	3815.98	7.42	.008
Nominal/Real Team x ERRTYPE x ORDER	1	208.42	0.41	.526
Nominal/Real Team x ERRTYPE x INDUSTRY x TASKPAIR = In/Out of Special. x ERRTYPE x INDUSTRY	1	1961.81	3.81	.055
Nominal/Real Team x ERRTYPE x INDUSTRY x ORDER = 1st v. 2nd	1	2049.78	3.98	.050
Nominal/Real Team x ERRTYPE x TASKPAIR x ORDER	1	.79	0.01	.941
Nominal/Real Team x ERRTYPE x INDUSTRY x TASKPAIR x ORDER	1	118.84	0.23	.632
Error (Nominal/Real Team x ERRTYPE.)	64	514.56		

REFERENCES

Alba, J. W. and A. Chattopadhyay. "Effects of Context and Part-Category Cues on Recall of Competing Brands." *Journal of Marketing Research* 22 (August 1985): 340-349.

American Institute of Certified Public Accountants. *AICPA Professional Standards*, Vol. 1. New York: AICPA, 1999.

Bamber, E. M. "Expert Judgment in Audit Teams: A Source Reliability Approach." *Journal of Accounting Research* 21 (Autumn 1983): 392-412.

Bamber, E. M., and J. H. Bylinski. "The Audit team and the Audit Review Process: An Organization Approach." *Journal of Accounting Literature* 1 (1982): 33-58.

Bamber, E. M., and R. J Ramsay. "An Investigation of the Effects of Specialization in Audit Workpaper Review." *Contemporary Accounting Research* 14 (Fall 1997): 501-14.

Diehl, M. and W. Stroebe. "Productivity Loss in Brainstorming Groups: Toward the Solution of a Riddle." *Journal of Personality and Social Psychology* 53 (3) (1987): 497-509.

Frederick, D. "Auditors' Representation and Retrieval of Internal Control Knowledge." *The Accounting Review* 65 (April 1991): 240-58.

Gramling, A. A. and D. N. Stone. "Industry Issues in Auditing." Working paper, University of Illinois, June 1999.

Harding, N. and K. T. Trotman. "Hierarchical Differences in Audit Workpaper Review Performance." *Contemporary Accounting Research* (Winter 1999): forthcoming.

Hill, G. W. "Group Versus Individual Performance. Are N+1 Heads Better Than One?" *Psychological Bulletin* (1982): 517-39.

Hoch, S. J. "Availability and Interference in Predictive Judgment." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10 (4) (1984): 649-662.

Johnson, P. E. and K. Jamal. "Human Judgment: Limitations and Opportunities in the Year 2000" 1987 Proceedings of the Arthur Young Professors' Roundtable. Edited by A. B. Bailey. Reston, VA: Council of Arthur Young Professors (1988).

Libby, R. "The Role of Knowledge and Memory in Audit Judgment." In *Judgment and Decision Research in Accounting and Auditing*, edited by A. H. Ashton and R. H. Ashton. New York: Cambridge University Press, 1995.

Libby, R. and J. Luft. "Determinants of Judgment Performance in Accounting Settings: Ability, Knowledge, Motivation, and Environment." *Accounting, Organizations and Society* 18(5) (July 1993): 425-450.

Messier, W. F. and R. M. Tubbs, "Recency Effects in Belief Revision: The Impact of Experience and the Review Process," *Auditing: A Journal of Practice & Theory* 13 (Spring 1994): 57-72.

Ramsey, R. "Senior/Manager Differences In Audit Workpaper Review Performances." *Journal of Accounting Research* 32 (Spring 1994): 127-35.

Rich, J. S., I. Solomon, and K. T. Trotman. "Audit Review Process A Characterization from the Persuasion Perspective." *Accounting Organization and Society* 22 (July 1997a): 481-505.

Rich. J. S., I Solomon, and K. T. Trotman. "Multi-Auditor Judgment Making Research: A decade Later." *Journal of Accounting Literature* 16 (1997b): 86-126.

Solomon, I. "Multi-Auditor Judgment/Decision Making Research." *Journal of Accounting Literature* 6 (1987): 1-25.

Solomon, I., M. D. Shields, and O. R. Whittington. "What Do Industry-Specialist Auditors Know?" *Journal of Accounting Research* 37 (Spring 1999): 191-208.

Stasser, G, and D. Stewart. "Discovery of Hidden Profiles in Decision-Making Groups: Solving a Problem Versus Making a Judgement." *Journal of Personality and Social Psychology* 63 (September 1992): 426-434.

Stewart, D. and G. Stasser, "Expert Role Assignment and Information Sampling During Collective Recall and Decision Making." *Journal of Personality and Social Psychology* 69 (October 1995): 619-628.

Taylor, D. W., P. C. Berry, and C. H. Block. "Does Group Participation When Using Brainstorming Facilitate or Inhibit Creative Thinking?" *Administrative Science Quarterly* 3 (1958): 23-47.

Taylor, M. H. "Bounded Rationality, Uncertainty, and Competence: The Effects of Industry Specialization on Auditors' Inherent Risk Assessments." Working paper, University of Nebraska, December 1998.

Trotman, K. T. "The Review Process and the Accuracy of Auditor Judgments." *Journal of Accounting Research* 23 (Autumn 1985): 740-52.

Trotman, K. T., and P. W. Yetton. "The Effect of the Review Process on Auditor Judgments." *Journal of Accounting Research* 23 (Spring 1985): 256-267.

Trotman, K. T., P. W. Yetton and I. Zimmer. "Individual And Group Judgments Of Internal Control Systems." *Journal of Accounting Research* 21 (1983): 286-292.