

**Fraud-risk Factors and Audit Planning:
The Effects of Auditor Rank**

David S. Kerr*

INTRODUCTION

External auditors have a responsibility to plan and perform the audit to detect material misstatements in financial statements caused by fraud (AICPA, 2002). This is a complex task requiring considerable professional judgment. Although many firms have computer software tools that assist in audit planning, and while AICPA publications and firms' audit manuals provide general guidelines for the planning of audits, the actual planning is left largely to the auditor's judgment.

Audit planning involves consideration and integration of several factors. As described in SAS No. 109 (AICPA, 2006), the auditor should consider the entity's environment, including both management tone-at-the-top and internal control. In addition, the auditor must consider the results of analytical procedures performed in the planning stage of the audit. An important purpose of these analytical procedures is to help the auditor identify unusual items or relationship that might suggest the presence of a material misstatement due to fraud (AICPA, 2002). These factors, along with others such as materiality and assessed control risk, are then evaluated and integrated by the auditor when assessing the risk of material misstatement due to fraud and when deciding the appropriate nature, extent, and timing of subsequent substantive tests.

*The author is Associate Professor of Accounting at University of North Carolina at Charlotte.

The objective of this research is to examine the association between auditor rank and outcome measures of judgment performance in audit planning, including a study of the relationship between rank and judgmental agreement (i.e., consensus) among auditors, a comparison of alternate measures of judgmental agreement, and an analysis of three specific components of judgmental agreement: factor-weighting agreement, judgment consistency, and agreement on the appropriate degree of configural factor processing.¹ Understanding the effects of auditor rank on audit planning judgments is important for several reasons. First, different judgments made by auditors at different ranks and with different amounts of professional audit experience may affect the effectiveness or efficiency of the audit as differing judgments lead some auditors to either overaudit or underaudit (Colbert, 1989; Kozloski, 2011). Advocating the use of rank as a factor in audit judgment research, Tan and Jamal (2006, 768) state, “Auditors of different ranks hold particular roles and are associated with general characteristics. For example, audit managers likely have better knowledge about technical issues, client management, and project management than do staff auditors.” Rose *et al.* (2010) find evidence that auditors at higher ranks (i.e., managers and partners) tend to be more trusting than audit seniors, staff, and students. Second, by studying the effects of rank on audit planning judgments, we hope to shed light on the stage at which auditors obtain sufficient expertise for convergence in their judgment strategies (and, by implication, for “correct” decisions). In addition, as pointed out by Abdolmohammadi and Wright (1987, 1), “knowledge of differences in the decision processes of experts vis-a-vis novices would be extremely valuable in designing decision aids (e.g., decision support and expert systems) and in developing staff training programs.”

¹ Configural factor processing involves consideration of the interactions or patterns of factors in combination rather than merely processing factors separately (Magro and Nutter, 2012).

This paper extends the audit judgment and decision making literature on fraud and audit planning in three important ways. First, to our knowledge, this is the first study to examine the effects of auditor rank on auditors' joint consideration of tone-at-the-top, internal control, and preliminary analytical procedures in the context of audit planning.

Second, to gain an understanding of the effects of rank (as a proxy for general audit knowledge and experience) on audit planning judgments and the degree of their consensus, this study compares the judgments of three groups of participants (audit managers, seniors, and auditing students) across two tasks varying in complexity (audit program planning and internal control evaluation). By including students in the participant pool, a comparison of the relative effects of classroom instruction and audit practice experience and training on judgmental agreement is possible.

Third, in this study, two measures of inter-auditor judgmental agreement are examined — the mean absolute difference between participants' pair-wise judgments, and the traditional Pearson product-moment correlation coefficient — allowing a comparison of the performance of alternate measures of judgmental agreement. Some interesting differences in results are found between the two measures.

BACKGROUND AND DEVELOPMENT OF HYPOTHESES

In recent years, academic researchers have begun to study “tone-at-the-top” as a risk factor and its relationship to audit decision making. Based on interviews with 27 audit partners and managers, Dickins *et al.* (2008) provide evidence that tone-at-the-top affects external auditors' perceived risk and, in turn, their audit fees. Extending research on tone-at-the-top to internal auditing, Hansen *et al.* (2009) report the results of a survey investigating how internal auditors assess management tone and how those assessments affect the extent of audit test work.

The study's findings led the authors to recommend that tone-at-the-top assessments be made by internal auditors with sufficient experience interacting with upper management rather than lower ranking auditors with less experience and fewer interactions with management. Kerr and Diaz (2009) explore the role of tone-at-the-top in audit planning along with the interaction of tone with other fraud risk factors. They find that auditors consider tone-at-the-top in combination with internal control and analytical procedures in audit planning. We extend this literature by addressing whether the auditor's position in the firm ("rank") has an effect on the auditor's consideration of the risk factors examined by Kerr and Diaz (2009).

Auditors of different ranks may use different judgment processes. For instance, a relatively new auditor with one year of professional experience may use different judgment strategies than an audit manager with ten years of experience. Similarly, auditing students with no professional experience may process information differently than experienced audit practitioners. Aspects of auditors' judgment strategies which may differ across rank include the relative weighting of fraud risk factors and consideration of the interactions among risk factors (i.e., the degree of configural factor processing). Consistent with prior auditing and nonauditing research, professional rank is used in this study as a proxy for task-specific knowledge and audit experience, as rank potentially is a more appropriate proxy for knowledge than is years of general audit experience (Tan, 2001; Tan and Jamal, 2006; Trotman *et al.*, 2009; Wright, 2001). The appropriateness of using professional rank as a proxy for knowledge is discussed by Abdolmohammadi (1999, 54): "...accounting firms make a determination of the promotability of an auditor to a professional rank based on performance measures, including technical knowledge. In many cases, professional rank is positively correlated with years of experience.

Further, some auditors are promoted earlier than others based on performance measures that signal the achievement of a level of competence necessary for the higher rank.”

Auditors’ roles and responsibilities typically differ across ranks. Given these differences in roles and responsibilities across rank, auditors of different ranks may weigh fraud risk factors differently. For instance, audit partners and managers have more responsibilities for interactions with their clients’ top management than audit seniors or staff (McKnight and Wright, 2011; Tan and Jamal, 2006). Accordingly, we would expect tone at the top to be given more importance (i.e., weighted more heavily) by higher-ranking auditors than lower-ranking auditors when planning an audit.

In addition, given the relative complexity of audit planning, audit experience / knowledge is expected to be associated with greater levels of configural factor processing in audit planning. Configural factor processing requires the consideration of the configuration, or pattern, of multiple factors such that the effect of one factor is altered by the condition of one or more other factors. Thus, configural processing requires the recognition of the interrelationships among various factors (Abdolmohammadi, *et al.*, 2004). Prior research has observed that configural information processing is a skill typically acquired with experience (Hammersley, 2006; Magro and Nutter, 2012), with less-experienced individuals exhibiting less configurality in their processing of information relative to more-experienced individuals. Therefore, in audit planning, we predict there will be a greater degree of configurality in the judgment processes of higher ranking auditors relative to lower-ranking auditors.

To examine the effects of auditor rank on auditors’ processing of risk factors in audit planning, the current study employs the “expertise paradigm” (Bonner, 1994; Tan and Jamal, 2006). That is, both rank and task complexity are varied, with the focus on comparisons of

performance differences of the groups of participants *across* task complexity. The concept of task complexity is operationalized in terms of task structure (Trotman, 2001). Structured, or simple, tasks are well-defined, relatively straightforward, and generally require little professional judgment. Unstructured, or complex, tasks require considerably more professional judgment and expertise, as they are less straightforward, with fewer if any guidelines available.

Audit planning, including consideration of fraud risk factors, is generally considered to be a relatively complex task requiring professional judgment (Bonner and Pennington, 1991), while internal control evaluation is considered a relatively structured task in that it is a “relatively straightforward, well-defined task” (Krogstad, *et al.*, 1984, 70). Given the structured nature of internal control evaluation, and given that students receive training in their auditing courses on which characteristics of internal control constitute strengths and weaknesses, one would not expect the judgments of experience practitioners to differ significantly from those of students in a simple internal control evaluation task (Davis and Solomon, 1989). In contrast, audit planning is considered to be a more complex task than internal control evaluation. Audit planning involves not only the evaluation of internal controls, but also entails other tasks such as the evaluation of preliminary analytical procedures and other fraud risk factors. As tasks increase in complexity, the greater the need for task-specific knowledge and skills to perform the tasks (Abdolmohammadi, 1999; Tan and Jamal, 2006). Accordingly, professional audit experience / knowledge is expected to play a more prominent role in audit planning than in the evaluation of internal control. Therefore, there should be a greater differential in the consensus levels of audit practitioners and auditing students when planning the extent of substantive tests of balances than when evaluating internal control, with practitioners exhibiting greater consensus levels than audit students when planning the extent of substantive tests of balances.

The preceding analyses lead to the following hypotheses:

- H1: In audit planning, tone at the top will be of greater importance to auditors of higher rank than to auditors of lower rank.
- H2: In audit planning, auditors of higher rank will exhibit greater configularity in their processing of fraud risk factors than lower ranking auditors.
- H3: There will be a greater difference between the level of judgmental agreement among senior-level auditors and the level of agreement among auditing students when planning the extent of substantive tests of details of balances than when evaluating internal control.

The third hypothesis is a prediction of a rank x audit-task interaction for audit seniors and student participants.

Prior research suggests that the effects of rank on judgmental agreement may diminish or reverse as higher ranks are achieved (Bedard, 1989; Libby, 1995; Tan and Jamal, 2006). As explained by Bedard (1989, 114), "... an auditor may have twenty years of auditing experience, but not exhibit more "expertise" in evaluating internal control than a senior because, of these twenty years of experience, only three involve evaluating internal control systems." Furthermore, the rank of the auditor defines the auditor's roles in the firm, with audit partners and managers having more responsibility for client development and retention than lower-ranking auditors. Since the knowledge related to audit planning is obtained early in auditors' careers and may gradually wane over time, auditors at higher rank might exhibit decreased levels of judgmental agreement.

While the knowledge obtained at the ranks of audit staff and senior in the early years of auditors' careers may serve to educate auditors in the manner in which their firm expects them to plan audits, resulting in a convergence in their judgment strategies, these rank effects on judgmental agreement among auditors can be expected to begin to reverse in later years of auditors' careers. Accordingly, in the current study, given that planning the extent of substantive tests and internal control evaluation are performed primarily at the senior level, auditors at higher

ranks beyond the senior level are expected to be associated with decreased levels of judgmental agreement when performing those tasks.² Since rank effects on judgment agreement and their reversals, if any, should be larger for relatively complex tasks requiring a greater degree of judgment than for structured, less complex, tasks, and since prior research suggests that audit planning is a more difficult, less structured task than internal control evaluation, any reversal of rank effects should manifest itself through a greater differential in the consensus levels of audit managers and audit seniors when planning the extent of substantive tests than when evaluating internal control, with managers exhibiting lower consensus levels than seniors when planning the extent of substantive tests. This leads to the fourth hypothesis:

- H4: There will be a greater difference between the level of judgmental agreement among manager-level auditors and the level of agreement among senior-level auditors when planning the extent of substantive tests of details of balances than when evaluating internal control.

The methodology used to test these hypotheses is described in the next section.

METHOD

Research Instrument and Administration of the Experiment

Participants included practicing auditors and accounting students. A partner in an international accounting firm assisted in obtaining audit managers and audit seniors with audit planning experience as participants. The research instruments were delivered in person to the partner, who in turn, distributed an instrument to each participant. After completing the experiment at their own pace, participants returned the instrument directly to the partner. Twenty audit seniors and 19 audit managers completed an audit-planning scenario, and the same 20 seniors and 18 of

² Prior research indicates that audit planning is usually performed by personnel at the senior level, with the manager acting primarily as a reviewer and in an advisory role when necessary (Abdolmohammadi, 1999; Trotman and Wright, 1996). In addition, discussions with personnel (an audit partner and an audit manager) in the firm that participated in this study indicated that audit planning, including setting time budgets and determining the extent of testing, is primarily the responsibility of seniors.

the managers completed an internal control evaluation scenario. The audit planning scenario used in this study is the same as that used in Kerr and Diaz (2009). All practitioner participants were CPAs. Twenty undergraduate accounting students enrolled in their first auditing course at a university in the U.S. also participated in both scenarios.³ The topics of audit risk, internal control evaluation, and the audit of the sales, receivables, and cash receipts cycle had been covered previously in class lecture. Students, like their practitioner counterparts, were given instructions concerning the importance of working independently.

The research instrument consisted of three parts. Part 1 included instructions and information about the client's business and industry. Part 2 consisted of the audit planning cases. In each of these cases, participants received information concerning the client's tone-at-the-top, internal controls, and results of preliminary analytical procedures. Each of these factors was treated as an aggregate variable and was operationalized as a repeated-measures variable. Tone at the top was operationalized by varying several potential "red flags" of fraud, including management's cooperativeness with the auditors, management's aggressiveness with respect to accounting policies, employees' competence, and results of the previous year's audit. Results of tests of controls were operationalized independently of tone at the top and manipulated as indicating either a weak control system in which several key controls were absent, or a strong control system in which controls were operating effectively. Results of preliminary analytical procedures were manipulated as suggesting either the possible presence of material misstatements in accounts receivable, or no indications of material misstatements in receivables.

³ Thirty-seven students completed both scenarios. Responses from four students with auditing experience from internships were discarded. Since an approximately equal number of participants at each experience level was desired, 13 of the remaining students' responses were eliminated randomly to arrive at a final student sample size of 20. These responses were discarded in a way that maintained the Latin-square design of the experiment.

The levels of each factor were varied dichotomously as either unfavorable (high fraud risk condition) or favorable (low fraud risk condition), resulting in eight audit planning cases for each participant to consider. Information about a fourth factor, the client's business stability, was presented as part of the background information in part 1 and was operationalized as a between-subjects variable.⁴ Other potentially relevant factors were held constant in the background information. For each case, participant were asked to decide the number of hours they would plan for substantive tests of details of accounts receivable and then indicate their planned extent of tests of details for receivables using a nine-point Likert scale anchored with "no testing of receivables" and "extensive testing of receivables."

Part 3 contained the internal control evaluation cases and concluded with a debriefing questionnaire. In each of the internal control evaluation cases, participants were presented with information concerning four separate control procedures related to the client's sales and cash collection cycle. The levels of each factor were varied dichotomously and indicated whether a particular control was operating effectively. The manner in which the control procedures were operationalized resulted in eight internal control evaluation cases per participant. For each case, participants evaluated the effectiveness of the internal control system and indicated their judgment on a nine-point Likert scale anchored with "extremely weak" and "adequate to strong."⁵

⁴ Two additional cases were included in both the audit planning cases and the internal control evaluation cases to explore a separate issue beyond the scope of this paper. This increased both the number of audit planning cases and the number of internal control evaluation cases per participant from eight to ten. It is unlikely that the content of the additional cases affected the data analyzed in this study for two reasons: (1) participants were instructed to treat each case independently from all other cases, and (2) the two additional cases presented no new information about the client and actually contained *less* information than any one of the eight main cases.

⁵ To control for possible practice effects, a Latin square design was used to systematically counterbalance the order in which each participant received the audit planning cases and the internal control evaluation cases.

RESULTS

To examine the relationship between audit rank and the manner in which auditors integrate tone at the top with other fraud risk factors when planning the extent of substantive testing, we performed separate ANOVAs for managers, seniors, and students. In each ANOVA, participants' extent-of-testing ratings were the dependent variable, and tone at the top, results of tests of controls, and results of preliminary analytical procedures were the independent variables. ANOVA results are presented in Table 1.

[Insert Table 1 here]

The ANOVA results reveal some interesting similarities and differences in the manner audit managers, audit seniors, and students made their extent-of-testing judgments. The main effects for tone at the top, tests of controls, and preliminary analytical procedures were statistically significant for all three ranks ($p=0.000$). When tone at the top was unfavorable, participants' extent-of-testing ratings were significantly higher than when tone was favorable, *ceteris paribus*.⁶ This finding held across participant ranks, indicating that participants of all ranks responded to the manipulations of the risk factors in the expected manner.

Analysis of Auditor Rank and the Relative Importance of Risk Factors

ANOVA's eta-squared index (η^2) was used to investigate the relationship between auditor rank and the relative importance (i.e., factor-weighting) participants placed on each of the fraud risk factors when making their extent-of-testing decisions. Similar to the squared multiple-correlation coefficient (R^2) in multiple-regression, this index provides a measure of each factor's comparative effect size on participants' judgments. Overall, practitioners weighted tone at the top

⁶ Similar results were observed for tests of controls and preliminary analytical procedures across the three ranks of participants. That is, ratings were higher when tests of controls were unfavorable than when they were favorable, and were higher when preliminary analytical procedures were unfavorable than when they were favorable.

more heavily than either of the other two risk factors, and tests of control procedures were weighted more heavily than preliminary analytical procedures. When analyzed by rank, tone at the top increased in weight as auditor rank increased, with managers ($\eta^2 = 0.469$) placing greater importance on tone at the top that did seniors ($\eta^2 = 0.387$), and seniors giving it more weight than students ($\eta^2 = 0.239$) (see table 1). These ANOVA results are consistent with H1. In contrast, analytical procedures decreased in weight, or importance, as rank increased. These findings may be attributable to differences in the job requirements and duties of higher ranking auditors, with managers focusing more on client relationships and retention compared to auditors of lower rank (Rose *et al.*, 2010; Trotman *et al.*, 2009).

Analysis of Auditor Rank and Consideration of Risk-Factor Interactions

Managers exhibited a higher degree of configural factor processing than seniors or students as indicated by a marginally significant three-way interaction among tone, tests of controls, and analytical procedures ($p=0.056$). The nature of this three-way interaction can be seen clearly in Figure 1, which illustrates a pair of two-way interactions between tests of controls and analytical procedures: one where tone at the top is unfavorable, and the other where tone is favorable.

[Insert Figure 1 here]

As seen in Panel A, when tone is unfavorable, there is virtually no interaction between the other risk factors ($p=0.957$). However, when tone is favorable, an interaction between tests of controls and analytical procedures is apparent ($p=0.011$). These findings are consistent with H2. The auditing literature suggests a moderating effect of internal controls on analytical procedures; specifically, the results of analytical procedures should be more trustworthy when the client's internal controls are strong rather than weak (AICPA, 1988; Arens, *et al.*, 2012; Bedard, *et al.*,

1999). The three-way interaction's pattern indicates that this moderating effect is especially strong in the presence of a favorable tone at the top, suggesting that managers are more willing to reduce the extent of testing when the combined results of tests of control procedures and analytical procedures are favorable and tone at the top is also favorable. It should be noted that this higher-order interaction between risk factors was exhibited only by audit managers; it was not observed in audit seniors' or auditing students' judgments. This finding is indicative of a greater degree of configularity in managers' judgment processes relative to seniors or students.

Analysis of Auditor Rank and Judgmental Agreement

Hypotheses three and four predict that, for the audit planning judgment examined, auditor rank will be associated with (1) increases in judgmental agreement in the early stages of auditors' careers, and (2) decreases in judgmental agreement in subsequent years. The mean absolute difference (MAD) for pair-wise judgments was used as the measure of judgmental agreement.

Correlation coefficients have been the traditional measure of inter-auditor judgmental agreement in most prior audit judgment studies. However, while correlations measure the extent to which changes in one person's judgments (relative to that person's average judgment) tend to be associated with changes in another person's judgments, they tell us nothing about the magnitude of the differences in those judgments.⁷

To identify the actual magnitude of judgment differences across participants in this study, pair-wise mean absolute differences (MAD) between participants' Likert-scale ratings in each of the eight audit planning cases and eight internal control evaluation cases were calculated. These

⁷ For instance, one participant's average budgeted audit hours was 36 hours with a variance of 262, while another participant's average budgeted hours was 656 with a variance of 42,455. Although there were very large differences in these participants' budgeted hours for each audit planning case, the correlation between their budgeted hours was very high (0.96).

MADs were calculated separately for each level of audit experience examined and are presented in Table 2.

[INSERT TABLE 2 HERE]

To provide a comparison of MADs with correlation coefficients, Pearson product-moment correlation coefficients between participants' pair-wise Likert-scale ratings in each of the eight audit planning cases and the internal control evaluation cases were also calculated. To compare the two measures of consensus, and to provide an initial test of the hypothesized effects of auditor rank, two sets of ANOVAs were performed — one set based on correlations as the dependent variable, and another set based on MADs. Within each set, one ANOVA was based on students' and seniors' responses; the second was based on seniors' and managers' responses. Scheffe's multiple comparisons tests were used to examine simple effects.

In neither ANOVA based on correlations was the auditor-rank x audit-task interaction significant ($p > 0.05$). As discussed below, this is inconsistent with the results based on pair-wise MADs. Given that the correlation coefficient removes the effects of differences in means and variances, and given that analysis of the actual magnitude of judgment differences is meaningful in the context of internal control evaluation and audit planning judgments, the correlation coefficient appears to be a less sensitive measure of differences in participants' judgments than the mean absolute difference.

In contrast to the correlation results, analyses of students' and seniors' MADs revealed a significant auditor-rank x audit-task interaction ($F=22.69, p < 0.001$). Scheffe's multiple comparisons tests and inspection of this interaction revealed that, as hypothesized, seniors had greater consensus among themselves than did students when planning the extent of testing

($p < 0.001$), and there was no significant difference in seniors' and students' consensus levels when evaluating internal control ($p = 0.163$). These results are consistent with hypothesis three.⁸

To examine whether additional audit experience beyond the senior level is associated with reduced levels of judgmental agreement when planning the extent of substantive tests, seniors' and managers' MADs were also compared. ANOVA revealed a significant auditor-rank x audit-task interaction in seniors' and managers' MADs ($F = 7.78$, $p = 0.006$). Scheffe's test and inspection of the interaction indicated that, while there was no difference in the judgmental agreement levels of seniors and managers in the internal control evaluation task ($p = 0.609$), managers agreed less among themselves concerning the appropriate extent of detailed tests of balances than did seniors ($p = 0.015$). These results are consistent with hypothesis four and suggest a reversal of rank effects beyond the senior level for the audit planning task examined.⁹

In summary, these results suggest that professional audit experience and/or training in the early years of auditors' careers increases the level of judgmental agreement among auditors in the context of the audit planning task examined. In the early stages of their careers, experience and training serve to educate auditors in the manner in which their firm expects them to perform. This education leads to increased judgmental agreement. However, for the audit planning task examined in this study, as auditors continue to gain more audit experience beyond the senior level, rank (experience) effects reverse, resulting in decreased judgmental agreement.¹⁰

⁸ To assess the extent to which results are driven by outliers, SPSS's *explore* procedure was used to identify "extreme values" in the data. Deletion of extreme values in the data set did not significantly alter the results.

⁹ Although participants' budgeted audit hours were not the primary variable of interest in this study, we examined budgeted hours for consistency with the findings based on Likert-scale extent-of-testing ratings. ANOVA revealed significant differences between MADs in budgeted audit hours for students and seniors (215.94 and 14.44, respectively; $F = 46.98$, $p < 0.001$), and for seniors and managers (14.44 and 26.85, respectively; $F = 18.23$, $p < 0.001$). These results are consistent with those based on extent-of-testing ratings.

¹⁰ The parametric tests performed assume that all groups come from normal populations with the same variance. Since the data show some instances of non-normality and heterogeneity of variance, non-parametric statistical tests—the Kruskal-Wallis test and the Mann-Whitney U test—were also performed to test the significance of differences in levels of judgmental agreement across auditor rank. The results were virtually identical to those observed using the parametric

Additional Analyses of Judgmental Agreement

To provide a clearer understanding of the observed rank effects on judgmental agreement, and to gain a better understanding of the similarities and differences in the judgment processes of auditors at differing ranks, the following three components of judgmental agreement were analyzed: factor-weighting agreement, judgment consistency, and agreement on configural factor processing (Bonner, 1990). These three components affect the degree to which the judgments of two auditors agree.

Factor-weighting agreement (G_{MAD}) is measured as the mean absolute difference between the predictions of linear regression models of participants i 's and j 's judgments. Judgment consistency (R_{MAD}) is measured as the mean absolute difference between participant i 's judgments and the judgments predicted by the regression model of participant i 's judgments. Agreement on configurality (C_{MAD}) is measured as the mean absolute difference between the residual variances that are unexplained by the linear regression models of participants i and j . Values for G_{MAD} , R_{MAD} and C_{MAD} across auditor rank and audit task are presented in Table 3.

[INSERT TABLE 3 HERE]

The major source of differences between the responses of participants at different ranks appears to arise from differences in the weight each participant placed on the risk factors in the experiment (i.e., factor-weighting agreement). Differences in judgment consistency, R_{MAD} , and differences in the degree of configurality in auditors' judgment processes, C_{MAD} , appear to have little influence. To study this further, we performed two sets of three 2 x 2 ANOVAs. One set consisted of three ANOVAs in which G_{MAD} , R_{MAD} , and C_{MAD} for students and audit seniors were

tests, with one minor exception. In the internal control evaluation task, the Mann-Whitney test reported the difference between students' and seniors' consensus levels to be statistically significant ($p=0.042$), whereas Scheffe's test reported this difference to be not significant ($p=0.163$).

the dependent variables, and rank and audit task were the independent variables. The second set of ANOVAs was based on G_{MAD} , R_{MAD} , and C_{MAD} for audit seniors and managers.

The results of the first set of ANOVAs revealed a significant rank x task interaction for factor-weighting agreement (G_{MAD}) ($F=25.89$, $p<0.001$). The pattern of this interaction is nearly identical to that based on students' and seniors' MADs, indicating that the extent to which students and audit seniors differ in their factor-weighting judgments is greater when planning the extent of testing than when evaluating internal controls. The ANOVAs based on R_{MADs} and C_{MADs} revealed no significant interaction effects, suggesting that the effects observed previously for audit seniors and students are not due to differences in judgment consistency or configural factor processing agreement.

The results of the set of ANOVAs for audit managers and seniors also revealed a significant rank x task interaction for factor-weighting agreement (G_{MAD}) ($F=9.33$, $p=0.002$). The pattern of this interaction is nearly identical to that based on audit seniors' and managers' MADs, indicating that the extent to which audit seniors and managers differ in their factor-weighting judgments is greater when planning the extent of testing than when evaluating internal controls. Results of the ANOVAs based on R_{MADs} and C_{MADs} revealed no significant interaction effects. These results suggest that, for the audit planning task examined in this study, the effects of auditor rank are manifested primarily from differences in factor-weighting agreement, with rank having little or no effect on judgment consistency or configurality.

Analyses of Debriefing Questions

An alternative explanation for the finding that managers exhibit less judgment agreement among themselves than do seniors in the audit planning task may be that seniors were comparatively more attentive to, or engaged by, the task itself than were managers. Seniors may

have been more attentive to the task because the task is more relevant, and thus more engaging, to auditors at the senior level, or they felt more implied pressure from their firm to attend carefully to the task. This possible alternative explanation was examined by analyzing participants' responses to four questions included in the debriefing questionnaire: (1) How realistic did you find the experiment? (scale anchored 1=very realistic, 4=very unrealistic), (2) How interesting did you find the experiment (scale anchored 1=very interesting, 4=very boring), (3) Would you like a copy of the results of this study, and (4) How long did it take you to complete the experiment (excluding any breaks taken).

As shown in Table 4, there were no significant differences between seniors and managers on any of the measures of attentiveness. Accordingly, the previously observed differences in managers' and seniors' levels of judgmental agreement do not appear to be driven by differences in the degree to which participants attended to the task.

[INSERT TABLE 4 HERE]

SUMMARY, IMPLICATIONS, AND CONCLUSION

Summary

The objective of this research was to examine the effects of auditor rank (as a surrogate for audit knowledge/experience) on outcome measures of judgment performance in audit planning. Results of the experiment reveal that, when planning the audit, managers place greater relative weight on tone-at-the-top than do senior auditors, and seniors give it greater weight than auditing students. Results also reveal a higher degree of configural processing of fraud risk factors by audit managers than lower-ranking participants. In addition, results support the hypotheses that, in the context of the audit planning task examined, (1) in the early stages of auditors' careers (i.e., lower ranks), audit experience/training tends to increase inter-auditor judgmental agreement

regarding the effects of tone at the top, internal controls, and analytical procedures, and, (2) later in auditors' careers, higher rank is associated with decreased judgmental agreement. Auditor rank had little impact on the more-structured internal control evaluation task.

Three components of judgmental agreement were examined to learn which component is most affected by rank. Results show that rank is associated with inter-auditor judgmental agreement mainly through its effect on factor-weighting agreement. Differences in judgment consistency and differences in agreement on configural factor processing had little effect on inter-auditor agreement.

Comparisons of the traditional measure of inter-auditor judgmental agreement — the correlation coefficient — with mean absolute differences demonstrated that mean absolute differences are more sensitive to differences in participants' responses than are correlations. For the type of tasks employed in this study, and in other contexts in which the degree of judgment consensus is meaningful, the mean absolute difference appears to be a more appropriate measure of inter-auditor judgmental agreement.

Implications for Auditing Practice

The results of the study have important policy implications for public accounting firms and auditors responsible for assessing fraud-risk factors when planning an audit. Audit planning is a relatively complex task requiring the auditor to evaluate and integrate multiple factors when determining the appropriate nature, extent, and timing of audit procedures. These factors include, among others, tone-at-the-top, results of preliminary analytical procedures, and results of tests of the client's internal control system.

One of this study's findings is that there are significant differences in the weights placed on each of the above factors by auditors of different ranks. For instance, managers placed more weight

on tone-at-the-top than auditors at lower ranks. However, audit planning is typically performed largely by senior-level auditors, with managers serving primarily as reviewers and advisors to the audit team (Abdolmohammadi, 1999; Trotman and Wright, 1996). Results of this study suggest that CPA firms should ensure that managers are actively involved in audit-planning tasks, such as when members of the audit team meet to exchange ideas regarding how and where material misstatements due to fraud might occur and their potential effects on the client's financial statements (AICPA 2002). When audit managers are involved in these "brainstorming" sessions, the managers' relevant experiences and professional judgment are made available to lower ranking, less experienced, staff members. In addition, managers' active participation in these tasks would likely result in the audit team placing greater weight on specific fraud-risk factors such as client management's tone-at-the-top.

Conclusion

This study focused primarily on the evaluation of various fraud-risk factors when planning an audit and the effects of professional rank on those evaluations. Results revealed that participants weighted tone-at-the-top more heavily than either tests of internal control activities or preliminary analytical procedures and weighted tests of controls more heavily than preliminary analytical procedures. In addition, results indicated that audit managers place more weight on tone-at-the-top and exhibit a greater degree of configularity in their processing of fraud-risk factors relative to audit seniors and auditing students.

The experiment performed in this study contains several limitations that are present in many behavioral studies. These limitations tend to restrict the ability to generalize the results of this study to other groups and decision environments. Since participants were not randomly selected from the entire population of auditors, caution should be exercised when generalizing the

results to other groups of auditors. In addition, while this study attempted to provide participants with realistic case situations, the realism in the auditors' decision environment was limited. Finally, it should be noted that any single study is limited in its ability to allow us to draw conclusions regarding the judgment processes of individuals. Replications of results are needed to assess the pervasiveness and robustness of results across different participants and settings.

REFERENCES

- Abdolmohammadi, M. J. 1999. A comprehensive taxonomy of audit task structure, professional rank and decision aids for behavioral research. *Behavioral Research in Accounting*, 11: 51-92.
- Abdolmohammadi, M.J., and A. Wright. 1987. An examination of the effects of experience and task complexity on audit judgments. *The Accounting Review*, (January): 1-13.
- Abdolmohammadi, M.J., D.G. Searfoss, and J. Shanteau. 2004. An investigation of the attributes of top industry audit specialists. *Behavioral Research in Accounting* (16): 1-17.
- American Institute of Certified Public Accountants (AICPA). 1988. *Analytical Procedures*. Statement on Auditing Standards No. 56. New York: AICPA.
- _____. 2002. *Consideration of Fraud in a Financial Statement Audit*. Statement on Auditing Standards No. 99. New York, NY: AICPA.
- _____. 2006. *Understanding the Entity and Its Environment and Assessing the Risks of Material Misstatement*. Statement on Auditing Standards No. 109. New York, NY: AICPA.
- Arens, A.A., Elder, R.J., & Beasley, M.S. 2012. *Auditing and Assurance Services: An Integrated Approach*. (14th Ed.). Upper Saddle River, New Jersey: Prentice Hall: 185.
- Bedard, J. 1989. Expertise in auditing: Myth or reality? *Accounting, Organizations and Society*, 14: 113-131.
- Bedard, J.C., Mock, T.J., & Wright, A.M. 1999. Evidential planning in auditing: A review of the empirical research. *Journal of Accounting Literature*, 18: 96-142.
- Bonner, S. 1990. Experience effects in auditing: The role of task-specific knowledge. *The Accounting Review* (January): 72-92.
- Bonner, S. 1994. A model of the effects of audit task complexity. *Accounting, Organizations and Society*, 19, 3: 213-234.
- Bonner, S., and N. Pennington. 1991. Cognitive processes and knowledge as determinants of auditor expertise. *Journal of Accounting Literature*, 1-50.
- Colbert J. 1989. The effect of experience on auditors' judgments. *Journal of Accounting Literature*, 137-149.
- Davis, J. and I. Solomon. 1989. Experience, expertise, and expert-performance research in public accounting. *Journal of Accounting Literature*, 150-164.

- Dickins, D.E., Higgs, J.L., & Skantz, T.R. 2008. Estimating audit fees post-SOX. *Current Issues in Auditing*, 2: A9-A18.
- Hansen, J., Stephens, N.M., & Wood, D.A. 2009. Entity-level controls: The internal auditor's assessment of management tone at the top. *Current Issues in Auditing*, 3 (1): A1-A13.
- Hammersley, J.S. 2006. Pattern identification and industry-specialist auditors. *The Accounting Review*, 81 (2): 309-336.
- Kerr, D., & Diaz, M.C. 2009. Auditor consideration of tone-at-the-top in audit planning: An experimental investigation. *Journal of Forensic & Investigative Accounting*, 1 (1).
- Kozloski, T.M. 2011. Knowledge transfer in the fraud risk assessment task. *Journal of Forensic & Investigative Accounting*, 3 (1): 49-85.
- Krogstad, J., R. Ettenson, and J. Shanteau. 1984. Context and experience in auditors' materiality judgments. *Auditing: A Journal of Practice & Theory*, (Fall): 54-73.
- Libby, R. 1995. The role of knowledge and memory in audit judgment. In R. Ashton and A. Ashton (eds.) *Judgment and Decision-making Research in Accounting and Auditing*, (Cambridge University Press).
- Magro, A.M., and S.E. Nutter. 2012. Evaluating the strength of evidence: How experience affects the use of analogical reasoning and configural information processing in tax. *The Accounting Review*, 87 (1): 291-312.
- McKnight, C.A., and Wright, W.F. 2011. Characteristics of relatively high-performance auditors. *Auditing: A Journal of Practice & Theory*, 30 (1): 191-206.
- Rose, A.M, J.M Rose, and M. Dibben. 2010. The implications of auditors' dispositional trust and career advancement opportunities for the detection of fraud. *Journal of Forensic & Investigative Accounting*, 2 (3): 141-163.
- Tan, H.T. 2001. Methodological issues in measuring knowledge effects. *International Journal of Auditing*, 5: 215-224.
- Tan, H.T., and Jamal, K. 2006. Managing perceptions of technical competence: How well do auditors know how others view them? *Contemporary Accounting Research*, 23 (3): 761-787.
- Trotman, K.T. 2001. Design issues in audit JDM experiments. *International Journal of Auditing*, 5: 181-192.
- Trotman, K.T. and A.M. Wright. 1996. Recency effects: Task complexity, decision mode, and task-specific experience. *Behavioral Research in Accounting*, 8: 175-193.

Trotman, K.T., A.M. Wright, and S. Wright. 2009. An examination of the effects of auditor rank on pre-negotiation judgments. *Auditing: A Journal of Practice & Theory*, 28 (1): 191-203.

Wright, W.F. 2001. Task experience as a predictor of superior loan loss judgments. *Auditing: A Journal of Practice and Theory*, 20 (1): 147-155.

Table 1
ANOVA Results for Audit Planning (by Auditor Rank)

Panel A: Audit Managers

Dependent Variable: Extent-of-Testing Ratings

Source	Mean Square	F	Eta Squared ¹	Significance (two-tailed)
Main Effects				
Tone at Top	164.24	126.97	0.469	0.000
Tests of Controls	45.32	35.04	0.196	0.000
Preliminary Analytical Procedures	16.45	12.72	0.081	0.000
Two-way Interactions				
Tone*Tests of controls	3.18	2.46	0.017	0.119
Tone*Analytical procedures	1.11	0.86	0.006	0.355
Tests of controls*Analytical procedures	5.16	3.99	0.027	0.048
Three-way Interactions				
Tone*Tests of controls*Analytical procedures	4.80	3.71	0.025	0.056

Panel B: Audit Seniors

Dependent Variable: Extent-of-Testing Ratings

Source	Mean Square	F	Eta Squared ¹	Significance (two-tailed)
Main Effects				
Tone at Top	71.54	85.93	0.387	0.000
Tests of Controls	60.71	72.92	0.349	0.000
Preliminary Analytical Procedures	38.54	46.29	0.254	0.000
Two-way Interactions				
Tone*Tests of controls	1.89	2.27	0.016	0.134
Tone*Analytical procedures	0.09	0.10	0.001	0.750
Tests of controls*Analytical procedures	0.39	0.47	0.003	0.500
Three-way Interactions				
Tone*Tests of controls*Analytical procedures	1.09	1.30	0.009	0.256

Panel C: Auditing Students

Dependent Variable: Extent-of-Testing Ratings

Source	Mean Square	F	Eta Squared ¹	Significance (two-tailed)
Main Effects				
Tone at Top	86.29	47.77	0.239	0.000
Tests of Controls	64.39	35.65	0.190	0.000
Preliminary Analytical Procedures	96.88	53.63	0.261	0.000
Two-way Interactions				
Tone*Tests of controls	3.75	2.08	0.013	0.152
Tone*Analytical procedures	1.14	0.63	0.004	0.428
Tests of controls*Analytical procedures	1.14	0.63	0.004	0.428
Three-way Interactions				
Tone*Tests of controls*Analytical procedures	3.16	1.75	0.011	0.188

¹ Greater eta-squared values indicate greater weight given to each risk factor.

Table 2
 Mean Absolute Differences:
 Managers, Seniors, and Students

		Task	
		<u>Audit Planning</u>	<u>Internal Control Evaluation</u>
Auditor Rank	Managers	1.30 (0.44) [1.25]	1.28 (0.59) [1.17]
	Seniors	1.06 (0.33) [1.00]	1.35 (0.50) [1.38]
	Students	1.43 (0.64) [1.31]	1.20 (0.43) [1.19]

Standard deviations are shown in parentheses; median values are shown in square brackets. Smaller mean absolute differences reflect greater consensus.

Table 3
 Mean Values for G_{MAD} , R_{MAD} , and C_{MAD}

		Cue Weighting Agreement (G_{MAD})	Judgment Consistency (R_{MAD})	Configurality Agreement (C_{MAD})
Audit Planning	Managers	1.20 (0.45) [1.19]	0.40 (0.22) [0.31]	0.53 (0.24) [0.50]
	Seniors	0.96 (0.37) [1.00]	0.34 (0.14) [0.38]	0.47 (0.15) [0.44]
	Students	1.33 (0.68) [1.19]	0.41 (0.13) [0.38]	0.53 (0.19) [0.53]
Internal Control Evaluation	Managers	1.16 (0.60) [1.06]	0.40 (0.14) [0.38]	0.56 (0.21) [0.53]
	Seniors	1.26 (0.50) [1.22]	0.33 (0.22) [0.25]	0.48 (0.27) [0.42]
	Students	1.05 (0.44) [1.00]	0.43 (0.27) [0.39]	0.58 (0.25) [0.56]

Standard deviations are shown in parentheses; median values are shown in square brackets. Smaller numbers reflect greater agreement/consistency.

Table 4
Measures of Task Attentiveness

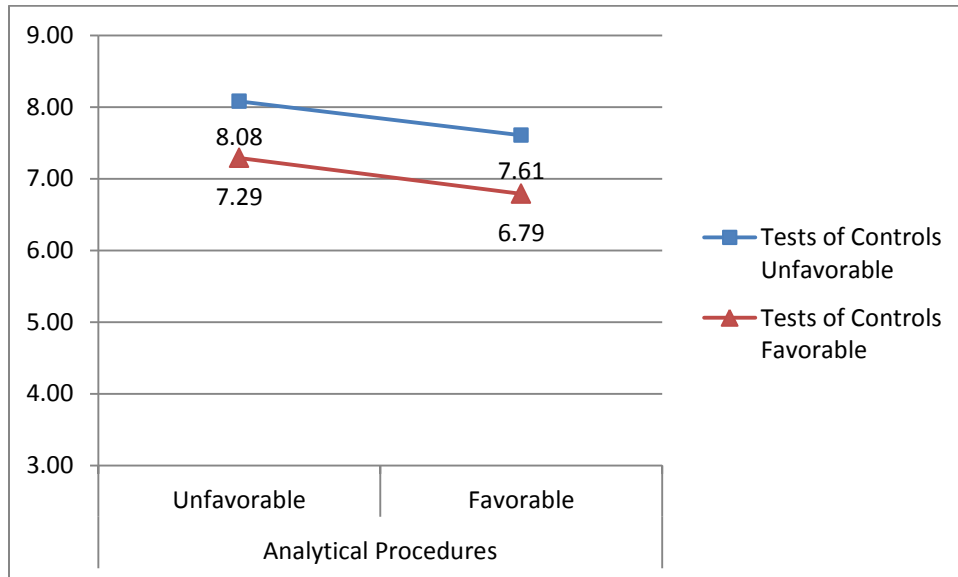
	<u>Level of realism</u> ¹	<u>Level of interest</u> ²	<u>Requested copy of results</u>	<u>Time (minutes) taken to complete experiment</u>
Managers	2.17	2.28	50%	59.13
Seniors	2.00	2.28	33%	55.31
Difference	0.17	0.00	17%	3.82
t-value (U-value)	0.77	0.00	(135.00)	0.579
Significance level	0.45	1.00	0.317	0.567

¹ 1=very realistic, 4=very unrealistic

² 1=very interesting, 4=very boring

Figure 1
Decomposition of Three-way Interaction:
Managers' Extent-of-Testing Judgments

Panel A: Unfavorable Tone at the Top



Panel B: Favorable Tone at the Top

