

THE EFFECT OF “TEAM” VERSUS “GROUP” AUDIT FORMAT CHARACTERISTICS ON RISK ASSESSMENTS MADE BY AUDIT PARTNERS

D. Donald Kent, Jr.

State University of New York at Brockport

Donald F. Arnold, Sr. and Josef Schmee

Union College, New York

ABSTRACT

Research has shown a recent tendency within auditing away from the use of a “team” approach (a hierarchical and sequential audit review process with a predominant emphasis on written communication) and towards a “group” approach (a co-acting process with greater staff interaction and a greater emphasis on oral communication and consensus building). The purpose of this research was to examine the effect of two important “team” versus “group” format characteristics (oral versus written discussion and consensus versus non-consensus reaching) upon partners’ risk assessments within the typical multiple employment level structure of audit firms. One hundred and twenty-four practicing auditors from 30 local public accounting firms were formed into 31 groups, each group consisting of four different hierarchical employment levels. Each group performed three audit risk assessment tasks, first as individuals and then in a group setting. Two group situations were used: written communication with justification but without consensus, representing the “team” format; and discussion to consensus, representing the “group” format. The results indicated that the partners’ final risk assessment decisions were more heavily influenced by the risk preferences of the non-partner staff within the group when the “group” format was used as compared to when the “team” format was used. The implications of the findings are that the format of communications utilised by the audit firm and the degree of staff/partner consensus-reaching sought by the partners can influence the partners’ final decisions regarding issues of risk assessment.

INTRODUCTION

In their summary of multi-auditor judgement/decision-making research, Rich, Solomon and Trotman (1997) documented a current trend within the audit profession towards placing greater emphasis on the utilisation of co-acting audit

"groups" versus sequentially acting "teams" in the conduct of an audit. Under the "team" approach, the audit staff would perform the audit procedure, document the evidence in the working papers, reach a conclusion based on the evidence and pass the working papers to the senior for review and approval. Once approved, the working papers would be passed to the audit manager and then to the audit partner for final review. The process relied extensively, though not exclusively, on written communications (the working papers) and was completed, primarily, in a sequential manner by employment level (staff to senior to manager to partner).

Under the "group" approach, individuals function more co-actively, providing for greater interaction along multiple dimensions within the audit function. Individuals operate more as a task or goal oriented group. Greater reliance is placed upon oral communication, joint concurrent efforts to solve problems in a consensus building manner and a higher degree of first-hand contact between various members of the audit team. Rich et al. (1997) observed that while the "team" approach still plays an important part within the auditing profession, the recent trend has been towards greater utilisation of the "group" dynamic. They noted that this trend has developed for various reasons, including a desire for increased efficiency within the audit's review process, a reduction of written working papers available to plaintiffs during litigation and the retention of top-qualified personnel by involving all staff levels in greater policy development and decision-making processes within the audit.

Research has been conducted on selected comparative characteristics of the two formats. For example, it has shown that auditors function with, and react to, audit evidence differently when working in multi-person units than when working as unrelated individuals (Solomon, 1987). Johnson (1995) observed that group effects offset the well established recency effect¹ found when examining individual auditors; Reckers and Schultz (1993) observed group versus individual auditor differences when examining factual answer versus judgement type of decisions; while Bambers, Watson and Hill (1996) observed differences in the quality of audit decisions when comparing groups versus individuals.

There has been only limited research, however, on the effect of the change from the "team" to the "group" format upon the important audit element of risk that the auditor assumes when exercising professional judgement within the audit process. While the final decision as to the level of such risk to be assumed must be made by the audit partner, the whole concept of a group participation (staff, senior, manager and partner) within an audit engagement clearly implies that the entire staff will provide input to and thus influence the final decision maker.

The goal of the current research was to examine the effect of two important differentiating characteristics of the "group" and "team" approaches upon selected risk assessments made when audits are conducted by groups composed of individuals at the multiple hierarchical employment levels typically found in public accounting. The differentiating characteristics examined were (a) the primary, though not exclusive, means of communication utilised within each approach - oral communication for the "group" approach and written communication for the "team" approach, and (b) the process of "consensus building",

more prevalent within the "group" approach. The risk assessments selected for the study related to evaluating assertions made by the client management regarding the appropriateness of specific account balances. Thus, the major research topic addressed in this study was to determine if the primary elements within each audit approach (the means of communication and the process of consensus building) have a significant impact upon audit partners when exercising their professional judgement on issues of risk assessments.

USE OF THE CHOICE-SHIFT TOOL FOR CURRENT EXAMINATION

When conducting this examination, the authors drew upon an important group behavioural dynamic relating to risk assessments made by groups of individuals (such as audit teams) known as the choice-shift phenomenon. This phenomenon, first observed by Stoner (1961), demonstrates the tendency for a group's risk decisions to gravitate towards the extreme ends, instead of the mean, of the individual risk choices provided by the participants of that group. That is, the dynamics of the group tend to pull that group's collective choice towards one end or the other of the range of risk choices provided by the individual participants when they came into the group.

Prior to Stoner's findings, it was generally thought that groups, such as committees, would have an averaging effect on decisions, with the group tending to gravitate towards the middle of varying individual positions. Hereafter, the group's movement towards one or other extreme position of the individual participants is referred to as the "traditional" pattern of choice-shift. Researchers have observed this choice-shift phenomenon in various groups, including: college students (Wallach and Kogan, 1965); middle-level executives and other professionals (Marquis, 1962; Siegel and Zajonc, 1967); men and women (Wallach, Kogan and Burt, 1967); and subjects from various Western European countries (Bateson, 1966; Fraser, Gouge and Billig, 1970; Vidmar, 1970).

While researchers have examined this phenomenon within an audit context (Barrett and O'Malley, 1976; Reckers, 1978; Schultz and Reckers, 1981), they have not utilised its full experimental potential to examine comparative characteristics of the "team" versus "group" audit approach and how these characteristics might impact on risk assessment. For example, these studies have been limited to: utilising either *oral* (a "group" characteristic) or *written* (a "team" characteristic) but not both communication formats; only requiring a *single forced group consensus* (a "group" characteristic) as opposed to *not requiring a forced consensus* (a "team" characteristic); and utilising, as subjects, personnel at only a *single employment level* within the firm (e.g. only partners), uncharacteristic of the composition of the typical audit team.

It is important to realise that the current study did not examine choice-shift *per se*. Instead, the principle of choice-shifting and the methodology used in examining this phenomenon provided a useful tool for the current study of selected characteristics of the "team" versus the "group".

RESEARCH DESIGN AND DEVELOPMENT OF HYPOTHESES

The current study followed a two-step process. The first step was to establish, as a benchmark, the shifting pattern, within an audit context, when the group consists of members at multiple hierarchical levels (partners, managers, seniors and staff). Is the pattern one of shifting towards the extremes of the participants, as might be implied by the traditional choice-shift pattern, or towards the risk preference of one of the employment levels (e.g. partners)? The second step then allowed us to focus on the primary research goals of determining if the findings of the first research step were affected by the format of the group interaction (i.e. the selected characteristics of "group" versus "team").

The influence of "employment level"

Shaw (1976) documented the existence of various effects of position and role of the participants on group dynamics. He found that, while high status persons (e.g. partners) are willing to conform on solidarity issues, they are less willing to conform to expectations of the low status person on issues of productivity. They see conforming as a loss of their power to direct outcomes and their prestige within the group. He noted that other high status persons, however, are more confident in their impressions when power relationships exist. In addition, research has shown that the reactions of the powerful person include the tendency to use power, to look at others as objects of manipulation and to demonstrate a preference towards keeping a psychological distance from the others (Shaw, 1976). Accordingly, position, as identified through the hierarchical employment level of the participants, could have an intervening effect on the choice-shift phenomenon within multi-level staffed audit groups.

There is also a considerable body of research showing hierarchical level differences in decision-making abilities that relate to risk. For example, research has shown that staff level accountants are more willing to assume additional risk than other auditors with the opposite being true of partners² (Martin and Previts, 1982; Marxen, 1990; Johnson, 1995). Furthermore, Pennington and Hastie (1988) found that decisions based on only a partial set of facts were influenced by expectations as to what made up the complete story and by knowledge about similar events. It is reasonable to assume that since individuals at higher levels of employment (e.g. partners) have had more and varied experience, such expectations and knowledge will vary among the different staff levels within the public accounting groups. Accordingly, it could be expected that when auditors make a group decision, the participant with the greatest amount of experience with or knowledge of similar events (generally the partner) would have the most influence on the others within the group.

Considering the partners' power and prestige, as well as their attitude towards risk, it would be normal to expect that the choice-shift would be towards that of the partner. Accordingly, the first hypothesis (stated in the null form), which forms our benchmark, is:

Ho1: When using groups of multiple employment level auditors, a shift towards the partner's preference will not occur.

The influence of oral versus written communication and consensus versus no consensus building alternatives

Traditional choice-shift experiments can utilise either oral (discussion) or written (ballot) forms of communication and can either require or not require the participants to reach a single group consensus. Two combinations of these alternatives frequently used in choice-shift studies are Discussion-to-Consensus (D-to-C) and Balloting-without-Consensus (B-w/o-C).

It can be seen that the B-w/o-C format parallels important characteristics within the "team" approach while the D-to-C format does the same for the "group" approach to auditing. The "team" format places greater emphasis on written communication (the audit working papers) and does not concern itself with a process of consensus building among team members. By contrast, the "group" approach places greater reliance upon oral communication and emphasises joint concurrent efforts to solve problems. Rich et al. (1997) note that the "team" approach places significant reliance upon a hierarchical sequential review of conclusions (staff up to partner level) written in the working papers, while the "group" approach uses a greater amount of first-hand oral communication contact among the hierarchical levels. A comparative examination of these two formats will provide a greater understanding of important differences between the "team" and the "group" audit approaches to the choice-shift aspect of risk assessment.

Bem, Wallach and Kogan (1965) noted that traditional shifting occurred when using discussion to reach a consensus but did not occur when using balloting (written communication) for the same purpose. They also found that shifting occurred when participants used discussion, even though the exercise did not seek consensus. Furthermore, Myers and Lamm (1976) suggest that the mode of discussion affects results as 'binding consultation creates a feeling of greater responsibility for outcomes which, in turn, creates a willingness to recognize the need to bend the rules' (Solomon, 1987, p. 15). Research has also shown that the effect of the format of the communication is dependent, in part, upon the experience of the auditors (Stone and Dilla, 1994). These findings would imply that the form of the communication (oral versus written) along with an attempt at consensus building can have significance for the choice-shift process when multiple hierarchical employment level subjects participate in the task.

From this comparison of these two formats, we developed the second hypothesis (stated in the null form):

Ho2: When using groups of multiple employment level auditors, the format used by the participants (D-to-C versus B-w/o-C) will not affect the final risk assessment decision made by the partner.

RESEARCH METHODOLOGY

Studies utilising a choice-shift framework usually follow a two-step process. First, the researcher presents a potentially risky scenario to a group of individuals. Each participant then identifies the level of risk he or she would be willing to assume individually for that scenario. The researcher averages the responses, forming a Pretest (average) score. Second, the group communicates with each other and reaches a single group decision as to the acceptable level of risk (Test score). The traditional choice-shift phenomenon occurs when the Test scores are significantly different from the average of the Pretest scores (see Wallach and Kogan, 1965, as an example).

The experimental procedures followed in the current research consisted of the standard choice-shift experimentation adapted to simulate an examination of the audit "group" and "team" environments. The researchers presented groups of auditors with the case scenarios prior to the group gathering. Each auditor's task was to read three different audit scenarios and reach an individual (personal) decision regarding a factor of audit risk. Then the individuals gathered in a group setting where the group exercise was conducted.

Case scenarios

The instrument consisted of three independent audit scenarios, each possessing the risk that the auditors may unknowingly fail to appropriately modify their audit procedures or opinion on financial statements that are materially misstated. In addition, the scenarios had to demand a degree of professional judgement in assessing that risk. For each scenario, the auditor had to decide the minimum degree of personal confidence he or she would require in an underlying assertion or assumption in the client's financial statements in order to eliminate the need for either further audit action or modification of those financial statements. The experiment used three different scenarios in order to enhance the generalisability of the results.

The scenarios had to meet the criteria of: high involvement of all hierarchical levels; high probability of a group decision being reached; and high levels of risk assessment. When developing the specific scenarios for this study, the researchers conducted a preliminary study³ to determine which general areas of the audit process met these criteria. This preliminary determination was necessary in order for the scenarios to be realistic to the participants at all four hierarchical employment levels.

Two types of decisions were found to fit the criteria: those involving a determination during the course of the audit whether (1) additional audit steps are necessary and (2) accounting estimates represented on the financial statements are acceptable. The researchers then developed three specific audit scenarios that utilised these two types of audit decisions. The scenarios involved account reclassification, bad debt provision and inventory valuation. Two partners of non-Big Five firms reviewed these scenarios for realism. Each scenario conformed to the format of the standard choice-shift dilemma whereby a person faces two alternatives, one of which is more risky than the other. After being presented with

the scenario, each participant was asked to estimate the level of confidence (measured as a percentage from 0 to 100) that they would want to have in the client's stated representation in order to be willing to reject further action (specified in the case) prior to signing off on that aspect of the audit. See **Appendix** for a summary of the scenarios and the full wording of the question asked⁴.

Participants

Sutton and Hayne (1997) noted that prior psychological research (Newcomb, 1961; Sherif, Harvey, White, Head and Sherif, 1961; McGrath, 1993) indicated that experimental results from natural versus ad hoc groups are frequently inconsistent. Research has shown that subjects' familiarity with each other affects group members' behaviour (Brock, 1973) and that a tendency towards balance and agreement in a collective system increases with acquaintance (Newcomb, 1961).

Research in auditing has tended to confirm the effect of participant familiarity on results. Asare and McDaniel (1996) found that reviewers act differently when they are more familiar with the preparers of the working papers. Bammers (1983) observed that audit managers' (seniors') perceptions of the reliability of auditor seniors' (staff) work was influenced by their familiarity with the audit seniors (staff). In addition, Nemiroff and King (1975) observed that having actually worked together allows groups to reach a tighter consensus. However, most multi-person audit research has not used subjects who have worked together ("practiced groups") prior to their participation in the research experiment. The current study overcame this problem by using subjects from smaller (local) accounting firms where professional staff members know and have worked extensively with each other.

The use of subjects from smaller (local) accounting firms offered us a second advantage of providing some empirical research on how smaller firms function. Most, if not all, prior significant research on audit-related issues has examined the topics by using as subjects staff members of the large international firms. While research (Zimmerman, Flaherty and Murray, 1999) has demonstrated the difference in the characteristics of small versus large firms (e.g. percentage of revenue from audit versus tax services, percentage of audit work done on opinion audits versus review engagements, percentage of clients whose stock is publicly traded), there has been no research examining differences between large and small firms in the conduct of the audit itself. Both large and small firms must follow the same generally accepted auditing standards and they both will usually work in small groups composed of audit staff from different hierarchical levels. Yet other than through anecdotal evidence or personal experience, we know very little about how the smaller firms function when conducting an audit.

We do know from research in the field of organisational behaviour (Nelson and Quick, 2003) that smaller organisations, in general, are less formalised and standardised, therefore providing a setting for a D-to-C audit format to have a greater potential impact on the ultimate audit decision. On the other hand, greater centralised control together with a flatter hierarchical authority tends to be a characteristic of smaller organisations (Nelson and Quick, 2003), and more

centralised organisations have a lesser degree of participative decision making (Robbins, 1996). These findings imply a tendency for the partner in charge of an audit at a smaller firm to take greater control in the decision making. Thus an examination of the impact of the two formats in question (D-to-C and B-w/o-C) has the potential to provide a greater degree of significance within the smaller accounting firms.

In summary, the utilisation of the smaller auditing firms afforded us the opportunity to use natural (versus ad hoc) groups, conduct some research on the sub-set of smaller firms seldom empirically examined, and provide us with an organisational setting that has the potential for significant impact based on the research decision utilised.

The participants were 31⁵ four-person audit teams (see Table 1 for demographic data) from 30 different "local" CPA firms⁶ located in various cities in the northeast region of the US (outside the metropolitan New York City or Boston area)⁷. The average number of partners and other professional staff of the participating firms was 5.5 and 17.4 respectively, while the average number of offices per firm was around 1.4. By using subjects from smaller firms, "practiced groups" formed the basis of the research. In each case, the four participants from each firm had significant levels of familiarity with the other group members.

TABLE 1: DEMOGRAPHIC DATA OF SUBJECTS AND FIRMS

	D-to-C	B-w/o-C	Total
Number of Firms	15	15	30
Number of Groups	15	16 ^a	31 ^a
Per firm Average Number:			
Partners	5.6	5.4	5.5
Other Professional Staff	20.0	14.9	17.4
Number of Subjects (4 per group)	60	64	124
Years in Public Accounting:			
Average	9.4 years	9.2 years	9.3 years
Range	0.5 – 35 years	0.5 – 33 years	0.5 – 35 years
CPAs	68%	67%	68%
Gender (Male)	70%	67%	69%
Graduate Degrees Held	7%	11%	9%

^a One firm had two groups.

Each group consisted of one partner, manager, senior and staff⁸, randomly assigned to one of two experimental treatments described below as Discussion-to-Consensus or Balloting-without-Consensus. Table 1 indicates that there were no significant differences in the relevant demographic characteristics of the two groups.

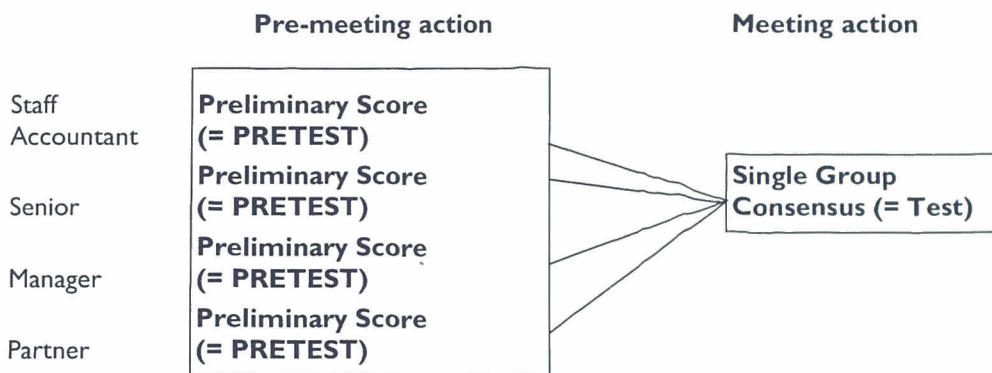
One to two days prior to the experiment, each subject received the three scenarios along with written instructions to read each scenario and write down, prior to attending the group session, his or her risk preference, defined as the percentage identified in the instructions (Preliminary score). The subjects were instructed twice (once by the contact person at the firm who did not participate in the research exercise and once in the written instructions) not to discuss the case⁹ with anyone else and were told that they would be participating in an audit study with a few other unnamed members of the firm. The subjects were not told about the actual exercise prior to coming to the group sessions, nor were they told that their assigned risk preference (Preliminary) scores would be shared.

On the day of the group session, the participants and one researcher gathered in a conference room at the respective firm's offices. In order to ensure consistency, the same researcher facilitated and monitored all of the group sessions but did not participate in the discussion.

Conduct of the exercise: D-to-C simulating the "group" multi-person audit format

For this experimental treatment, each group discussed each case in order and reached a single group decision (a single percentage of confidence) on that case (see **Figure 1**). One member of the group wrote that percentage on a piece of paper provided by the researcher. The Preliminary scores formed the Pretest scores while the single group decision formed the Test score¹⁰.

FIGURE 1: PRETEST-TEST DEVELOPMENT PROCESS: DISCUSSION-TO-CONSENSUS



Conduct of the exercise B-w/o-C simulating the "team" multi-person audit format

While the process followed prior to the group sessions for the B-w/o-C groups was the same as for the D-to-C groups, the group sessions were conducted in a different manner. Each session began with the audit junior staff member writing down on a blank sheet of paper his or her Preliminary score along with a reason for the score. The length of the reason given was to be similar to that provided in the audit working papers. This sheet of paper represented the working papers found within the typical "team" audit process. As noted by Ballau (1997),

reviewers’ conclusions on work done by subordinates is dependent, in part, on whether the subordinates had also provided written notes.

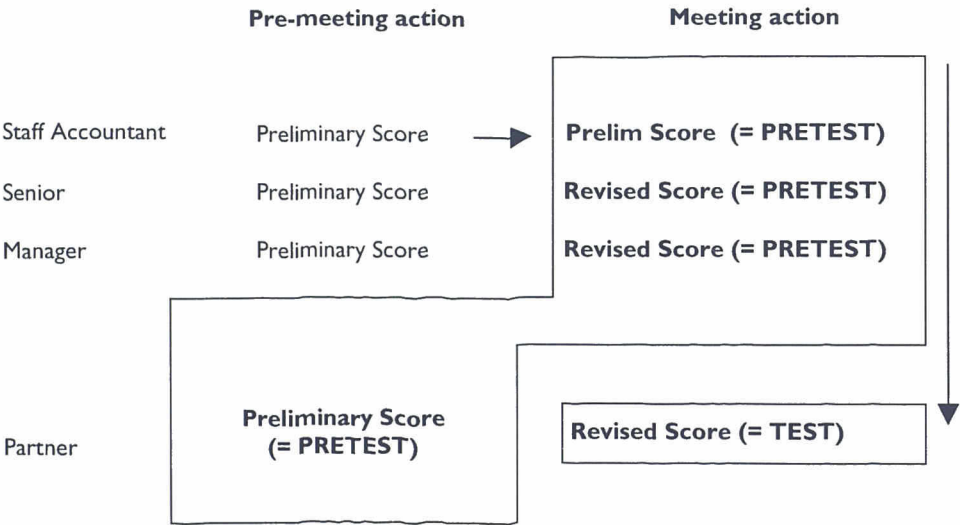
This paper with the score and reason on it was then passed to the audit senior. The senior would read what the junior staff member had written, review their own Preliminary score, form a current opinion (possibly revised from his or her Preliminary score) and write that current opinion score (Revised score), along with a brief reason, on the paper. The senior then passed the paper to the audit manager who repeated the same task. The manager passed the paper, with all three scores and reasons, to the partner who also repeated the task. This process simulated the working paper creation and flow found within the typical “team” audit approach, going from staff to senior to manager to partner.

This process was utilised in order to focus, exclusively, on the characteristic of written communication which is a primary, though not exclusive, means of communication within the “team” approach. While oral communication often does occur within the “team” approach, the goal of the research was to examine the impact of written (versus oral) communication on the process. Thus it was necessary that these participants limit themselves to written communication.

All participants remained in the room with the researcher present until the participants completed the process, guaranteeing that there was no oral communication between the subjects during the experiments.

Since the primary concern of this research was to observe the effect of the format on the final audit decision (and not necessarily on a complete choice-shift model format where all participants know the Pretest opinions of all other participants), the basis of comparison must focus on the partner’s final choice.

FIGURE 2: PRETEST–TEST DEVELOPMENT PROCESS: BALLOTING-WITHOUT-CONSENSUS



While in the D-to-C format the partner agrees with and accepts the group's Test score, there is no single agreed group score in the B-w/o-C format and the partner's Revised score is the final score accepted within the audit. Accordingly, the Pretest scores for the B-w/o-C teams consisted of the three scores written on the paper by the non-partner staff members and the Preliminary score brought to the group by the partner, while the partner's Revised score (written on the paper) was used as the Test score¹¹ (see **Figure 2**). Once this was completed for the first case, the participants followed the same process for the remaining two cases.

Recall that the two treatments (D-to-C and B-w/o-C) are not intended to be perfect replications of the "group" and "team" audit approach. Instead, the utilisation of these two treatments allows us to focus on two of the important characteristics of these approaches, those characteristics being the role of extensive interactive (oral) discussion (group) versus "upward" (or "downward") written communication (team) and consensus reaching (group) versus non-consensus reaching (team).

ANALYSIS

An analysis of all 372 data points (31 groups x 3 scenarios x 4 employment levels) indicated the presence of one outlier (a partner score on one scenario for the D-to-C treatment)¹². The scores of all 4 employment levels for this one scenario of the affected firm were deleted, allowing the analysis to be made on 368 data points. **Table 2** presents the average Pretest scores by employment level and average Test scores for the scenarios.

TABLE 2: AVERAGE (AND STANDARD ERROR) OF PRETEST AND TEST SCORES BY CASE, TREATMENT AND EMPLOYMENT LEVEL

	Discussion-to-Consensus			Balloting-without-Consensus		
	Account re-classification	Bad debt provision	In-ventory	Account re-classification	Bad debt provision	In-ventory
PRETEST						
Partner	87.33 (3.71)	68.67 (3.89)	77.33 (5.56)	89.69 (3.11)	63.13 (5.38)	75.94 (3.00)
Manager	88.67 (2.74)	65.67 (7.27)	77.00 (3.90)	89.38 (2.23)	60.31 (7.81)	69.06 (5.25)
Senior	92.00 (2.79)	66.33 (5.86)	80.00 (5.16)	93.75 (2.02)	56.88 (7.05)	65.63 (5.68)
Staff	86.67 (3.15)	70.00 (5.32)	73.67 (4.10)	90.00 (2.24)	54.38 (7.41)	66.25 (6.18)
TEST						
	89.33 (2.48)	67.14 (3.63)	79.33 (4.08)	91.88 (2.03)	61.25 (5.62)	72.81 (4.23)

The analysis began with a search for any movement towards either the maximum or minimum Pretest score, independent of employment level or experimental treatment. This would be the movement expected for the traditional choice-shift phenomenon. Table 3 summarises the differences between the average Pretest score and Test score¹³.

TABLE 3: SHIFT TOWARDS MINIMUM OR MAXIMUM PRETEST SCORES WITHIN GROUPS FOR ALL SCENARIOS

Scenarios	Treatment and basis of comparison	(A) Average pretest score ^a	(B) Average test score ^b	(C) = (A – B) Difference ^c
All Scenarios	D-to-C (Min.)	61.59	78.86	-17.27
Inventory	D-to-C (Min.)	59.33	79.33	-20.00
Acct reclass.	D-to-C (Min.)	76.67	89.33	-12.66
Bad debt prov.	D-to-C (Min.)	47.86	67.14	-19.28
All Scenarios	B-w/o-C (Min.)	54.27	75.31	-21.04
Inventory	B-w/o-C (Min.)	47.19	72.81	-25.62
Acct reclass.	B-w/o-C (Min.)	81.88	91.88	-10.00
Bad debt prov.	B-w/o-C (Min.)	33.75	61.25	-27.50
All Scenarios	D-to-C (Max.)	91.25	78.86	12.39
Inventory	D-to-C (Max.)	90.67	79.33	11.34
Acct. reclass.	D-to-C (Max.)	97.00	89.33	7.67
Bad debt prov.	D-to-C (Max.)	85.71	67.14	18.57
All Scenarios	B-w/o-C (Max.)	87.60	75.31	12.29
Inventory	B-w/o-C (Max.)	84.69	72.81	11.88
Acct. reclass.	B-w/o-C (Max.)	98.75	91.88	6.87
Bad debt prov.	B-w/o-C (Max.)	79.38	61.25	18.13

192 data points for B-w/o-C = ((16 groups)(3 scenarios)(4 staff levels))

176 data points for D-to-C = ((14 groups)(3 scenarios)(4 staff levels) + (1 group)(2 scenarios)(4 staff levels))

^a Average of each group's minimum (Min.) or maximum (Max.) Pretest score

^b Average of the Test scores for the respective Treatment

^c All differences were significant at the .0001 level

Because firms were assigned either to D-to-C or to B-w/o-C, but not to both treatments, the firms in the two treatment groups are mutually exclusive. In order to take account of this statistical design aspect, one needs to treat firms as being

nested into treatment. Firms and treatment were crossed with scenarios, because all firms and all treatment tested the same three scenarios. Crossed designs are far more common. In the present case, unless the factor firm was treated as nested, the standard errors and the F-ratios of a standard ANOVA table would have given potentially misleading results. Note that a small observed mean difference (column C on **Table 3**) would indicate that the Test score did, indeed, move towards the minimum (maximum) Pretest score.

The results show that the mean difference for each individual scenario and for the scenarios combined are all significantly different from zero at the .0001 level, indicating that the groups *did not* shift towards either the minimum or maximum Pretest score, independent of employment level and format of experimental treatment.

Next we turned to an examination of the movement towards the partner’s predisposition (Pretest scores) as implied by the first hypothesis. This hypothesis was tested by calculating the absolute difference between each group’s Test score and the Pretest scores assigned by the various employment levels observed¹⁴.

Recall that a small observed absolute mean difference would indicate that the Test score moved towards that of a specified Pretest score being observed¹⁵. **Table 4** shows the results of these differences by employment level.

TABLE 4: SHIFT TOWARDS PARTNER’S PRETEST SCORE

Level	Mean ^a of absolute differences of individuals’ pretest scores and their respective group test scores	
	<u>D-to-C</u>	<u>B-w/o-C</u>
Staff	14.03	18.44
Senior	11.08	16.35
Manager	11.65	13.65
(A)		
All non-partner levels	12.25	16.14
(B)		
Partners	5.28	4.06
(C) = (A – B)		
Difference of means	6.97 ^b	12.08 ^b

48 (= 16 groups x 3 scenarios) data points per level for B-w/o-C

44 (= (14 groups x 3 scenarios) plus (1 group x 2 scenarios)) data points per Level for D-to-C

^a All reported means and difference of means reported are significant at the .0001 level

^b The difference between the D-to-C and the B-w/o-C “Difference of Means” is significant at the .044 level.

The results reveal that, while the absolute differences for all employment levels are significantly greater than zero, there exists a smaller difference for the partners (5.28 and 4.06), indicating a greater movement *towards* the partners’ Pretest scores than for any other individual employment level or for all non-partner levels

combined. This result is true for both “group” and “team” format treatments. Furthermore, the difference between the partners’ movement towards their Pretest scores and the combined non-partners’ movement towards their Pretest scores is significantly different from zero ($p < .0001$) for both the D-to-C treatment (6.97) and the B-w/o-C treatment (12.08). These findings would indicate that we are able to reject the first hypothesis, implying that the partners’ personal preferences are the dominant ones within the group.

If the partner’s Pretest score had consistently been the maximum (or minimum) Pretest score, the shifting indicated in **Table 4** would not necessarily be indicative of movement towards the partner – i.e. it could simply represent movement towards the maximum (minimum) Pretest scores. However, further testing of the data (see **Table 5**) shows a lack of predominance of the minimum or maximum scores by any one employment level. The one exception was with the Account Reclassification scenario. Even here the partner Pretest score was not the predominant one. Therefore, the movement towards the partner’s Pretest score observed in **Table 4** is independent of the relative standing of the partner’s Pretest score within the group.

TABLE 5: PERCENTAGE OF TIMES INDIVIDUAL GROUP MAXIMUM AND MINIMUM PRETEST SCORES WERE ASSIGNED BY EACH RANK ^a

Level	All scenarios ^b		Account reclass. ^c		Bad debt provision ^b		Inventory ^b	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
Partner	20.31	27.46	22.01	25.00	24.32	25.58	14.63	32.56
Manager	27.34	22.54	31.93	17.86	21.62	30.23	26.83	20.93
Senior	21.88	28.17	16.02	37.50	27.03	18.61	24.39	25.58
Staff	30.47	21.83	30.04	19.64	27.03	25.58	34.15	20.93
TOTAL	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

^a If two or more members of a group had the same minimum (maximum) score, they were both counted in this calculation and the total observations increased proportionately

^b Unable to reject the null hypothesis that there is no difference in proportions within each column (Min. and Max.)

^c The null hypothesis that there is no difference in proportions with each column (Min. and Max.) is rejected at the .03 level for Min. and .01 level for Max. However, in both cases, the significantly disproportionate item was not the Partner but the Manager for the Min. and the Senior for the Max.

The second hypothesis addressed the issue of the format (treatment) used to arrive at the final risk assessment decision (Test score). Here a comparison was made of the relative difference in partner versus all non-partner shifts (as reported in **Table 4**) by treatment (D-to-C and B-w/o-C). Recall that in **Table 4** we calculated the shifting (Pretest versus Test score) and observed a significant predominance of the shifting towards the partner.

The results demonstrate a greater relative shift towards the partner’s score for the B-w/o-C (as compared to the D-to-C) treatment in two ways. First, the amount

of shifting under B-w/o-C from Pretest to Tests score for non-partner levels (16.14) is 3.98 times that for the partners (4.06), while it is only 2.32 times greater (12.25 versus 5.28) under the D-to-C treatment. Furthermore, a treatment-based comparison of the two sets of partner versus non-partner differences (12.08 versus 6.97) also shows a significant difference ($p < .05$). Accordingly, we were able to reject the second hypothesis, providing evidence that oral discussion with consensus has a significantly greater relative impact on the modification of the partner's Preliminary opinion than when using written communication (balloting) without consensus.

In summary, the data shows that the partner's personal preference regarding risk assessment remains the dominant one within the group. More importantly, however, the format of the experiment (simulating the "group" and "team" audit approaches) had a significant influence on the relative degree of that predominance. In other words, discussion aimed at consensus does provide a greater relative impact on the partners' decisions regarding risk assessment than does written communication not aimed at consensus.

IMPLICATIONS OF THE FINDINGS FOR THE PRACTICE OF PUBLIC ACCOUNTING WITHIN IRELAND

While there has been very little research on methods of audit practice – as it relates to group decision making – within Ireland, one might consider the applicability of the findings of this research to Irish practice by examining similarities in relevant cultural characteristics of both the US and Ireland. The classic measurements often used for such cross-country cultural comparisons are the cultural constructs developed by Hofstede (1980).

Hofstede (1991, p. 112) defines culture as a 'set of likely reactions of citizens with a common mental programming... [Exact] reactions [to a common situation] need not be found within the same persons, but only statistically more often in the same society'. He believes that culture represents a system of shared values and beliefs. The cultural constructs he identifies were the result of sampling over 100,000 employees of a large multi-national corporation in 53 countries.

Two of the relevant constructs that might be used to compare Ireland with the US on the issues involved in this research are Power Distance and Uncertainty Avoidance. Hofstede (1984, pp. 83–84) describes these two constructs as:

Power Distance: The extent to which the members of a society accept that power in institutions and organizations is distributed unequally. . . . The fundamental issue addressed by this dimension is how society handles inequalities among people when they occur.

Strong versus Weak Uncertainty Avoidance: Uncertainty Avoidance is the degree to which the members of a society feel uncomfortable with uncertainty and ambiguity. Weak Uncertainty Avoidance societies maintain a more relaxed atmosphere in which practice counts more than principles and deviance is more easily tolerated.

Both of these constructs are applicable to the issues addressed in this research. Power Distance addresses the issues of the hierarchical status of the group's participants in the study while Uncertainty Avoidance addresses the issues of needing to feel confident in the client management's representation prior to signing off on the audit.

Further examination of these constructs by Hofstede (1980) show that, in lower Power Distance countries, managers are seen as making decisions after consulting with subordinates, managers are more satisfied with participative superiors and employees are less afraid of disagreeing with their bosses (p. 199). In lower Uncertainty Avoidance countries, there is more acceptance of dissent (p. 184).

Both Ireland and the US score relatively low on both the Power Distance and Uncertainty Avoidance scales and are within one standard deviation of each other when compared to all countries observed by Hofstede. From this, one can infer a relative similarity between the two countries on these relevant constructs, with the potential implication that the findings of this research are applicable to Ireland. Obviously, considerable further research would be needed to verify this.

DISCUSSION OF FINDINGS

The findings demonstrate a significant difference in the impact of the discussion and consensus building format versus the sequentially reviewed written work paper without consensus format upon the relative modification of partners' attitudes towards the risk assessment. While both formats demonstrated a predominance of movement towards the Preliminary opinion of the partners – a movement that might well be expected – the results demonstrated that the modification of partners' Preliminary opinions, relative to the modification for the non-partners, was significantly smaller when relying upon a written communication, non-consensus building format.

These findings indicate that in the multi-person, multi-employment level audit review process common to most audits, partners still play the dominant role in the process. However, in such a setting, the format of oral communication and consensus building – important characteristics of the “group” approach – has a greater impact on the partner's risk assessment decisions as compared to the format of written communication and no consensus building – important characteristics of the “team” approach. That is, it would appear that partners listen, react to and internalise the thoughts and opinions of the staff to a greater extent when such views are expressed orally with a consensus reaching goal as opposed to being written in the working papers and without a consensus reaching goal.

This clearly implies that in addition to the nature of the investigative (audit) procedures used when conducting an audit, the mere format of communications utilised and the degree of staff/partner consensus reaching sought by the partners can also influence the partners' final decisions regarding issues of risk assessment. Furthermore, if firms wish to utilise their staff as more than mere fact gatherers (which some, though not all, audit partners might think is wise), then the oral discussion and consensus building format becomes a valuable tool. These findings

are important as we seek to gain greater understanding and appreciation of the movement towards the “group” and away from the “team” approach to auditing.

The findings also demonstrated an interesting conclusion for research in other auditing areas that utilise the choice-shift framework, that being the dominance of the hierarchical employment level effect over the risk extreme effects of the traditional choice-shift model. This is important in interpreting all prior choice-shift research within the audit domain, independent of its relationship to the “group” versus “team” issue. Since a multiple hierarchical employment level composition is the basis of most multi-person audit groups, previous research findings based on single-hierarchical employment level subjects (e.g. groups consisting of only managers or only partners) might be of limited value and not applicable to the multi-hierarchical employment level setting.

The sample was limited to subjects from local public accounting firms, providing us with a greater opportunity to use natural (versus ad hoc) groups and to study the characteristics of a subset of audit firms rarely examined. However, further research would be needed before these findings could be definitively applied to the larger international accounting firms. Furthermore, while inter-country cultural characteristics would imply that the findings of the current study might well be applicable in Ireland as well as the US, further research would be needed before a definitive conclusion on this could be reached. As with all empirical research, certain limitations should be noted. First, due to time restrictions, the research design did not utilise both the D-to-C and the B-w/o-C format within each group of subjects. Second, the firms themselves selected the individual subjects, and non-participation bias was not examined. Suggested further research would thus include the utilisation of subjects from the larger international firms and cross-country examinations along with experimentation using both treatments within each subject group.

NOTES

- ¹ The recency effect notes that the importance placed on individual pieces of information is a function, in part, of the order in which the information was presented to the decision maker.
- ² “Hierarchical level” refers to the employment levels within a public accounting firm – i.e. partner, manager, senior accountant and staff level accountant.
- ³ Through discussion with one Big Five firm partner and a partner at a smaller firm, as well as from the personal experience of the researchers, a list of 22 different types of audit decisions was constructed. Copies of this list were sent to 169 offices of CPA firms with the request that it be evaluated by at least one person currently employed in each of the four hierarchical levels at the firm. The respondents were asked to evaluate each type of audit decision on three criteria: (a) their perception of its level of risk; (b) whether one person or a group usually makes this decision; and, (c) in the case of group decisions, the number and employment levels of the participants. Subjects were given room to list additional audit decisions which groups might make. The response rate of

approximately 24 per cent (160 subjects) is consistent with other mail surveys (Emory, 1985).

- 4 Materiality was controlled by stating that the situations would be material without designating a level of high or low. The situations implied that if an error were made as to the correct decision, the results would be material and would be of sufficient magnitude to be of concern to the decision maker(s). Each case explicitly stated that no fraud was suspected.
- 5 A pilot study was used to determine the anticipated variance in the subject responses which, in turn, determined the necessary sample size. College students at the senior undergraduate and graduate level were used as subjects for the pilot study.
- 6 "Local" is defined as having three or fewer offices and all are located in one state.
- 7 The basis of firm participation was not random. Instead it consisted of those firms which volunteered when asked to participate.
- 8 The managing partner or another contact person within each firm made the subject selection. That selection was based upon the availability and willingness of the subjects to participate.
- 9 While there was no further control to ensure zero communication between subjects prior to the experimental session, the discussions observed by the researcher during the "D-to-C" group exercise indicated that prior discussion had not taken place for these subjects. Furthermore, discussions with the contact person indicated to us that they usually did not even inform individual participants of the names of the other participants.
- 10 While not participating in the groups' discussions, the researcher present monitored the sessions to ensure that all participants remained in the group setting, no other members of the firm joined the group, and that participants followed the instructions. These initial instructions included the requirement that the cases be discussed in order and that once the group decided a Test score for a single case, they would not return to it. After the groups completed all cases, they placed the paper with the results written on it in a folder and gave it to the researcher.
- 11 While the Preliminary scores represented the initial opinions brought to the process by the individual participants of the B-w/o-C process, the Pretest scores formed the data base that was passed onto the partners as they made their final decision (Test scores). Since a goal of the experiment was to examine the effect of the treatment processes on the final decision, it was necessary to base the analysis on the data that the decision maker(s) used when making the decision. All tests were also run using Preliminary scores instead of Pretest scores for the B-w/o-C group. All conclusions reached were the same as when using the Pretest scores.
- 12 Cook's D statistics and leverage plots identified this one outlier. Cook's D statistic was three times as large as the next largest value for the rejected firm. Eliminating this one scenario of this one firm did not change the overall conclusions.
- 13 Since the Test score is always greater than the Pretest minimum and less than the Pretest maximum, it was not necessary to calculate absolute values. Furthermore, since there was one minimum (maximum) Pretest score and one Test score per group, the difference between the Pretest score average of all the groups combined and the Test score average of all groups combined also equals the average of the differences from each individual group.

14
$$\frac{| [\text{Test Score}_{i,j} - \text{Partner (Manager, Senior, Staff) Pretest Score}_{i,j}] |}{n}$$

where: i = ith team or group; jth scenario; n = number of observations

15 Since the observed differences might have been either positive or negative, it was necessary to use an absolute value. Accordingly, the difference between average Pretest score (by level) and the Test scores for the entire population becomes irrelevant.

APPENDIX

Summary of question asked and scenarios presented

Following is the exact wording of the question asked for each scenario:

Based upon your subjective beliefs, what percentage would you insert in the blank space of the following sentence: “Based on my experience as an auditor, I would want to be at least ____% personally confident that client management’s representation for the [the subject item of that scenario] was adequate in order to reject [the action recommended by the auditor as specified in the scenario]” NOTE: A 100% response indicates that in order to signoff on the representation for [the specific subject item of the scenario], you will NOT ACCEPT ANY UNCERTAINTY in that representation. Thus, a high percentage indicates that you would demand a high personal level of confidence before you would signoff. A lower percentage indicates that you would be willing to signoff with a lower level of personal confidence.

Following is a summary of the three scenarios:

The first scenario involved the proper valuation for ending inventory. Consigned goods have been commingled with regular inventory both physically and in the accounts. The auditor had to decide whether to accept the 10 per cent adjustment suggested by the client or do extensive additional audit work. This additional audit work would exceed budgeted hours, not be paid by the client, and result in the loss of revenue for the auditor’s firm. In addition, doing the additional work would cause a delay in the issuance of the audit report, creating an unfavourable response from the client’s creditors. The client threatened to find another auditor if a delay occurred. The perception of both audit risk (giving a clean opinion on potentially incorrect financial statements) and litigation risk (from investors who rely on those statements and from the client for a delayed issuance of the financial statements) was high in this scenario.

The second scenario involved a not-for-profit hospital that held a large federal grant that provided reimbursements for allowable expenses. Certain non-allowable expenses, expected to be recorded in a “non-reimbursable” account, had become commingled in the “reimbursable” account (i.e. account reclassification). The auditors discovered two misclassified expenses in the sample selected with this error rate exceeding statistical limits. The client contended that the account balance for the reimbursable account would be correct after adjusting for these errors. The auditor could only be sure by doing extensive additional auditing. If the client incorrectly

stated the "reimbursable" account balance and the Federal authorities later discovered this error, the auditor's firm could be barred from auditing not-for-profit organisations that receive Federal grant programmes. Since such audits comprise over 50 per cent of the firm's revenue base, the auditors classified the audit risk as being high. However, auditors perceived litigation risk as being low.

The third scenario involved the estimation of bad debt expense of a new client. The client contended that its percentage was within industry standards, and it was also within the "statistically acceptable bounds" based on a normal distribution and given standard deviation. However, a slightly smaller true standard deviation would have pushed it outside these bounds. The client later agreed to increase the percentage for the provision by an additional one per cent but refused to go as high as the auditor recommended. Even with the one per cent increase, a material difference still existed in relation to the auditor's recommendation. In addition, the client threatened to use another auditor if an agreement could not be reached. The client represents an expansion of the auditor's business into a new industry. Expansion into this industry could significantly affect the firm's total revenue, and the firm does not want to hurt its reputation within this industry. The loss of this client over a disagreement on a valuation dependent in part upon industry norms may signal incompetence to the other members of this industry. The auditors perceived both audit risk (giving a clean opinion on potentially incorrect financial statements) and litigation risk (investors relying on those statements) as high in this scenario.

REFERENCES

- Asare, S. and McDaniel, L. (1996). 'The Effects of Familiarity with the Preparer and Task Complexity on the Effectiveness of the Audit Review Process', *The Accounting Review*, Vol. 71, No. 2, pp. 139-160.
- Ballau, B. (1997). 'The Effect of Cognitive Limitations on the Risk of Accepting Insufficient Evidence in Audit Working Paper Review', *Working Paper*, Auburn University.
- Bambers, E. (1983). 'Expert Judgment in the Audit Team: A Source Reliability Approach', *Journal of Accounting Research*, Vol. 21, No. 2, pp. 396-412.
- Bambers, E., Watson, R. and Hill, M. (1996). 'The Effects of Group Support System Technology on Audit Group Decision Making', *AUDITING: A Journal of Practice & Theory*, Vol. 15, No. 1, pp. 122-134.
- Barrett, J. J. and O'Malley, T. (1976). 'Choice-Shifts and Auditor Opinions', Symposium on Auditing Research, Champaign, IL. University of Illinois, pp. 187-204.
- Bateson, N. (1966). 'Familiarization, Group Discussion, and Risk-Taking', *Journal of Experimental Social Psychology*, Vol. 2, No. 2, pp. 119-129.
- Bem, D., Wallach, M. and Kogan, N. (1965). 'Group Decision-Making Under Risk of Aversive Consequences', *Journal of Personality and Social Psychology*, Vol. 1, No. 5, pp. 453-460.
- Brock, T. (1973). 'Communicator - Recipient Similarity and Decision Change', in *Advances in Communication Research*, Mortensen, C. and Sereno, K. (eds.), New York: Harper & Row.
- Emory, C.W. (1985). *Business Research Methods* (3rd edn), Homewood, Illinois: Richard D. Irwin, Inc.
- Fraser, C., Gouge, C. and Billig, M. (1970). 'Risky Shifts, Cautious Shifts, and Group Polarization', *European Journal of Social Psychology*, Vol. 1, No. 1, pp. 7-30.

- Hofstede, G. (1980). *Culture's Consequences: International Differences in Work Related Values*, Beverly Hills: Sage.
- Hofstede, G. (1984). 'Culture Dimensions in Management and Planning', *Asia Pacific Journal of Management*, Vol. 1, No. 1, pp. 81-99.
- Hofstede, G. (1991). *Cultures and Organizations: Software of the Mind*, New York: McGraw Hill.
- Johnson, E. (1995). 'Effects of Information Order, Group Assistance, and Experience on Auditors' Sequential Belief Revisions', *Journal of Economic Psychology*, Vol. 16, No. 1, pp. 137-160.
- Marquis, D. (1962). 'Individual Responsibility and Group Decisions Involving Risk', *Industrial Management Review*, Vol. 3, No. 2, pp. 8-23.
- Martin, J. and Previts, G. (1982). 'The Risk Preference Profiles of Practising CPAs: Some Tentative Results', *Accounting and Business Research*, Vol. 13, No. 49, pp. 21-28.
- Marxen, D. (1990). 'A Behavioral Investigation of Time Budget Preparation in a Competitive Audit Environment', *Accounting Horizons*, Vol. 4, No. 2, pp. 47-57.
- McGrath, J. (1993). 'The JEMCO Workshop - Description of a Longitudinal Study', *Small Group Research*, Vol. 24, No. 3, pp. 285-306.
- Myers, D. and Lamm, H. (1976). 'The Group Polarization Phenomenon', *Psychological Bulletin*, Vol. 83, No. 4, pp. 602-627.
- Nelson, D. and Quick, J. (2003). *Organizational Behavior: Foundations, Realities, and Challenges* (4th edn). Cincinnati, Oh: South-Western.
- Nemiroff, P. and King, D. (1975). 'Group Decision Making Performance as Influenced by Consensus and Self-Orientation', *Human Relations*, Vol. 28, No. 1, pp. 1-21.
- Newcomb, T. (1961). *The Acquaintance Process*, New York: Holt, Rinehart and Winston.
- Pennington, N. and Hastie, R. (1988). 'Explanation-Based Decision Making: Effects of Memory Structure on Judgment', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 14, No. 3, pp. 521-533.
- Reckers, P. (1978). *An Examination of the Choice-Shift Phenomenon in an Auditing Framework*, Ph.D. Dissertation, University of Illinois at Urbana-Champaign.
- Reckers, P. and Schultz, J. (1993). 'The Effects of Fraud Signals, Evidence Order, and Group-Assisted Counsel on Independent Audit Judgment', *Behavioral Research in Accounting*, Vol. 5, pp. 124-144.
- Rich, J., Solomon, I. and Trotman, K. (1997). 'Multi-Auditor Judgment/Decision Making Research: A Decade Later', *Journal of Accounting Literature*, Vol. 16, pp. 86-126.
- Robbins, S. (1996). *Organizational Behavior: Concepts, Controversies, Applications* (7th edn), Englewood Cliffs, NJ: Prentice Hall.
- Schultz, J. and Reckers, R. (1981). 'The Impact of Group Process on Selected Audit Disclosure Decisions', *Journal of Accounting Research*, Vol. 19, No. 2, pp. 482-500.
- Shaw, M. (1976). 'The Social Environment: Group Structure', in *Group Dynamics*, New York: McGraw-Hill Book Co., pp. 244-250, 262-272.
- Sherif, M., Harvey, O.J., White, B.J., Head, W.R. and Sherif, C.W. (1961). *Intergroup Conflict and Cooperation*, Norman, OK: Institute of Group Relations.
- Siegel, S. and Zajonc, R. (1967). 'Group Risk-Taking in Professional Decisions', *Sociometry*, Vol. 39, No. 4, pp. 339-350.
- Solomon, I. (1987). 'Multi-Auditor Judgment/Decision Making Research', *Journal of Accounting Literature*, Vol. 6, pp. 1-25.
- Stone, D. and Dilla, W. (1994). 'When Numbers are Better than Words: The Joint Effects of Response, Representation and Experience on Inherent Risk Judgments', *AUDITING: A Journal of Practice & Theory*, Vol. 13, Supplement, pp. 1-19.

- Stoner, J. (1961). 'A Comparison of Individual and Group Decisions Involving Risk', MA Thesis, School of Industrial Management, Massachusetts Institute of Technology.
- Sutton, S. and Hayne, S. (1997). 'Judgment and Decision Making, Part III: Group Processes', in Behavioral Accounting Research, Arnold, V. and Sutton, S. (eds.), Sarasota, FL: American Accounting Association.
- Vidmar, N. (1970). 'Group Composition and the Risky Shift', *Journal of Experimental Social Psychology*, Vol. 6, No. 2, pp. 153-166.
- Wallach, M. and Kogan, N. (1965). 'The Roles of Information, Discussion, and Consensus in Group Risk Taking', *Journal of Experimental Social Psychology*, Vol. 2, No. 1, pp. 1-19.
- Wallach, M., Kogan, N. and Burt, R. (1967). 'Group Risk Taking and Field Dependence-Independence of Group Members', *Sociometry*, Vol. 30, No. 4, pp. 323-339.
- Zimmerman, R., Flaherty, D. and Murray, M.A. (1999). 'The National Results of the 1998 Management and Accounting Practice Survey', *The CPA Journal*. July, pp. 40-45.