Department of Commerce

Touche Ross

WORKING PAPER SERIES

INDIVIDUAL VERSUS GROUP PERFORMANCE IN FINANCIAL DECISION MAKING: A TEST OF THREE THEORIES

by

*Errol R. Iselin

1987 No. 1

Touche Ross International

INDIVIDUAL VERSUS GROUP PERFORMANCE IN FINANCIAL DECISION MAKING: A TEST OF THREE THEORIES

by

*Errol R. Iselin

1987 No. 1

*Senior Lecturer, Department of Commerce, University of Queensland.

I would like to thank Ken Trotman and Ian Zimmer for kindly permitting me to use their case study material, and John Hoggett, Ken Leo and Trevor Wilkins for assistance in obtaining participants for the research. I would also like to thank Paul Ryder, Ian Zimmer, and participants at a workshop at Griffith University for their helpful comments on an earlier draft of the paper.

REG. DON.
1988
ECOM

INDIVIDUAL VERSUS GROUP PERFORMANCE IN FINANCIAL DECISION MAKING: A TEST OF THREE THEORIES

1. INTRODUCTION

Several accounting researchers[1] have studied aspects of the relative decision-making performance of individuals and groups in financial analysis tasks. Following earlier work in psychology (for example Einhorn, Hogarth and Klempner [1977]), the accounting studies have usually involved audit and financial distress (predicting impending bankruptcy or loan default) tasks. These studies can be related to practice because audit and loan decisions in practice are made by individuals and groups (interacting and noninteracting).

The research studies typically compare the performance of individuals, composites, and interacting groups. A composite decision is the mean decision of a set of individuals who do not interact. When the decision alternatives are binary, as in predicting bankruptcy and loan default, the mean decision is a majority rule decision.[2] Composites are also known as equal weighted composites, staticized sets of individuals, statistical committees, and composite groups.

The findings in the accounting literature regarding the performance of individuals and composites have been relatively consistent. Einhorn *et al.* [1977] argue that there are two types of error in individual decision making – random error around the mean and systematic bias. Systematic bias is the difference between the mean of a population of individual decisions and the correct value. Einhorn *et al.* [1977] show that composites reduce the random error or inconsistency in individual performance, and under certain conditions,[3] increase the mean accuracy of decisions. The findings in accounting have largely supported this theory. Since this issue seems fairly well settled in accounting it will not be pursued in this research. Suffice is to say that the findings here[4] were in accordance with the Einhorn *et al.* [1977] theory.

However, a major issue which is still unresolved concerns the relative performance of composites and interacting groups. Consequently, this paper focuses on that issue.[5] The findings here are quite inconsistent. Libby, Trotman and Zimmer [1987] cite three reviews in the psychology literature which present all three possible outcomes for relative performance. Hackman and Morris [1975] conclude that "for many tasks" composites outperform interacting groups; Fischer [1981] states that in probability forecast tasks there is "little or no difference" in performance; while Rohrbaugh [1979] concludes that interacting groups are better than composites. In accounting the findings are similarly inconsistent. For example, Trotman, Yetton and Zimmer [1983] found composites outperformed interacting groups; Trotman and Yetton [1985], Libby *et al.* [1987], and Chalos [1985] found no significant differences; while Trotman [1985] found interacting groups were better than composites.

In view of these conflicting findings, it should not be surprising that at least two different theories about the relative performance of composites and interacting groups can be found in the accounting literature - the expertise theory and the information load theory. Briefly, the expertise theory states that in certain conditions[6] interacting group members will follow the most expert member and as a result interacting groups will outperform composites. The information load theory states that interacting groups can process high information loads better than individuals and as a result, where information loads are high interacting groups outperform composites. These theories are not necessarily inconsistent, yet they have been presented as distinct explanations. The theories will be described in detail and reviewed in the next section where it will be argued that there are problems in the empirical tests of both. The review of the expertise theory leads to a third theory - the equivalence theory. This theory states that experiments supporting the expertise theory contain a learning confounding, and when this confounding is controlled there will be no difference in the performance of interacting groups and composites (i.e. they will have equivalent performance).

Since the evidence supporting each of the three theories is not conclusive, and since the equivalence theory is inconsistent with the other two, further research is needed to clarify the area. This paper describes an experiment designed to make some contribution in that regard. The experiment provides a test of each of the theories. Briefly, the experiment finds support for the equivalence theory but no support for the expertise or load theories.

The paper proceeds as follows. The next section presents the theories in more detail, reviews them, and develops theoretical propositions which will be tested in the empirical part of the research. Following sections present the empirical research method, results and discussion, and finally, the conclusions that might be drawn from the study.

2. THEORY AND PROPOSITIONS

The expertise theory has been advanced in the accounting literature by Libby, Trotman and Zimmer (Libby *et al.* [1987], Trotman [1985]). In detail the theory states that, where,

(1) a decision task produces a reasonable level of systematic bias, and permits interacting group members to identify relative expertise, and

(2) the interacting group members have reasonably different levels of expertise,

interacting group members will follow the decision of the most expert member, and as a result, interacting groups will outperform composites. The theory argues that a reasonably complex task is required to satisfy condition (1). Such a task produces a reasonable level of systematic bias. Einhorn *et al.* [1977] show that, where the probability of identifying the best member in a group is less than one (this condition will apply in this research), a standardized bias[7] level greater than one is necessary to give groups operating under the expertise theory the potential to outperform composites. The higher the standardized bias the greater the opportunity for superior performance by groups. A reasonably complex task is also necessary to permit group members to identify relative expertise. Complex tasks normally require the division of the problem into subproblems and the making of various calculations. The discussion of these matters should help group members identify relative expertise. Condition (2) is necessary because if the most expert group member is only marginally better than the others, following the expert's decision will produce a result only marginally better than the average for the individuals (the composite result). However, if the most expert member is significantly better than the others and they follow the expert's decision, the group result will be significantly better than the composite (average individual) result.

Trotman [1985], in an audit review task, found interacting groups performed significantly better than composites. This finding was advanced as support for the expertise theory. However, a close examination of Trotman's [1985] experiment reveals a learning or testing confounding. Trotman's [1985] subjects performed his audit review task twice. They first performed the task as individuals, with this performance being the base for the calculation of composite performance. They then performed the task a second time in interacting groups. According to Campbell and Stanley [1966] this procedure results in a testing confounding because "people usually score higher when they take an achievement or an intelligence test the second time." (Huck, Cormier and Bounds [1974] p. 235). In other words, subjects <u>learn</u> from the first administration of a test and consequently perform better on a second administration. This learning alone could have caused Trotman's [1985] interacting groups to perform better than his composites. Thus the equivalence theory stated in the previous section is an alternative explanation of Trotman's [1985] results (i.e. an alternative to the expertise theory).

In order to test these two theories this research manipulated two independent variables - learning trials and decision-making unit. Learning trials was manipulated by having subjects make the same set of decisions twice i.e. in two trials in a repeated measures fashion. The decision-making unit was manipulated as follows. In the first trial all subjects made the decision set as individuals. For the second trial, subjects were partitioned (randomly) in the ratio one third to two thirds. The one third subset then made the decisions again as individuals. In this "individual" subset, composites of three were constructed randomly. The two-thirds subset made the decisions in the second trial as interacting groups. There were two types of interacting group - homogeneous and diverse. Homogeneous groups consisted of three individuals who were either all expert or all novice financial decision makers. There were an equal number of expert and novice groups. Diverse groups consisted of three individuals - one expert and two novices. The dependent variable in this research was decision accuracy. The decisions made had correct answers.

Given the independent variable manipulations described in the previous paragraph, the equivalence theory leads to the following research propositions:

P1(a): Learning will occur from the first trial composite decisions to the second trial decisions (composite decisions for individuals; interacting group decisions for homogeneous and diverse groups). Learning will be the same for all three decision-making units, i.e. the variables learning and decision-making unit will not interact.

P1(b): In the second trial, there will be no difference in the performance of diverse groups, homogeneous groups, and individual composites, after adjusting for any first trial composite performance differences.

Proposition 2 follows from the expertise theory.

P2: In the second trial, diverse groups will outperform homogeneous groups and individual composites, after adjusting for any first trial composite performance differences.

In P1(b) and P2 it is necessary to adjust for first trial performance differences because subjects cannot be randomly assigned to expert and novice categories. As a result the three decision-making units may not be exactly equivalent in the first trial. This adjustment will be described in detail in the method section.

The argument advanced by the expertise theory to support P2 is as follows. In the reasonably complex decision-making task used, the novice decision makers in the diverse groups will recognize the expert and follow his/her decisions. As a result, the diverse group's decisions in the second trial will be better than their composite decisions in the first trial. However, this improvement cannot occur with homogeneous groups because there, all group members perform about the same. There is no scope for poorer members to follow a better member. This type of improvement also cannot occur in the individual decision-making unit because there is no interaction in which novices could follow experts.

The information load theory comes from the work of Chalos and Pickard [1985]. Chalos and Pickard [1985] studied the relative performance of individuals and interacting groups and found that the latter outperformed the former. These performance differences were due to two factors: (1) groups had improved decision consistency, and (2) groups could process high information load better than individuals. While they do not discuss how these findings might apply to a <u>composite</u> v. group comparison, inferences can be made. As shown by Einhorn *et al.* [1977], composites reduce the random error or inconsistency in individual performance. Consequently, at least some of the Chalos and Pickard [1985] consistency effect would <u>not</u> apply to a composite v. group comparison. It is not clear if some of the effect would remain in such a comparison. As stated in the introduction, the research described in this paper will focus on effects on means and will not study random error (consistency). The information load effect found by Chalos and Pickard [1985] did apply to means. It will therefore be investigated here as a possible explanation of the superior performance of interacting groups over composites.

However there were problems in Chalos and Pickard's [1985] test of the theory. First, the statistical significance testing of the findings was incomplete. Second, there is an important external validity problem in the experiment. In the low load condition, each subject received three selected cues for each firm and on the basis of these cues made a loan default decision. In the high load condition, subjects received the three cues plus full financial statements. It would seem that making loan default decisions on the basis of only three cues would be a very difficult task, and one unrepresentative of practical loan default decision making where many more cues would be available. Hence, further testing of the load theory is necessary.

This research provides a further test of the theory. To do so, information load is included in the research as an independent variable. Proposition 3 follows from the load theory.

P3: The information load and decision-making unit variables will interact as follows. In the high load condition, in the second trial, homogeneous and diverse groups will outperform individual composites, after adjusting for any first trial composite performance differences (homogeneous and diverse groups will perform at the same level). This effect will either not occur at the low load level or will occur to a significantly reduced extent.

This proposition follows from the load theory's argument that: (1) interacting groups can outperform individuals and composites at high loads, (2) such an effect will not occur, or will be reduced at low loads and (3) the effect is independent of the type of group.

3. METHOD

3.1 Subjects

Three hundred and sixty volunteer subjects were used. All had practical administrative decision-making experience, which ranged from 12 months to 20 years with a mean of 4.9 years. Subjects were employed in business and government administration and all had completed at least one course in accounting. Subjects had a diversity of experience and were not all loan officers or auditors as in other research in the area (e.g. Trotman [1985], Libby *et al.* [1987]). A diversity of experience was necessary to classify subjects into expert and novice categories, to operationalize the decision-making unit variable.

3.2 Design

A 2 x 3 x 2 factorial design was used where there were: (1) two levels of information load, (2) three levels of decision-making unit, and (3) two levels of task learning. Task learning was a repeated measures factor. Subjects were randomly assigned to the two levels in variable 1 and to the individual and interacting group conditions in variable 2. Subjects were assigned to the three types of interacting group in variable 2 on the basis of their expertise. The procedure will be described below.

3.3 Experimental Task

This section discusses in turn (1) the general nature of the experimental task, (2) the manner in which the independent and dependent variables were operationalized, and (3) the experimental procedure.

3.3.1 General Nature of the Task

The task was an adaptation of the bankruptcy prediction task developed by Libby *et al.* [1987]. Libby *et al.* [1987] constructed financial profiles on a population of firms that was defined as follows: companies listed on the Sydney Stock Exchange between January 1970 and December 1979, that were; (1) classified as land developers in the Australian Stock Exchange Journal; and (2) survived for at least five years. There were 39 firms in total. Twelve failed prior to December 1979. Libby *et al.* [1987] tested the ability of a discriminant analysis model using the Lachenbruch cross-validation procedure to predict the failure/non-failure of the firms. Using five financial ratios for the year prior to failure for failed firms, or for a random year for non-failed firms, the model correctly predicted 84.1% of cases.

For the research described in this paper, a random sample of ten firms was selected from the population described in the previous paragraph. Three of these firms failed prior to December 1979. A financial profile was constructed for each firm. The type of information included in each profile will be described in the next section. The information was for the following period: for failed firms - the year prior to the year of failure, and the year three years prior to the year of failure; and for non-failed firms - the year prior to a random year in the 1970's, and the year three years prior to the random year.

The ten firms were randomly divided into two subsets of four (one failure) and six (two failures). A decision-making booklet was constructed for each subset. The booklet contained the financial profiles for the firms which were in random order and unnamed. A cover sheet defined the population, the method of sample selection, the period covered by the financial profiles, and the decision-making task required which was to decide whether, in one year's time, each firm would have failed or survived (i.e. a dichotomous decision was required). Subjects were advised that their performance would be measured in terms of the number of correct decisions i.e. type I and II errors would have equal cost. Subjects were not advised of the proportions of failure/non-failure in the sample or the population. As in Libby *et al.* [1987] the population proportion was regarded as "a component of the expertise that the participants could bring to the task." (recall that subjects were given the population definition).

The experimental task used here is similar to those used by a number of other studies e.g. Libby and Blashfield [1978], Zimmer [1981], Chalos [1985], Chalos and Pickard [1985], and Libby *et al.* [1987]. However, unlike all of these studies except Libby *et al.* [1987], this research uses representative failure/non-failure proportions which overcomes external validity problems.

3.3.2 Operationalization of Variables

Information load was manipulated over two levels. For each firm (i.e. decision) the low load level comprised five financial ratios for each of the two annual periods described in the previous section. The ratios were: (1) earnings (before interest and taxes) over total tangible assets, (2) cash flow (profit plus depreciation) over total liabilities, (3) current assets over current liabilities, (4) total liabilities over shareholders funds, and (5) retained earnings over total tangible assets. Hence low load subjects received ten cues for each decision. According to Streufert [1972] this is about the optimum level for human information processing. High load subjects received for each firm for each of the two periods, low load information plus a balance sheet and a profit and loss statement. The format of these two statements was standardized for all firms. Footnotes were omitted. This operationalization of information load using ratios and financial statements (without footnotes) is similar to the one employed by Chalos and Pickard [1985]. Subjects were randomly allocated to the two load conditions. Hence, in respect of this variable we have a true experimental design (Campbell and Stanley [1966]).

Decision-making unit was operationalized to three levels (1) individual, (2) interacting homogeneous group, and (3) interacting diverse group. This was done as follows. In a first trial (trial 1) subjects made individual decisions from the booklet containing four firms. They were not at this stage given feedback about the accuracy of those decisions. Then in a second trial (trial 2) subjects made individual decisions from the booklet containing six firms. Again they were not given feedback. Subjects were then randomly divided into two subsets in the ratio 1:2. Those in the two-thirds subset were further divided into expert and novice categories in the ratio 5:7. The basis of the classification was the total number of correct decisions in trials 1 and 2. The expert and novice subjects were then assigned to homogeneous and diverse interacting groups such that homogeneous groups contained either three experts or three novices and diverse groups contained one expert and two novices. An equal number of homogeneous and diverse groups were constructed. In the homogeneous groups, there were an equal number of expert and novice groups. In a third trial (trial 3) the homogeneous and diverse groups made the same decision set as in trial 2 in a fresh booklet as an interacting group. The one-third subset of subjects which was not formed into groups made the same decision set as in trial 2 (in a fresh booklet) once again as individuals. They were instructed to try to improve the decisions made in trial 2. These participants who completed all three trials as individuals were randomly assigned to composites of three individuals. For each composite, composite scores for sessions two and three were obtained. For the interacting groups a composite score for trial 2 was obtained (from the individual trial 2 scores of those in the group).

The group size used in this research (3) is the same as that used in a number of other studies in the area e.g. Chalos and Pickard [1985], Libby *et al.* [1987]. Einhorn *et al.* [1977] found that in the types of groups being studied in this research, performance improved very little as group size increased beyond the small group.

Expertise was measured here over two trials to help minimize a regression to the mean (statistical regression) confounding. According to Campbell and Stanley [1966] this confounding arises when subjects are classified as high or low performers on the basis of only one test. One test normally measures performance with error, and as a result, some subjects classified as high (or low) performers would not be consistently so and would regress to the mean in a second test. Libby *et al.* [1987] measure expertise using only one test, and consequently their work contains this confounding. Really several tests are necessary to find consistent high and low performers. The use of two tests here is, therefore, not a perfect solution to the problem, but is better than using only one. The sole purpose of trial 1 in this research was the minimisation of the statistical regression confounding.

Expertise theory requires interacting group members to have reasonably different levels of expertise. In this experiment subjects' total scores on trials 1 and 2 ranged from 3 to 10 (only two obtained 10). Hence this research has captured a reasonably broad range of expertise. This should be sufficient for the expertise theory to apply to the interacting diverse groups. Expertise theory also requires a decision task that produces a reasonable level of systematic bias. In this research the standardized systematic bias in trial 2 was 2.39 for the low load condition and 2.61 for the high load condition. According to Einhorn *et al.* [1977] these levels are sufficient to permit diverse groups operating under the expertise theory to outperform composites.

In respect of the decision-making unit variable, we do not have a true experimental design. This is because although there were some random allocations in the operationalization of the variable, subjects could not be randomly allocated to the expert and novice categories. In this case, a nonequivalent control group design is appropriate (see Campbell and Stanley [1966], and Cook and Campbell [1979]), and such a design is used here. This design collects both pretest and posttest scores, and adjusts the latter for any differences in the former. ANCOVA is an appropriate method of analysis (see Reichardt [1979]), and it is used here. Trial 3 scores (the posttest) will be the dependent variable and trial 2 scores (the pretest) the covariate.

Task learning was operationalized by comparing performance on trials 2 and 3. Thus it can be established if learning occurs (i.e. performance improves) from trial 2 to trial 3. Note that subjects had the same sample of firms in both these trials.

The dependent variable in this research was decision-making performance. This was operationalized to the number of correct decisions in each trial. There was a correct answer for each decision.

3.3.3 Procedure

Trials 1, 2 and 3 were run in three separate sessions about a week apart. About 15 to 30 subjects participated at a time. Interacting group and individual decision making were not conducted in the same room at the same time. Subjects were instructed not to talk to their colleagues about the experiment until it had been completed. They were instructed not to compare their decisions with those of their colleagues because they did not all have the same sample of firms. They did have the same sample, this deception being necessary to prevent collusion among subjects between sessions. Participants were advised of the deception and its necessity in the debriefing session. Subjects did not retain any experimental materials outside the decision-making sessions. Finally, after all three sessions had been completed, participants attended a debriefing session in which the experiment, the results, and the correct decisions were explained.

4. RESULTS AND DISCUSSION

Propositions 1, 2, and 3 refer to first and second learning trials. Note that in the operationalization of variables described in the previous section, these became trials 2 and 3 respectively. Trial 1 was a preliminary trial used only in connection with the measurement of expertise. Trials 2 and 3 are used to measure learning and as the pretest and posttest respectively in the nonequivalent control group design.

To test propositions P2 and P3 the data were analysed using ANCOVA. The dependent variable was trial 3 decision accuracy (group performance for groups, and composite performance for individuals[8]). The covariate was trial 2 decision accuracy (composite performance for both groups and individual composites[9]). Covariate analysis was used because both P2 and P3 require the adjustment of trial 3 performance (the posttest) for any trial 2 performance (the pretest) differences by the three decision-making units. SPSS (Hull and Nie [1981]) ANCOVA was employed. The ANCOVA assumption of homogeneity of regression coefficients was tested and found to hold ($F = 1.20$; $df = 2/111$; $p = .305$).

Table 1 shows the ANCOVA results. The table reveals that there are no significant effects. The expertise theory (see P2) predicted that diverse groups would outperform homogeneous groups and individual composites. If this prediction were supported by the data, Table 1 would reveal a significant main effect for decision-making unit. However the table shows that this effect is not significant. Hence the results do not support the expertise theory.

The information load theory (see P3) predicted that the interacting groups would outperform individual composites in the high load condition, and that such an effect would either not occur at the low load level or would occur to a significantly reduced extent. If this prediction were supported by the data, Table 1 would show a significant interaction between information load and decision-making unit. Since this interaction is not significant the results do not support the information load theory.

To test proposition P1(a), the data were analysed using ANOVA. Again SPSS was employed. A mixed design was appropriate. Learning trials (trials 2 and 3) was a within subject factor, and information load and decision-making unit were between subjects factors. The dependent variable was decision-making performance i.e. accuracy - in trial 3, group performance for groups and composite performance for individual composites; in trial 2, composite performance for all three decision-making units. Table 2 shows the ANOVA table and Table 3 the means for significant effects. The equivalence theory predicts (see P1(a)) that learning will occur from trial 2 to 3 and that the learning will be the same for all three decision-making units. If this prediction is supported by the data, Table 2 will reveal (1) a significant main effect for learning trials, and (2) no significant learning trials by decision-making unit interaction. This has occurred. Table 3 shows that the means for the learning effect are as predicted. Decision accuracy has improved from trial 2 to 3. Hence the results support proposition 1(a).

This learning effect probably would have been greater from trial 1 to 2. This is because the learning curve, with number of trials on the X axis and performance on the Y axis, normally has a greater slope (indicating greater learning) at lower trial levels. The learning effect from trial 1 to 2 cannot be measured in this research because different samples of firms were used in the two trials and they may not be equivalent in terms of difficulty of predicting bankruptcy.

Proposition 1(b) was tested with the ANCOVA analysis in Table 1. The proposition predicts no significant differences in trial 3 performance of the three decision-making units after adjusting for trial 2 performance differences. Table 1 shows that this has occurred. Hence both propositions derived from the equivalence theory are supported.

Table 2 also shows an information load main effect.[10] The means in Table 3 indicate that subjects have lower decision accuracy in the high load condition. This effect has occurred in all decision-making units. This finding is in accord with past research. As noted earlier, psychologist Streufert [1972] has found that decision making performance decreases as load increases beyond ten cues. Other psychologists (e.g. Brehmer [1976], Ogilvie and Schmitt [1979]) have reported similar effects. Iselin [in press] has generalized this finding to the financial decision-making area. However, this load main effect is not of interest in our study of the load theory of composite versus group decision-making performance. That theory predicts a load by decision-making unit interaction which is not supported by the results here, as noted above.

Some may argue that the lack of support in this research for the expertise theory and/or the information load theory is due to insensitivity in the dependent variable. Decision accuracy was measured in trials 2 and 3 over six firms and hence on a seven point scale (0-6). Note that this scale has been sensitive enough to reveal, (1) a learning main effect, (2) an information load main effect, and (3) a random error reduction effect for composites v. individuals.[11] As noted above, all three findings are in accord with existing theory. In addition, note that Chalos [1985] used a similar nine point scale (eight firms) in his research and that seven point rating scales are widely used in behavioral science research.

5. CONCLUSION

The purpose of this paper was to investigate empirically three alternative theories which could explain the findings in the accounting literature regarding the relative performance of composites and interacting groups in financial decision making. The theories were the expertise theory, the information load theory, and the equivalence theory. The results do not support the expertise or load theories. The equivalence theory is supported.

Briefly the research has found no evidence to suggest that, in an interacting group, novice decision makers will follow an expert, resulting in the interacting group's decision performance being superior to that of a composite of individuals. There was also no evidence here that interacting groups could process high information loads better than composites. It has, however, been found that (1) individuals and interacting groups can learn over time (i.e. increased trials) such that their decision-making performance improves, (2) the improvement of both individuals and groups is about the same, and still exists in a comparison between a second and a third trial, and (3) when learning is controlled, there is no difference in the performance of composites and interacting groups. This result suggests that Trotman's [1985] finding that interacting groups outperformed composites, may be due to a learning confounding and not the operation of the expertise theory.

Although the expertise and information load theories have not operated in this research, there may be conditions in which they do. That is for future research to determine. It is important that such research avoid the learning and statistical regression confoundings that have occurred in prior studies.

The findings in this research may have implications for practice. In practice, financial decisions are often made by interacting groups (e.g. loan and audit review committees). The interaction process takes time which has an opportunity cost. These costs may return no compensating benefits because the groups may perform no better than composites which do not incur interaction costs. However, these are only tentative suggestions at this stage. Further research is necessary to investigate the findings in this research more comprehensively, as acknowledged in the previous paragraph.

TABLE 1

ANCOVA on Trial 3 Performance with the Covariate of Trial 2 Performance.

Source of Variation	S.S.	df	M.S.	F	p
Within cells	53.66	113	.47		
Decision-making unit (A)	.38	2	.19	.40	.67
Info. load (B)	.37	1	.37	.78	.38
A by B	.20	2	.10	.21	.81

TABLE 2

ANOVA on Decision-Making Performance

Source of Variation	S.S.	df	M.S.	F	p
Between Ss					
Within cells	131.24	114	1.15		
Decis.-making unit (A)	.31	2	.15	.13	.88
Info. load (B)	4.34	1	4.34	3.77	.05
A by B	2.30	2	1.15	.10	.37
Within Ss					
Within cells	31.01	114	.27		
Learning trials (C)	1.67	1	1.67	6.13	.01
A by C	.31	2	.15	.57	.57
B by C	.01	1	.01	.04	.84
A by B by C	.01	2	.00	.01	.99

TABLE 3

Means for Significant Effects

Effect	Level	Decision Accuracy
Learning trials	trial 2	3.81
	trial 3	3.97
Information load	low load	4.02
	high load	3.75

FOOTNOTES

1 For example, Schultz and Reckers [1981], Solomon [1982], Uecker [1982], Trotman, Yetton and Zimmer [1983], Trotman and Yetton [1985], Trotman [1985], Chalos [1985], Chalos and Pickard [1985], Libby, Trotman and Zimmer [1987].

2 In this case, it is necessary to have an odd number of individuals per composite to produce determinate outcomes.

3 Conditions of low standardized systematic bias. Standardized bias is defined as: $B = (x_t - \mu)/\sigma$

Where

B	=	standardized bias
x_t	=	the correct decision value
μ	=	the mean of the population of individual decisions
σ	=	the standard deviation of the population of individual decisions.

4 Composites reduced random error but did not improve mean accuracy. Given that the standardized bias levels in this research were 2.39 - 2.61 (see later), these findings are in accordance with Einhorn *et al.* [1977].

5 The paper will only concern itself with mean differences between composites and interacting groups. It will not study random error around the mean.

6 These conditions will be specified in the next section.

7 See footnote 3.

8 Recall that individuals were randomly assigned to composites of three.

9 Recall that all subjects made trial 2 decisions as individuals. A trial 2 composite score was obtained for individuals formed into groups and for individuals randomly assigned to composites.

10 Note that the load effect occurs here and not in Table 1 because the ANCOVA analysis in Table 1 removes trial 2 differences from the trial 3 data. Since the load effect in Table 2 was a main effect it occurred in both trials 2 and 3. Hence when the trial 2 effect is removed from trial 3 data no load effect remains.

11 See footnote 4.

REFERENCES

BREHMER, B. "Social Judgment Theory and the Analysis of Interpersonal Conflict." Psychological Bulletin (1976): 985-1003.

CAMPBELL, D.T., and J.C. STANLEY. Experimental and Quasi-Experimental Designs for Research. Boston: Houghton Mifflin, 1966.

CHALOS, P. "Financial Distress: A Comparative Study of Individual, Model, and Committee Assessments." Journal of Accounting Research (Autumn 1985): 527-543.

CHALOS, P., and S. PICKARD. "Information Choice and Cue Use: An Experiment in Group Information Processing." Journal of Applied Psychology (1985): 634-641.

COOK, T.D., and D.T. CAMPBELL. Quasi-Experimentation. Boston: Houghton Mifflin, 1979.

EINHORN, H.J., R.M. HOGARTH, and E. KLEMPNER. "Quality of Group Judgment." Psychological Bulletin (1977): 158-72.

FISCHER, G.W. "When Oracles Fail - Comparison of Four Procedures for Aggregating Subjective Probability Forecasts." Organizational Behavior and Human Performance (1981): 96-110.

HACKMAN, J.R., and C.G. MORRIS. "Group Tasks, Group Interaction Process and Group Performance Effectiveness: A Review and Partial Integration." In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol. 8). New York: Academic Press, 1975.

HUCK, S.W., W.H. CORMIER, and W.G. BOUNDS. Reading Statistics and Research. New York: Harper & Row, 1974.

HULL, C.H. and N.H. NIE. SPSS Update 7-9. New York: McGraw-Hill, 1981.

ISELIN, E.R. "The Effects of Information Load and Information Diversity on Decision Quality in a Structured Decision Task." Accounting, Organizations and Society (In press).

LIBBY, R., and R.K. BLASHFIELD. "Performance of a Composite as a Function of the Number of Judges." Organizational Behavior and Human Performance (1978): 121-129.

LIBBY, R., K.T. TROTMAN, and I. ZIMMER. "Member Variation, Recognition of Expertise, and Group Performance." Journal of Applied Psychology (1987): 1-5.

OGILVIE, J.R., and N. SCHMITT. "Situation Influences on Linear and Non-Linear Use of Information." Organizational Behavior and Human Performance (1979): 292-306.

REICHARDT, C.S. "The Statistical Analysis of Data from Nonequivalent Group Designs." In T.D. Cook, and D.T. Campbell (Eds.), Quasi-Experimentation. Boston: Houghton Mifflin, 1979.

ROHRBAUGH, J. "Improving the Quality of Group Judgment: Social Judgment Analysis and the Delphi Technique. Organizational Behavior and Human Performance (1979): 73-92.

SCHULTZ, F.J., and P.M.J. RECKERS. "The Impact of Group Processing on Selected Audit Disclosure Decisions." Journal of Accounting Research (Autumn 1981): 482-501.

SOLOMON, I. "Probability Assessment by Individual Auditors and Audit Teams: An Empirical Investigation." Journal of Accounting Research (Autumn 1982): 689-710.

STREUFERT, S.C. "Cognitive Complexity: A Review." Technical Report #2. Purdue University, 1972.

TROTMAN, K.T., "The Review Process and the Accuracy of Auditor Judgments." Journal of Accounting Research (Autumn 1985): 740-752.

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

TROTMAN, K.T., P.W. YETTON, and I.R. ZIMMER. "Individual and Group Judgments of Internal Control Systems." Journal of Accounting Research (Spring 1983): 286-292.

UECKER, W.C. "The Quality of Group Performance in Simplified Information Evaluation." Journal of Accounting Research (Autumn 1982): 388-402.

ZIMMER, I. "A Lens Study of the Prediction of Corporate Failure by Bank Loan Officers." Journal of Accounting Research (Autumn 1980): 629-636.