The Review Process and the Accuracy of Auditor Judgments
Author(s): Ken T. Trotman
Source: *Journal of Accounting Research*, Vol. 23, No. 2 (Autumn, 1985), pp. 740-752
Published by: Blackwell Publishing on behalf of The Institute of Professional Accounting, Graduate School of Business, University of Chicago
Stable URL: http://www.jstor.org/stable/2490836
Accessed: 14/10/2008 03:13

Journal of Accounting Research
Vol. 23 No. 2 Autumn 1985
Printed in U.S.A.

The Review Process and the Accuracy of Auditor Judgments

KEN T. TROTMAN*

1. *Introduction*

The review process is an integral part of the standard operating procedures of audit firms. While its importance has been noted (for example, see Mautz and Sharaf [1961] and AICPA [1981 sec., 320.02]) there has been little research or discussion on this review process (Bamber and Bylinski [1982]). One exception is Trotman and Yetton [1985] (hereafter TY) who examined the effect of the review process on the consensus of internal control evaluations. They found that although the level of consensus was significantly improved by the use of the review process, similar improvements could be obtained by the use of an interacting or composite group. TY suggested that the addition of a second opinion, regardless of its form, seemed to improve "performance."

Consistent with most previous studies on internal control judgments in which no criterion variable is available, TY used consensus as a surrogate for accuracy. However, as they noted, consensus is only a reasonable surrogate for accuracy if the expected value across all judges is an unbiased estimate of the correct judgment. Otherwise, group and/or review processes may simply suppress outliers, thereby inflating consensus without a corresponding increase in accuracy.

The main objective of the present study was to overcome this limitation by examining accuracy instead of consensus. This examination of the

* Senior Lecturer, University of South Wales. I wish to acknowledge the financial assistance of a Coopers & Lybrand research grant and the helpful comments of Robert Libby, Ron Weber, Philip Yetton, and Ian Zimmer. I also wish to thank Ron Weber for kindly lending me his case study material and simulation program. [Accepted for publication November 1984.]

effect of the review process on the accuracy of individual judgments is important given that a basic purpose of the review process is to reduce errors in judgment (Anderson [1977] and Arens and Loebbecke [1980]).

To determine accuracy, I used the case study developed by Weber [1978] which required auditors' judgments of the expected dollar error in an inventory system. These judgments were compared to the known error distribution generated by a system simulation model. The assessment of overall systems reliability in terms of dollar error is generally a critical audit decision and a necessary input to the auditor's materiality decision (see Weber [1978] and Mock and Turner [1981]).

The present study also extended the TY study in two other ways. First, it used a different operationalization of the review process. In TY, managers reviewed seniors' judgments in their absence. This excluded any interaction between the manager and the senior. In addition, managers were not aware of which seniors made each judgment, although they did know the senior's approximate status level within the organization. In this study, managers reviewed the judgments with the senior present. This allowed interaction between the managers and the seniors and (automatically) ensured that the managers knew which seniors made the judgment. This opportunity to interact provides a higher level of potential performance although it may result in certain process losses (Steiner [1972]). Also, since the managers knew which seniors made the judgment to be reviewed, they were in a better position to differentiate relative expertise between their judgments and those of the senior. Previous studies (for example, Einhorn, Hogarth, and Klempner [1977], Yetton and Bottger [1982], and Trotman, Yetton, and Zimmer [1983]) show that this ability to differentiate expertise can have an important influence on the group output.

Second, I examined judgments within a more complex task. The estimate of dollar errors in a system is less routine and involves more operations than the evaluation of the strength of an internal control system. Einhorn et al. [1977] suggested that task complexity may affect the level of systematic bias, which in turn affects the relative performance of alternative social decision schemes. For example, a greater number of operations in a task provides a greater opportunity for group members to show their expertise, particularly if these operations vary in difficulty.

Given these design changes, I found that while the review process significantly increased the accuracy of prereviewed judgments, there was no significant difference between the performance of the review process and an interacting group of two seniors. However, contrary to the TY consensus results, the composite group was significantly less accurate than the interacting group of seniors.

Section 2 develops the propositions to be tested and section 3 provides a description of the research methods. Section 4 contains the results and section 5 the summary and conclusions of the study.

2. Propositions

Einhorn et al. [1977] suggested that two sources of error in individual judgments in complex tasks are random error around the mean and systematic bias, where systematic bias is defined as the difference between the mean of a set of judgments and the "true value."

In this section I consider the possible reductions in these errors as the audit team changes from an individual to a composite group, then to an interacting group of seniors, and finally to a review process.

2.1 EFFECTIVENESS OF INTERACTING GROUPS

Random error can be reduced by the use of a composite group. For example, in situations where there is no systematic bias, Einhorn et al. [1977] state that the expectation of the means of groups of size N, drawn from the distribution of individual scores, will be less than the distribution of individual scores. Thus, a using composite group results in a tighter distribution around the mean value with a resulting increase in performance. However, they also show that this performance increase obtained by adding individuals to the composite group decreases as systematic bias increases, up to the point where at high levels of bias the composite group does no better than a randomly selected individual.

An alternative way to improve audit performance is to use an interacting group of seniors. In order for an interacting group to outperform a composite group, it is necessary for the former to tighten the distribution and/or reduce systematic bias further. For a situation of no bias, the only possibility is for the distribution of the judgments of the interacting group to be tighter than the distribution of unit weight composite judgments. Einhorn et al. [1977] illustrated that this is possible by showing that the composite performance is always below the performance of the best-member model.[1] If interaction between group members allows a group correctly to differentiate expertise, the interacting group performance can equal the best member.

As systematic bias in individual judgments increases, the gap between the performance of the composite group and the best-member model increases (Einhorn et al. [1977]) because a best-member strategy reduces systematic bias whereas the composite strategy does not. If an interacting group can correctly differentiate relative expertise, it allows the group to reduce the level of systematic bias by placing more weight on the judgments of the more expert individual. However, Einhorn et al. warn that if groups reverse the relative weights (due to nonvalid social cues that influence perceived expertise), an equal-weighted composite will outperform an interacting group. That is, it is better to weight group members equally than to assign high weights to those group members with comparatively poor judgments.

Thus the critical issue in the comparative performance of interacting

[1] The best-member model assumes that the group can recognize the best member with certainty.

and composite groups is the ability of group members to differentiate relative expertise. The limited previous evidence on this question has been mixed. Yetton and Bottger [1982] showed that groups can identify their best members with sufficient expertise to enable the performance of that best-member strategy to equal the performance of an interacting group. Miner [1984], using a similar task, found that while groups appeared to be able to identify their better members, they were little better than random chance at identifying the best member. On an audit task, Trotman et al. [1983] found that subjects could not differentiate relative expertise. However, those subjects were sampled from a restricted range of expertise since they were all enrolled in the same audit course. Finally, TY concluded that the inability of their interacting groups to show higher consensus than their composite groups was because the audit seniors could not differentiate relative expertise on their particular task.

In the present study, I hypothesized that interacting groups would outperform their equivalent composites for two reasons. First, relative expertise in this study was easier to differentiate than in previous group audit studies (for example, Trotman et al. [1983] and Trotman and Yetton [1985]) because (*a*) group members had worked together on previous audit engagements and therefore had prior knowledge of each others' relative expertise, (*b*) given the complex nature of the task, the range of expertise was likely to be higher than on frequently performed structured tasks, and (*c*) individuals had the opportunity to illustrate their calculations to the other group member, and to check each others' approach, whereas this opportunity did not exist in the more global judgments made in previous studies. Second, Einhorn et al. [1977] suggested that increased complexity increases the level of bias. This will decrease the relative performance of composite groups so that an interacting group should require less ability to differentiate relative expertise in order to outperform the equivalent composite. For example, Einhorn et al. varied the certainty with which the best member could be identified in considering their composite, best member, and proportional models.[2] When standardized bias[3] was zero, the composite model was substantially better than the proportional model, but inferior to the best-member model. When the standardized bias equaled 0.5 the composite only slightly outperformed the proportional model and was far inferior to the best-member model. As standardized bias increased to one, the composite model was outperformed by the proportional model.

---

[2] The proportional model considers how well the group would perform if it only had a certain probability of picking the best member.

[3] The level of standardized bias is calculated using the following formula:

$$\beta = (x_t - \bar{x})/\sigma$$

where β = standardized bias,
x_t = true value,
$\bar{x}$ = mean of the judgments made,
σ = standard deviation of the judgments made.

PROPOSITION 1. Interacting group judgments are more accurate than composite group judgments.

2.2 EFFECTIVENESS OF THE REVIEW PROCESS

The central issue in this study was the relative effectiveness of the review process in improving individual judgments, compared to the performance of an interacting group which consisted of a manager and a senior. It differed from an interacting group of two seniors primarily with regard to the rank structure of the group members. Given that the review process is a form of interacting group, the analysis in section 2.1 would suggest that the review process would decrease both the random error and systematic error. The extent of the reductions would depend on the ability of the managers to weight their judgments compared to the seniors.

PROPOSITION 2. Reviewed judgments are more accurate than individual judgments prior to review.

The relative performance of the review group compared to an interacting group of two seniors depends on three main factors: (*a*) the relative expertise of the senior to the manager on the task; (*b*) the ability of the manager to determine the seniors' relative expertise compared to that of the manager; and (*c*) the ability of the seniors to differentiate each others' relative expertise.

With respect to the first factor, Weber [1978] found that audit experience did not affect decision accuracy for the possible dollar error in the system. Consequently, on average, managers might not have had any additional expertise on this task as compared to seniors. If this was the case, the review process would not be expected to outperform interacting groups unless the relative expertise of group members could be more accurately determined by the review process than by interacting groups. With regard to the comparative ability of the two types of groups to differentiate relative expertise there was no previous empirical evidence. However, I expected that because of the different rank structures, there would be different processes used to allocate weights. For example, the process would be expected to be more of a joint decision in the interacting group, whereas in the review process the manager would be expected to allocate weights. These different processes may have had different respective effects on any reductions in random error or reductions in bias. However, since the relative strengths of these reductions are unknown, the null hypothesis was tested with respect to the comparative accuracy of an interacting group of seniors and the review process.

PROPOSITION 3. There is no difference in the accuracy of judgments made by an interacting group of two seniors and those of a review process.

3. *Methodology*

3.1 EXPERIMENTAL CASE

Subjects (consisting of practicing auditors—see below) were asked to assume the duties of another auditor who had just been transferred out

of their office to work on a special project. The subjects assumed duties just prior to the company's physical stock-take (inventory) which occurred on October 31. The financial year-end was December 31. The subjects were supplied with a description of the inventory accounting and control system plus a set of associated interim audit working papers prepared up to the date of the previous auditor's departure. The set of working papers, which were those used by Weber [1978], identified the strengths and weaknesses of the system. The weaknesses are shown in table 1.

A system simulation program modeled the inventory system's internal controls and accounting procedures and generated transaction data with predefined error frequencies to provide a data base for preparing the audit working papers. These working papers reported the results of an attribute sampling plan used to test the sampled transactions for errors. These sampling results provided a basis for inference about the error frequencies in the processed transactions (Weber [1978]).[4]

3.2 TASK

Subjects performed two tasks. First, they estimated the expected dollar error for total inventory as of September 30, and then they estimated the audit hours still required to perform the necessary audit activities up to and including year-end work.

3.3 SUBJECTS

Eight of the "Big Nine" Australian audit firms plus three medium-sized firms agreed to participate in the experiment. Five of the firms provided three teams, two firms provided two teams, and four of the firms each provided one team of auditors. A team consisted of three seniors, with at least three years experience, and one manager. This yielded a total of 91 subjects (one senior was ill on the day of the experiment). The mean experiences were 4.5 years (median = 4 years) for the seniors and 11.2 years (median = 10 years) for the managers.

Subjects were not selected randomly from within each firm. For five of the teams, a partner at the respective firms arranged the time of the experiment and selected the participants. For the other 18 teams, a partner selected which manager to contact and the managers in turn selected three seniors each and the date on which the experiment was to be carried out.

3.4 PROCEDURE

Four copies of the case study were given to each of the participating managers three days prior to the date of the experiment. They kept one copy and distributed the other three to the seniors. All subjects were asked to read the case study in order to become familiar with the

[4] The differences between the error rates in the statistical sample and the values used for error parameters in the simulation are shown in table 1.

TABLE 1
System Weaknesses

Ref. No.	Description of Weakness	Nature	Error Values Used in System Simulation	Errors Disclosed by Interim Audit: Mean Error Rate	Errors Disclosed by Interim Audit: Upper Limit of 95% Confidence
1	Error in receiving and inspection count operation	5% understatement	.10	.10	
2	Error in applying standard material price to shipment	Raw Mat. 1 priced as 3 Raw Mat. 2 priced as 4 Raw Mat. 3 priced as 1 Raw Mat. 4 priced as 2	.04	.08	.173
3	Error in reporting production count in Department I	5% overstatement	.08	.08	
4	Error in reporting production count in Department II	2% overstatement	.04	.04	
5	Error in applying standard direct labor hours and burden rates to time tickets	Std. for product 1 applied to 3 Std. for product 2 applied to 4 Std. for product 3 applied to 1 Std. for product 4 applied to 2	.03	.02	.093
6	Error in applying standard direct labor rates to time tickets	Rate for previous year applied 50% of time Rate for incorrect Department applied 50% of time	.04	.10	.198
7	Error in applying standard to raw materials requisitions processed	Std. for product 1 applied to 3 Std. for product 2 applied to 4 Std. for product 3 applied to 1 Std. for product 4 applied to 2	.04	.06	.148
8	Weigh-count correction	10 units overstatement			
9	Error in applying standards to production orders transferred to finished goods	Std. for product 1 applied to 3 Std. for product 2 applied to 4 Std. for product 3 applied to 1 Std. for product 4 applied to 2	.03	.02	.092

company's internal controls, the audit working papers, and, in particular, the weaknesses in the system described. However, they were not told what questions would be asked about the case during the experiment.

3.4.1. Individual and group judgments. I met with each audit team in a conference room at their place of employment.[5] At the start of the experiment all subjects were briefed on the nature of the experiment and the tasks. In particular, the seniors were informed that their work would be reviewed by their manager. After all of the seniors had completed the tasks, one of the three seniors was selected randomly for review and the other two seniors formed an interacting group. The two-person group was asked to come to a group decision on the tasks and then complete a clean answer booklet. Total time to complete the above procedures varied from 3¼ hours to 4½ hours.

3.4.2. Review process. At the start of the experiment, managers were told, in private, that they would be asked to review the work of one senior in detail, and that the other two seniors would concurrently complete a group judgment. Managers were asked to do sufficient work to put themselves in a position to carry out that review effectively. After both the senior and the manager had completed their initial tasks, the managers carried out the review and completed an answer booklet. These booklets were identical to those used for the individual judgments. This review process took approximately the same time as the interacting group procedures.

3.5 MEASUREMENT OF VARIABLES

Recall that by using Weber's [1978] simulation I could obtain an estimate of the distribution of the errors in the inventory system. This information was used to obtain a "true measure" of the errors which I then used to assess the accuracy of subjects' estimates of the errors. The measure of accuracy was the absolute difference between an auditor's estimate of the expected dollar error in the system and the mean of the distribution of simulated errors.

4. Results

Table 2 reports the level of accuracy for all individual seniors, for seniors who formed the groups, for seniors whose individual judgments were later reviewed, and for composites, interacting groups, and reviewers. For total inventories, interacting group judgments were more accurate than the judgments of their equivalent unit weight composite (Proposition 1: matched pairs t test; $t = 2.42$, $p \leq 0.05$).[6] The review process

---

[5] In most cases I singularly ran each team through the experiment. However, on two occasions, three and two teams respectively were processed at the same time. The rooms were large enough to divide the subjects up into separate groups after individual judgments had been completed.

[6] As all of these tests are not independent, the type I error may be inflated.

TABLE 2
"True" Errors–Expected Errors

	N	TOTAL INVENTORIES $\bar{x}$	TOTAL INVENTORIES s.d.	RAW MATERIALS $\bar{x}$	RAW MATERIALS s.d.
Individuals	68	54423	73684	26436	30313
Individuals who formed groups	44	48448	71637	22160	19970
Composite group	22	45535	42714	20118	13168
Interacting group	22	23196	19940	15197	10318
Individuals prior to review	23	67327	78784	35036	43598
Reviewed	23	31367	36151	16138	13915

significantly increased the accuracy of individual judgments (Proposition 2: matched pairs t test; $t = 2.19$, $p \leq 0.05$), but there was no significant difference between the accuracy of the review process and interacting groups (Proposition 3: matched pairs t test; $t = 0.93$, n.s.).[7]

Since the task was extensive and covered a considerable time period, fatigue could have occurred prior to the completion of the task. To check this, I tested the same propositions using only the estimates for raw materials. These estimates were typically completed first. The results for the accuracy of the raw material judgments were similar to the above with the exception that the difference between interacting and composite groups was only marginally significant ($t = 1.62$, $p = 0.06$).[8] This difference could have been due to different levels of systematic bias and/or the change in task complexity. Einhorn et al. [1977] suggested that the performance of a composite group decreases as systematic bias increases. In this study, the standardized level of systematic bias for the expected error in total inventories was 0.432, compared to 0.252 for the expected error in raw materials. The lower bias for estimates of raw materials may have resulted in the composite group performing comparatively better in the case of raw material estimates than for total inventory estimates. Alternatively, these results could have resulted from the increased task complexity which would have made differential expertise easier to identify.

Although there was no statistical difference between the accuracy of the reviewers and the interacting groups, the direction of the results favored the interacting groups. However, recall the method of obtaining the inputs for the review process and the interacting groups. From each set of four subjects (three seniors, one manager) one senior was randomly selected for review. The other two seniors formed an interacting group.

---

[7] The groups and reviewers were matched on the basis of their four-person team origin. Since the members of each team were from the same firm, this should have reduced the experimental errors caused by different firm training and procedures (see Nunnally [1975, pp. 211–12]).

[8] The propositions were also examined for the subsets of work-in-process and finished goods. All results were in the same direction, although in some cases the differences were not significant at the 5% level.

Despite this random allocation, the mean accuracy score for the seniors reviewed (average absolute error = 67,327) was lower than for the seniors who comprised the interacting groups (average absolute error = 48,448) (t = .98; n.s.). As a result the interacting groups had an advantage over the reviewers because they received better input data. Percentage improvements were therefore calculated. These show a 52% improvement in accuracy for interacting groups and a 53% increase for reviewers (t = .09; n.s.).

In the above analysis I considered mean absolute errors. Below I also examine mean errors which provide a measure of Einhorn et al.'s [1977] systematic bias. By examining the changes in mean errors and standard deviations of the responses, it is possible to divide the improved performance between reduction in random error and reduction in systematic bias as outlined in the proposition section.

Table 3 shows the level of systematic bias and the standard deviations for the judgments made by individuals who formed the groups, composite groups, interacting groups, individuals whose judgments were reviewed, and the review process. It shows that for all team formats the direction of the systematic bias was positive; that is, there was generally an overestimate of the error. In addition, the table shows a reduction in systematic bias for the review process and interacting groups as well as a reduction in random error for composite, interacting, and review groups. The review process was particularly effective in reducing systematic bias, showing the lowest level of systematic bias (11,629) and the highest percentage improvement (74.8%). On the other hand, interacting groups were the most effective in reducing random error with the lowest standard deviation (24,950) and the highest percentage improvement (63.5%). These contrary effects are consistent with the nonsignificant performance difference between interacting groups and review groups shown earlier (Proposition 3).

It could be suggested that the above reduction in bias was due to reviewers and interacting groups being less conservative than seniors. If this was the case, these process gains might be turned to process losses

TABLE 3

Systematic Biases and Standard Deviations

		Systematic Bias	Standard Deviations
Individuals who formed groups	44	37578	68376
Composite groups	22	37578	43157
Interacting groups	22	17983	24950
Individuals prior to review	23	46149	93888
Reviewed	23	11629	46842
% Improvement—individuals/composites		0	36.9%
% Improvement—individuals/int. groups		52.2%	63.5%
% Improvements—prior to review/reviewed		74.8%	50.1%

in situations of negative bias. To examine this generalizability issue, the ten individual judgments that underestimated the expected value figure were examined. For eight of these cases, the review process or interacting group reduced the bias, suggesting that the result was not merely caused by this conservatism effect.

Table 3 also explains two earlier results of this study. First, reviewed judgments decreased the systematic bias (74.8%) and standard deviations (50.1%) of individuals' judgments prior to review. This explains the significant performance difference in Proposition 2. Second, the significant difference in performance between interacting groups and composite groups (Proposition 1) can be explained by the interacting group reducing both systematic bias and random error by more than the equivalent composite.

Finally, I should point out that the differences in standard deviations shown in table 3 between interacting and composite groups ($t = 2.68$, $p \leq 0.05$) and interacting and review groups ($F = 3.52$, $p \leq 0.05$) are inconsistent with the TY results.[9] TY compared the standard deviations of composite, interacting, and review groups over their 15 cases and found no significant differences.

4.1 ADDITIONAL FINDINGS

I also examined the relationship between the evaluation of the internal control system and the estimates of hours for substantive testing provided by the various subjects. Correlation coefficients between the number of hours proposed for substantive testing and the absolute expected dollar error were calculated for the individuals, the interacting groups, and the reviewers. These were negative and insignificant (ranging from −0.02 to −0.18). As under- and overestimates may have different effects on sample sizes, these correlation coefficients were recalculated for only those cases where the auditors estimated that the expected dollar error in total inventories was an overstatement. Similar results were obtained. This finding of a lack of a significant relationship between the extent of the audit plan and the expected errors in the system is consistent with Weber [1978].

In addition, based on a finding by Weber [1978], I investigated the correlation between the level of audit experience and the number of audit hours proposed for substantive testing. Contrary to his results, my correlations were not significant ($r = 0.13$). However, the range of experience of my subjects was narrower and this could have deflated the correlation coefficient (Nunnally [1975]). As an alternative, I also tested the changes made by experienced managers to the judgments made by seniors. Here, audit managers significantly increased the seniors' estimated audit hours to complete the audit ($t = 2.06, p \leq 0.05$).

[9] The t test is used in the case of the correlated variances and the F test for the independent variances.

Finally, I assessed the effect of the review process and the use of interacting groups on the variability of estimates of audit hours required to complete the audit. Although interacting group judgments showed less variation than individual judgments ($F = 2.03$, $p \leq 0.05$), there was no significant difference between individual and reviewed judgments ($F = 1.14$, n.s.).

5. *Summary and Conclusions*

This paper reports the first empirical evidence on the effects of the review process on the accuracy of auditor judgments. Table 2 shows that judgments after review were significantly more accurate (53% on average) than those prior to review. The review process reduced both the level of systematic bias and the variance in the individual judgments. However, there was no significant difference between the accuracy of interacting groups of equal rank and the review process. While the review process was more effective in reducing systematic bias, the interacting group was more effective in reducing random error.

The increased accuracy from the review process is consistent with TY's consensus results. In addition, the reduced variation in individual judgments is consistent with TY's results which included an analysis of within-case variance. They also found that the variance of the judgments was reduced by the review process. However, inconsistent with the present study, they found no significant difference between the variance of the judgments of the review process and that of the interacting group of two seniors. The latter inconsistency may have resulted from the different operationalization of the review process in the two studies, or the differences in complexity between the tasks used in the two studies which affects ability to recognize expertise.

This study also showed that in reviewing seniors' proposed substantive test, managers significantly increased the number of audit hours proposed by the senior. This is consistent with Weber's [1978] findings that the extent of the proposed audit plan increased with increasing audit experience. However, although the review process increased the level of substantive testing, it did not decrease significantly the variability in sample sizes. This is consistent with Mock and Turner [1981]. Also consistent with Weber, I found that the extent of the audit plan proposed was insensitive to the size of the estimated dollar errors in inventory.

The final result relates to the comparative performance of interacting and composite groups. Interacting groups significantly outperformed their equivalent composite by reducing both the level of systematic bias and variance. Although the overall increase in accuracy was consistent with Solomon [1982], the reduction in variance was inconsistent with both TY and Trotman et al. [1983]. However, both of these latter studies used structured tasks involving the evaluation of the strengths of a number of internal control systems. Consequently, the opportunity to

demonstrate detailed calculations did not exist. On the other hand, the present study provided individuals with the opportunity to illustrate their calculations to the other group member, and to check each other's approach. This provided a greater opportunity to differentiate relative expertise.

Finally, a number of limitations of this study should be noted. First, auditors from most firms are not formally required to make the specific judgment required in this study. However, they are required to determine if weaknesses in internal control can lead to a material error in the accounts. This study, therefore, required auditors to provide a formal articulation of some of their thought processes when evaluating a system of internal control. Second, the use of only one case restricts the generalizability of the results. Social psychology literature (for example, Herold [1979]) has emphasized the importance of the task in group performance. Such factors as the complexity of the system and the overall strength of the internal control system might affect the relative performance of the alternative social decision schemes.

REFERENCES

AMERICAN INSTITUTE OF CERTIFIED PUBLIC ACCOUNTANTS. *Statement of Auditing Standards No. 1.* Chicago: Commerce Clearing House, 1981.

ANDERSON, R. J. *The External Audit 2: Organization and Procedures.* Toronto: Pitman, 1977.

ARENS, A. A., AND J. K. LOEBBECKE. *Auditing: An Integrated Approach.* 2d ed. Englewood Cliffs, N.J.: Prentice-Hall 1980.

BAMBER, E. M., AND J. H. BYLINSKI. "The Audit Team and the Audit Review Process: An Organizational Approach." *Journal of Accounting Literature* 1 (1982): 33–55.

EINHORN, H. J., R. M. HOGARTH, AND E. KLEMPNER. "Quality of Group Judgment." *Psychological Bulletin* (January 1977): 158–72.

HEROLD, D. M. "The Effectiveness of Work Groups." In *Organizational Behavior*, edited by S. Kerr. Columbus, Ohio: Grid Publishing Co., 1979.

MAUTZ, R. K., AND H. A. SHARAF. *The Philosophy of Auditing.* Sarasota, Fla.: American Accounting Assn., 1961.

MINER, F. C. "Group Versus Individual Decision Making: An Investigation of Performance Measures, Decision Strategies, and Process Losses/Gains." *Organizational Behavior and Human Performance* 33 (1984): 112–24.

MOCK, T. J., AND J. L. TURNER. *Internal Accounting Control Evaluation and Auditor Judgment.* Auditing Research Monograph no. 3. New York: AICPA, 1981.

NUNNALLY, J. C. *Introduction to Statistics for Psychology and Education.* New York: McGraw-Hill, 1975.

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

STEINER, I. D. *Group Process and Productivity.* New York: Academic Press, 1972.

TROTMAN, K. T., AND P. W. YETTON. "The Effect of the Review Process on Auditor Judgments." *Journal of Accounting Research* (Spring 1985): 256–67.

TROTMAN, K. T., P. W. YETTON, AND I. R. ZIMMER. "Individual and Group Judgments of Internal Control Systems." *Journal of Accounting Research* (Spring 1983): 286–92.

WEBER, R. "Author Decision Making on Overall System Reliability: Accuracy, Consensus, and the Usefulness of a Simulation Decision Aid." *Journal of Accounting Research* (Autumn 1978): 368–88.

YETTON, P. W., AND P. C. BOTTGER. "Individual Versus Group Problem Solving: An Experimental Test of a Best-Member Strategy." *Organizational Behavior and Human Performance* (June 1982): 307–21.