The Effect of Task Complexity and a Decision Aid on Decision Accuracy, Consensus, and Efficiency

Maureen Francis Mascha, Ph.D., CPA
Assistant Professor of Accounting
Elmhurst College
190 Prospect Avenue
Elmhurst, IL 60126-3296
(630) 617-3108
maureenm@elmhurst.edu

and

Cathleen L. Miller, Ph.D., CPA
Assistant Professor of Accounting
University of Michigan-Flint
303 E. Kearsley Street
Flint, MI 48502-1950
(810) 762-3278
catmillr@flint.umich.edu

September 2000

ABSTRACT

This study proposes that a mediating variable, task complexity, interacts with a decision aid to impact decision accuracy, consensus and efficiency. This study specifically examines the two components of task complexity proposed by Bonner's Model (1994), namely quantity and clarity, to determine if different combinations of these variables affect decision-making differently.

One hundred thirty-two senior accounting students participated in an experiment evaluating payroll internal controls. Task complexity is manipulated four ways and half of the subjects received a decision aid and half did not, resulting in eight treatment conditions.

The results indicate that when the task is "mixed" a decision aid significantly improves accuracy and consensus and marginally improves efficiency. Contrary to theoretical predictions, no significant improvement in accuracy, consensus, or efficiency was found with decision aid use when the task is complex, although for accuracy and efficiency the means were in the predicted direction. The findings also show that a decision aid *reduces* efficiency.

I. INTRODUCTION

Libby and Lewis (1982) state that one goal of applied decision research is to improve decision-making. Decision aids offer such a potential. Decision aids are tools that assist the user in performing a task (Eining and Dorr 1991) by offering a structured approach to problem solving (Bell and Wright 1995). Thus decision aids may reduce hours required to perform a task, resulting in more efficient use of staffing time in an era when competition and fee pressure are intense (Brown and Murphy 1990).

In an audit context, improved decision-making occurs when decision accuracy is improved and/or decision consensus is increased. Decision aid research on decision accuracy and decision consensus shows mixed results. Some studies show increased decision accuracy (e.g. Libby and Libby 1989; Butler 1985), while others show decreased decision accuracy (e.g. Jiambalvo and Waller 1984). Likewise, some studies show increased decision consensus (e.g. Libby and Libby 1989; Lewis, Shields, and Young 1983), while others show decreased decision consensus (e.g. Kachelmeier and Messier 1990; Ashton and Willingham 1988). This study proposes that a mediating variable, task complexity, interacts with a decision aid to impact decision accuracy and consensus and may shed light on the mixed results of previous studies.

Another important decision-making improvement measure is efficiency - the amount of time it takes to perform a task accurately. In the current audit environment of increasing fee competition and fee pressure to perform faster (Brown and Murphy 1990), public accounting firms are increasingly turning to automated decision aids to enhance both the accuracy and efficiency of their work. Unfortunately research on the effects of decision aids on efficiency is sparse and inaccurately measures efficiency as the time to complete a task rather than the joint effects of time and accuracy. This study examines the effects of decision aids and task complexity on efficiency in addition to the effects on decision accuracy and consensus.

The following sections review the prior literature to develop hypotheses, describe the experimental design, present results and discuss conclusions.

II. PRIOR LITERATURE AND HYPOTHESES

The prior accounting research reports contradictory results with regard to a decision aid on decision accuracy and consensus. One possible suggestion for the difference in results may be the difference in tasks used in the studies. In their meta-analysis of audit studies investigating internal control judgment, Trotman and Wood (1991) note studies examining the effect of a decision aid on accuracy and consistency employed different tasks with different degrees of complexity.[11] Thus, failure to hold task complexity constant may have contributed to the contradictory findings. Controlling for task complexity ex ante should also assist in reducing sampling error, another design problem noted by Trotman and Wood (1991).

Task Complexity

Most researchers who study task complexity agree that the more judgment required, the more complex the task (Abdolmohammadi and Wright 1987; Simon 1973). Bonner (1994), however, offers a model that recognizes task complexity as more than simply the amount of judgment involved. She posits that task complexity is a function of both the

---

[1]Task complexity plays a significant role in decision outcome as it affects degree of effort required (Asare and McDaniel 1996) and processing involved (Foos and Campbell 1992). Task complexity affects efficiency in that more complex tasks require additional time (Asare and McDaniel) and additional cognitive resources (Foos and Campbell).

clarity of the information presented to the decision maker and the amount of the information the decision maker must process in order to perform the task. Bonner's (1994) model allows manipulation of task complexity through manipulation of one or both of these components.

Bonner's model also helps to explain the contradictory results found in previous decision aid studies. For example, Butler (1985) found a decision aid increased consensus between subjects, while Kachelmeier and Messier (1990) found exactly the opposite. Analyzing the two different tasks using Bonner's model reveals that Butler's task was high in clarity and low in amount of information presented, i.e. low in complexity. Kachelmeier and Messier's task lacked clarity (subjects had to derive the necessary information themselves) and contained a large amount of information for the subjects to process, i.e. high in complexity. Thus, the conflicting results of decision aid effectiveness on auditor consensus may be a function of task complexity.

Additional evidence of the effect of task complexity on judgment is found in the expert systems literature, one type of decision aid (Eining and Dorr 1990; Murphy 1990). Hornick & Ruf (in print) noted that as their experimental task decreased in clarity (i.e. became more complex), subjects in the expert system treatment group exhibited greater decision accuracy than subjects in the non expert system group. Similar to the current study, the measured decision accuracy as the degree to which the subjects' evaluations matched those of the expert system. Thus, task complexity may affect decision accuracy in the presence of a decision aid.

Task complexity and decision accuracy

More complex tasks require greater cognitive effort (Asare and McDaniel 1996) and more cognitive processing (Foos and Campbell 1992) to result in an accurate decision. A decision aid provides additional information for improving decision accuracy by giving the decision maker the rules or steps to follow to complete the task accurately (Joyce 1976). The decision aid provides a resource to reduce the amount of cognitive effort and processing required for complex tasks, thus improving decision accuracy for complex tasks.

Hornick and Ruf (1997) find support that as tasks increase in complexity the use of a decision aid improves decision accuracy. Likewise Kachelmeier and Messier (1990) found that a decision aid increased decision accuracy for their complex task. Also, Caplan and Schooler (1992) found that a decision aid coupled with a simple task results in *decreased* accuracy. They posit that one reason for this result may be that simple tasks require few cognitive resources, and the decision aid simply overloaded or bored the subject. This discussion and research suggest that complex tasks performed with a decision aid will result in more decision accuracy than simple tasks performed with a decision aid. The following hypothesis is proposed:

> **H1: For a "complex" decision task, overall accuracy for subjects using a decision aid will be significantly *greater* than overall accuracy for subjects not using a decision aid; however, for a "simple" decision task, there will be no difference in overall accuracy for subjects using a decision aid and subjects not using a decision aid.**

Task complexity and decision consensus

In general, more complex audit tasks require more audit judgment, which can result in more decision variability because of different interpretations of information and the volume of information to be processed. One measure for decision variability is auditor consensus. Therefore, more complex tasks lead to more variability, which means less consensus.

Comparison of two early studies supports this prediction. Ashton's (1974) semi-complex task of internal control review of a payroll system was made more complex in Joyce's (1976) study by adding audit planning to the task. Ashton (1974) found a high degree of consensus, whereas Joyce (1976) found a low degree of consensus. Comparing the findings of these studies reveals that consensus varied significantly between the two studies where the tasks varied in complexity. Joyce's (1976) more complex task led to less consensus. However, these studies do not examine the effects of decision aids, complexity and consensus.

Kachelmeier and Messier (1990) examine the effects of a decision aid on consensus. They find that the decision aid resulted in less auditor consensus. As previously discussed, their task was complex (subjects had to select three judgment variables for use in the decision aid as compared to the subjects selecting one judgment outcome without the decision aid). Their finding suggests that combining a decision aid with complex tasks leads to less auditor consensus. Thus hypothesis two is stated as follows.

> **H2: For a "complex" decision task, consensus for subjects using a decision aid will be significantly *less* than consensus for subjects not using a decision aid; however, for a "simple" decision task, there will be no difference in consensus for subjects using a decision aid and subjects not using a decision aid.**

Levels of task complexity

Bonner's (1994) model posits that both clarity and the amount of information affect task complexity.[2] However, the model does not postulate whether clarity or amount of information dominates in determining task complexity level. Her model simply states that the overall complexity of the task is a function of how the pieces interact.

Separating the components of task complexity is important in designing a decision aid aimed at improving judgment. *Clarity* of the task differs from *amount* or *volume* of information provided to the subject in how they affect information processing and acquisition of knowledge. Clarity requires the subject to process each piece of information, retrieve from long-term memory the salient knowledge stored in long-term memory associated with the information being processed in working memory, and then attempt to integrate the new information with existing knowledge to reach the judgment required by the task (Anderson 1993; 1983; 1976). Thus, the *similarity* of each new piece of information to its predecessor (i.e. clarity) reduces the amount of retrieval from long-term memory and thus reduces the processing involved.

The *amount* or *volume* of information, however, not only affects the processing and retrieval from long-term memory, but also affects short-term or working memory. Iselin (1992) noted in his study of information overload that as the amount of information increased, so did working memory decay, fatigue, and accuracy. If information does not make it to working memory, it cannot be processed or added to long-term memory as new knowledge.

As hypotheses 1 and 2 predict, decision aid effectiveness on decision accuracy and consensus may be related to task complexity. Since tasks vary in degrees of complexity, knowing how decision aids interact with different levels of task complexity is

---

[2] The most complex task exists when clarity is low and amount of information is high; the least complex task exists when clarity is high and amount of information is low.

necessary to design effective decision aids. Therefore, this study examines how decision aids interact with clarity and amount of information separately in addition to their combinations.

***Clarity of information**.* Complex tasks are associated with low levels of clarity while simple tasks are associated with high levels of clarity. Based on the research discussed in regards to decision aid effects on decision accuracy and consensus for complex and simple tasks, the following two hypotheses are predicted.

> **H3a: For low clarity decision tasks, overall accuracy for subjects using a decision aid will be significantly *greater* than overall accuracy for subjects not using a decision aid; however, for high clarity decision tasks, there will be no difference in overall accuracy for subjects using a decision aid and subjects not using a decision aid.**

> **H3b: For low clarity decision tasks, consensus for subjects using a decision aid will be significantly *less* than consensus for subjects not using a decision aid; however, for high clarity decision tasks, there will be no difference in consensus for subjects using a decision aid and subjects not using a decision aid.**

***Amount of information**.* Complex tasks are associated with high amounts of information presented while simple tasks are associated with low amounts of information presented. Based on the research discussed in regards to decision aid effects on decision accuracy and consensus for complex and simple tasks, the following two hypotheses are predicted.

> **H4a: For a high amount of information decision tasks, overall accuracy for subjects using a decision aid will be significantly *greater* than overall accuracy for subjects not using a decision aid; however, for a low amount of information decision tasks, there will be no difference in overall accuracy for subjects using a decision aid and subjects not using a decision aid.**

> **H4b: For a high amount of information decision tasks, consensus for subjects using a decision aid will be significantly *less* than consensus for subjects not using a decision aid; however, for a low amount of information decision tasks, there will be no difference in consensus for subjects using a**

decision aid and subjects not using a decision aid.

Combinations of clarity and amount of information. Hypotheses 1 and 2 test the combinations of low clarity, high amounts of information ("complex" tasks) and high clarity, low amounts of information ("simple" tasks). The next four hypotheses relate to the combinations of low clarity, low amounts of information and high clarity, high amounts of information. Since these combinations combine elements of both "complex" and "simple" tasks and Bonner's model does not state whether clarity or amount of information plays the larger role, the hypotheses are stated in the null form.

> **H5a: There is no difference in overall accuracy for subjects using a decision aid and subjects not using a decision aid for a low clarity, low amount of information decision task versus "complex", "simple", or high clarity, high amount of information decision tasks.**

> **H5b: There is no difference in overall accuracy for subjects using a decision aid and subjects not using a decision aid for a high clarity, high amount of information decision task versus "complex", "simple", or low clarity, low amount of information decision tasks.**

> **H5c: There is no difference in consensus for subjects using a decision aid and subjects not using a decision aid for a low clarity, low amount of information decision task versus "complex", "simple", or high clarity, high amount of information decision tasks.**

> **H5d: There is no difference in consensus for subjects using a decision aid and subjects not using a decision aid for a high clarity, high amount of information decision task versus "complex", "simple", or low clarity, low amount of information decision tasks.**

Task complexity and decision efficiency

Most decision aid research examining efficiency defines efficiency as the total time to complete a task. However, simply spending the least amount of time on a task does not embrace the underlying concept that an efficiently performed task must be an *accurately* performed task. Using time to completion, decision aid research finds that decision aids often result in the subject taking *more* time to complete the task rather than less time (McDaniel 1990; Ashton and Willingham 1988). It is unclear, however, whether this additional time results in increased accuracy.

Eining and Dorr (1991) found subjects in the expert systems groups (a form of decision aid) significantly outperformed subjects in the control group even though subjects in the former group took longer to evaluate the cases. However, Kachelmeier and Messier (1990) found an increase in time to completion with a decrement in performance. Because there is no clear precedent for a decision aid's effect on efficiency, the following hypotheses are stated in the null form.

H6a: Subjects who perform the task in the presence of a decision aid will not significantly differ in their efficiency score than subjects who do not use the decision aid.

H6b: There is no difference in efficiency for subjects using a decision aid and subjects not using a decision aid for "complex" and "simple" decision tasks.

III. EXPERIMENTAL METHOD

The experimental design is a 2 x 2 x 2 between subjects laboratory experiment manipulating decision aid two ways (present/absent), clarity of the task two ways (high/low), and the amount of information two ways (high/low) for a total of eight cells. The clarity component is manipulated by the amount of similarity between internal control cues. High clarity is defined as all cues provide a similar signal regarding the state of the internal control environment, e.g. all strong cues or all weak cues. Low clarity is defined as a mix of cues such that no clear signal can be determined. The amount of information manipulation is determined by the number of cues provided to the subject. The high amount of information subjects receive 12 cues, whereas the low amount of information subjects receive 5 cues. The cues may relevant or irrelevant information, depending on the clarity manipulation for the case.

Experimental Task

The experimental task consists of the evaluation of internal controls in a payroll environment. Subjects begin by reading a brief case scenario describing the circumstances surrounding the payroll environment. Then they review four independent payroll case scenarios and provide internal control evaluations. The subjects also complete a short attitude toward evidence questionnaire before the fourth case[3], a demographics questionnaire at the start of the experiment, and a brief post-experimental questionnaire at the end of the experiment.

The experimental materials are adapted from Eining and Dorr's (1991) expert system study. Each case presents descriptions of internal payroll controls for five components - initial hiring and termination, recording of time worked, calculation and preparation of payroll, payment and distribution of wages, and other controls. The subjects provide quantitative evaluations (from 0 = absence of all controls to 100 = all controls are present and working as designed) and qualitative evaluations (high, medium, or low)[4] for each of the five component sections as well as an overall evaluation of the payroll controls.

The first case evaluation represents the subject's pre-experimental evaluations. The next two cases were practice cases. During this practice stage, one-half of the subjects received the decision aid with the two cases and were told to study the decision aid before

[3] The attitude questionnaire is completed before the fourth case evaluation to remove any recency effects from the two practice cases.
[4] The subjects are told that high represents a quantitative rating greater than 60, medium represents a quantitative rating between 40 and 60, and low represents a quantitative rating

completing the cases. The final case represents the subject's post-experimental evaluations and is completed without the use of the decision aid.

Decision Aid

The decision aid in this experiment is a manual aid. For each of the five component sections, the aid describes a good payroll internal control system and provides specific payroll procedures and controls for that component section. The authors created the decision aid based on work experience and textbook descriptions.

Subjects

This study uses senior level auditing students from three Midwest institutions as experimental subjects. The use of students is supported with three reasons. First, Eining and Dorr (1991) note that decision aids are most often employed for beginning auditors or novices. Accounting students represent a fair proxy for beginning auditors, especially senior or graduate level accounting students. Second, students do not present any danger of confound caused by prior exposure to decision aids as practitioners might (Tybout, Calder, and Philips 1989) and therefore are more homogenous subjects. Finally, internal control evaluation over a payroll system is found to be semi-complex for beginning auditors (Abdolmohammadi and Wright 1987). With some variation in the case materials, this task can be made “complex” for the students; however, this manipulation might not work for more experienced subjects.

Dependent Variables

Three dependent variables are used in this study. The first, decision accuracy, is defined as the absolute difference between the subject’s evaluation of internal control and the evaluation determined by an expert system. This definition is consistent with prior

---

less than 40. These rating definitions are similar to prior research in this area.

research examining decision accuracy for judgment tasks (Odom and Dorr 1995; Eining and Dorr 1991; Ashton and Brown 1980). The smaller the absolute difference, the more accurate is the decision.

The second dependent variable is consensus. Consensus is the extent to which subjects agree with each other (Bell and Wright 1995) and is measured by the correlation between subjects' evaluations (Joyce 1976). The higher the correlation between the subjects' evaluations, the more consensus among subjects exists.

The third dependent variable, efficiency, is defined as the ratio of decision accuracy to time of completion. Since decision accuracy is the absolute difference between the subject's evaluation and that of the expert system, larger ratios represent more efficient evaluations.

IV. ANALYSES AND RESULTS

Demographics

One hundred forty one students participated in the experiment. Nine subjects were excluded from the analyses due to missing data, resulting in a final sample size of 132 comprised of 61 males (47%) and 70 females (53%). A Chi-square test shows no significant differences between the number males and females by treatment group (χ^2 = 1.99, p = .96). Table 1 shows the number of subjects, average age, average GPA overall and major), average credit hours and average accounting experience by treatment group. Analysis of variance and Tukey t-tests show no significant differences between treatments on these demographic variables, suggesting successful randomization between treatment groups.

[Please insert Table 1 here]

Hypotheses testing - accuracy

Hypothesis 1 predicts that in the presence of a decision aid, overall accuracy will be greater for a complex decision task than for a simple decision task. Accuracy is first calculated as the difference between the subject's internal control evaluation and the internal control evaluation from an expert system for each of the five internal control evaluations for the pre- and post-cases. Then the average of the five individual internal control evaluations is calculated for the pre- and post-cases. Finally, overall accuracy for hypothesis testing is calculated as the difference between pre- and post-case accuracy means. Table 2 presents these pre- and post-case and overall accuracy means for each treatment group.

[Please insert Table 2 here]

Table 2 shows that using a decision aid for a simple task leads to lower accuracy between pre- and post-cases. It also shows that for the complex task, both groups - those with and without a decision aid - increased accuracy from pre- to post cases, with the no decision aid group experiencing more accuracy improvement.

Hypothesis 1 predicts the overall accuracy mean difference between the decision aid subjects and no decision aid subjects for group 3 (complex task) will be significantly improved over the overall accuracy mean difference between the decision aid subjects and no decision aid subjects for group 1 (simple task). To test this hypothesis, multivariate analysis of variance[5] contrasts were developed to compare the difference in overall accuracy means for groups 1 and 3. For both the simple and complex task groups the

[5] The student evaluations of the five internal control procedures are not independent of each other; therefore, multivariate analysis of variance is appropriate.

difference between the no decision aid group and the decision aid group means is not significant (Group 1: F = 1.16, p value = .33; Group 3: F = 0.55, p value = .74). Therefore, although the means are in the predicted direction, there is no statistical support for hypothesis 1.

Hypothesis 3a examines the effect of clarity of information on accuracy and predicts that using a decision aid will improve the accuracy of subjects receiving low clarity (more complexity) information more than the subjects receiving high clarity information. To test this hypothesis, multivariate analysis of variance contrasts were developed to compare groups 1 (low quantity, high clarity) and 4 (low quantity, low clarity) and groups 2 (high quantity, high clarity) and 3 (high quantity, low clarity). Both contrasts are not significant (Groups 1 & 4: F = 0.85, p value = .52; Groups 2 & 3: F = 0.98, p value = .43). Therefore, there is no support for hypothesis 3a.

Hypothesis 4a examines the effect of amount of information on accuracy and predicts that using a decision aid will improve the accuracy of subjects receiving high amounts (more complexity) of information more than the subjects receiving low amounts of information. To test this hypothesis, multivariate analysis of variance contrasts were developed to compare groups 1 (low quantity, high clarity) and 2 (high quantity, high clarity) and groups 3 (high quantity, low clarity) and 4 (low quantity, low clarity). Both contrasts are not significant (Groups 1 & 2: F = 0.90, p value = .48; Groups 3 & 4: F = 1.12, p value = .35). Therefore, there is no support for hypothesis 4a.

Hypotheses 5a and 5b examine the effect of combining clarity and amount of information on accuracy. These hypotheses are stated in the null form, such that is no significant difference is predicted between the mixed groups (Groups 2 and 4) and the simple and complex tasks (Groups 1 and 3). To test these hypotheses, multivariate analysis of variance contrasts were developed to compare group 4 (low quantity, low clarity) with the other three groups and group 2 (high quantity, high clarity) with the other three groups. Comparing group 4 with the other three groups results in a significant contrast (F = 2.46, p value = .04), rejecting the null hypothesis that there is no difference. The presence of a decision aid significantly improved the accuracy of this mixed group - low quantity, low clarity of information. Comparing group 2 with the other three groups does not result in a significant contrast (F = 1.82, p value = .11); therefore, the null hypothesis is not rejected. In the presence of a decision aid, there is no difference in accuracy of this mixed group - high quantity, high clarity - compared to the other groups.

Hypotheses testing - consensus

Hypothesis 2 predicts that in the presence of a decision aid, overall consensus will be lower for a complex decision task than for a simple decision task. Consensus is measured by comparing the dispersion of overall accuracy between the treatment groups. Table 3 presents the dispersion variables - standard deviation and variance - by treatment group.

[Please insert Table 3 here]

Table 3 shows that using a decision aid for a simple task leads to greater dispersion and, therefore, less consensus. However, it also shows that for the complex task, both groups - those with and without a decision aid - increased consensus, with the no decision aid group experiencing more consensus.

To test the hypotheses concerning consensus, Levene's (1960) test of differences in dispersion is used. This test examines the absolute deviations from the treatment mean. After transforming the accuracy variable by the treatment mean, a standard two-sample t-test is computed.

Hypothesis 2 predicts that in the presence of a decision aid, the dispersion of accuracy scores will be greater for the complex task than the simple task. A t-test between groups 1 and 3 show that the dispersion differences between no decision aid and decision aid subjects are not significantly different (Group 1: t = -0.30, p value = .76; Group 3: t = -1.18, p value = .25). Therefore, although the dispersion variables are in the predicted direction, there is no statistical support for hypothesis 2.

Hypothesis 3b examines the effect of clarity of information on consensus and predicts that using a decision aid reduces consensus of subjects receiving low clarity (more complexity) information more than the subjects receiving high clarity information. To test this hypothesis, analysis of variance contrasts were developed using the transformed accuracy variable to compare groups 1 (low quantity, high clarity) and 4 (low quantity, low clarity) and groups 2 (high quantity, high clarity) and 3 (high quantity, low clarity). Both contrasts are not significant (Groups 1 & 4: F = 0.11, p value = .74; Groups 2 & 3: F = 0.43, p value = .51). Therefore, there is no support for hypothesis 3b.

Hypothesis 4b examines the effect of amount of information on consensus and predicts that using a decision aid reduces consensus of subjects receiving high amounts (more complexity) of information more than the subjects receiving low amounts of information. To test this hypothesis, analysis of variance contrasts were developed using the transformed accuracy variable to compare groups 1 (low quantity, high clarity) and 2 (high quantity, high clarity) and groups 3 (high quantity, low clarity) and 4 (low quantity, low clarity). Both contrasts are not significant (Groups 1 & 2: F = 1.08, p value = .30; Groups 3

& 4: F = 0.00, p value = .96). Therefore, there is no support for hypothesis 4b.

Hypotheses 5c and 5d examine the effect of combining clarity and amount of information on accuracy. These hypotheses are stated in the null form, such that is no significant difference is predicted between the mixed groups (Groups 2 and 4) and the simple and complex tasks (Groups 1 and 3). To test these hypotheses, analysis of variance contrasts were developed using the transformed accuracy variable to compare group 4 (low quantity, low clarity) with the other three groups and group 2 (high quantity, high clarity) with the other three groups. Comparing group 4 with the other three groups results in a significant contrast (F = 6.35, p value = .01), rejecting the null hypothesis that there is no difference. The presence of a decision aid significantly increased the consensus of this mixed group - low quantity, low clarity of information. Comparing group 2 with the other three groups also results in a significant contrast (F = 7.40, p value = .01), rejecting the null hypothesis. The presence of a decision aid significantly increases consensus in this mixed group - high quantity, high clarity - compared to the other groups.

Hypotheses testing - efficiency

Efficiency is measured as accuracy (absolute difference between the subject's internal control evaluation and the expert system's internal control evaluation) divided by the reported time to complete the case evaluation (measured in minutes). An efficiency score close to zero represents the most efficiency. Table 4 presents the mean efficiency scores - pre- and post-cases and overall difference - by treatment group.

[Please insert Table 4 here]

Table 4 shows that all treatment groups decreased in efficiency from pre- to post-case evaluation. Table 2 shows that accuracy increased from pre- to post-case evaluation.

A review of the average times to complete the case evaluations shows that all treatment groups decreased the amount of time taken to complete the case evaluations from pre- to post-cases by approximately 3 minutes. Although subjects improved on accuracy and did so in less time, the ratio of accuracy improvement to time reduction results in *less* efficiency overall.

Hypothesis 6a examines whether the use of a decision aid will affect efficiency. Stated in the null form, the hypothesis states there will be no difference in efficiency between no decision aid and decision aid treatment groups. To test this hypothesis, multivariate analysis of variance contrasts were developed to compare the no decision aid treatment group with the decision aid treatment group within each of the four experimental groups. For all four groups the contrasts are not significant (Group 1: F = 0.96, p value = .44; Group 2: F = 0.81, p value = .55; Group 3: F = 0.23, p value = .95; Group 4: F = 0.77, p value = .57). Therefore, hypothesis 6a is not rejected; there is no difference in efficiency when a decision aid is used or not.

Hypothesis 6b examines the difference in efficiency between the simple and complex tasks in the presence of a decision aid. Again the hypothesis is stated in the null form; in the presence of a decision aid, there is no difference in efficiency between a simple and a complex task. To test this hypothesis, a multivariate analysis of variance contrast was developed to compare the difference in efficiency means for groups 1 (simple task) and 3 (complex task). The contrast is not significant (F = 0.78, p value = .57); therefore, there is no statistical support to reject hypothesis 6b. Although not statistically significant, the means in Table 4 for groups 1 and 3 show that a decision aid improves the efficiency for the complex task but decreases the efficiency for the simple task, similar to the accuracy means in Table 2.

Based on the results of the accuracy and consensus testing, additional post-hock contrasts were performed for effects of task complexity on efficiency. The results of these contrasts are consistent with the accuracy and consensus findings. Contrasts comparing groups 1 and 4 and groups 2 and 3 for the effects of clarity of information are not significant (Groups 1 and 4: $F = 1.34$; p value = .25; Groups 2 and 3: $F = 0.63$; p value = .68). Contrasts comparing groups 1 and 2 and groups 3 and 4 for the effects of amount of information are also not significant (Groups 1 and 2: $F = 0.94$, p value = .46; Groups 3 and 4: $F = 0.85$; p value = .52). Finally, contrasts comparing group 4 with the other three groups and group 2 with the other 3 groups for the combined effects of clarity and amount of information are statistically significant, consistent with the accuracy and consensus findings (Group 4: $F = 2.22$, p value = .06; Group 2: $F = 4.01$, p value = .00).

V. DISSCUSSION AND CONCLUSION

Previous research on the performance of decision aids on auditor decision-making accuracy and consensus found mixed results. This study expands previous research in three key ways. First, it examines a potential mediating variable, task complexity, as a possible explanation for the mixed results. Second, it specifically examines the two components of task complexity proposed by Bonner's Model (1994), namely quantity and clarity, to determine if different combinations of these variables affect decision-making differently. Finally, it investigates the effect of a decision aid on efficiency while controlling for task complexity.

Results indicate that when the task is "mixed" (i.e. quantity and clarity of information are in opposing directions such as low quantity, low clarity; high quantity, high clarity), a decision aid significantly improves accuracy and consensus and marginally improves efficiency. However, contrary to theoretical predictions, the results indicate no significant improvement in accuracy, consensus, or efficiency as a result of decision aid use when the task is complex, although for accuracy and efficiency the means were in the predicted direction. Interestingly, the findings indicate that when the task is complex, decision accuracy does significantly improve, for *both* the decision aid and non decision-aid groups.

Taken together these results suggest that task complexity may explain the previous mixed findings. Specifically, that task complexity is a function of both amount and content of information and that each contributes differently in affecting decision aid efficacy. These findings also note that a decision aid *reduces* efficiency. Taken together these findings suggest that task complexity needs to be considered before implementing a decision aid and that given the *loss* of efficiency, a decision aid should only be implemented when the complexity is “mixed”.

With any study, there are limitations. One is that accuracy, consensus, and efficiency were measured in a one-session setting, making it impossible to speculate on the longer-term effects of a decision aid. Future research needs to examine whether longer-term effects of a decision aid are different from the shorter-term, “immediate” effects. Second, students were used as subjects. While students are generally considered adequate surrogates for beginning auditors (Ashton and Brown 1980), results derived from using students may not be generalizable to auditors as a whole. Future research needs to examine whether these findings extend to practicing auditors, specifically whether experience interacts with task complexity in affecting decision aid efficacy. Finally, only one audit task was used. Future studies need to consider how task complexity and a decision aid affect decision-making in other contexts, such as risk assessment, audit planning, etc.

Table 1

Age, GPA, Credit Hours and Experience by Treatment Group

Group	No. of subjects	Avg. Age (yrs)	Average Overall GPA (4.00=A)	Average Major GPA (4.00=A)	Avg. Credit Hours	Average Accounting Experience (yrs)
1 - Simple[a] - No Decision Aid	16	24	3.33	3.35	110	0.53
Simple - Decision Aid	18	26	3.18	3.18	117	0.65
2 - Mixed[b] - No Decision Aid	14	26	3.29	3.36	129	0.43
Mixed - Decision Aid	18	25	3.22	3.26	102	0.59
3 - Complex[c] - No Decision Aid	15	24	3.28	3.18	127	0.47
Complex - Decision Aid	17	24	3.24	3.40	98	0.53
4 - Mixed[d] - No Decision Aid	17	27	3.25	3.24	115	0.41
Mixed - Decision Aid	17	29	3.26	3.06	124	0.59

[a]Group 1 Simple consists of low quantity and high clarity of information.
[b]Group 2 Mixed consists of high quantity and high clarity of information.
[c]Group 3 Complex consists of high quantity and low clarity of information.
[d]Group 4 Mixed consists of low quantity and low clarity of information.

Table 2

Mean Accuracy Scores by Treatment Group

Group	No Decision Aid				Decision Aid			
	N	Pre Case	Post Case	Over all	N	Pre Case	Post Case	Over all
1 - Simple[a]	16	30.17	20.00	10.17	18	25.42	25.35	0.07
2 - Mixed[b]	14	28.28	24.66	3.62	18	27.86	24.72	3.14
3 - Complex[c]	15	19.32	17.64	1.68	17	21.75	17.75	4.00
4 - Mixed[d]	17	18.31	18.36	-0.05	17	18.85	14.19	4.66

[a]Group 1 Simple consists of low quantity and high clarity of information.
[b]Group 2 Mixed consists of high quantity and high clarity of information.
[c]Group 3 Complex consists of high quantity and low clarity of information.
[d]Group 4 Mixed consists of low quantity and low clarity of information.

Table 3

Standard Deviations and Variances of Overall Accuracy by Treatment Group

Group	No Decision Aid			Decision Aid		
	N	Standard Deviation	Variance	N	Standard Deviation	Variance
1 - Simple[a]	16	16.09	259	18	16.53	273
2 - Mixed[b]	14	21.85	478	18	15.44	238
3 - Complex[c]	15	7.37	54	17	11.90	142
4 - Mixed[d]	17	10.70	114	17	7.72	60

[a]Group 1 Simple consists of low quantity and high clarity of information.
[b]Group 2 Mixed consists of high quantity and high clarity of information.
[c]Group 3 Complex consists of high quantity and low clarity of information.
[d]Group 4 Mixed consists of low quantity and low clarity of information.

Table 4

Mean Efficiency Scores by Treatment Group

Group	No Decision Aid				Decision Aid			
	N	Pre Case	Post Case	Over all	N	Pre Case	Post Case	Over all
1 - Simple[a]	16	5.76	8.60	-2.84	18	5.39	9.00	-3.61
2 - Mixed[b]	14	4.86	6.95	-2.09	18	4.28	7.82	-3.54
3 - Complex[c]	15	3.29	5.28	-1.99	17	3.74	4.67	-0.93
4 - Mixed[d]	17	2.99	5.27	-2.28	17	3.35	3.48	-0.13

[a]Group 1 Simple consists of low quantity and high clarity of information.
[b]Group 2 Mixed consists of high quantity and high clarity of information.
[c]Group 3 Complex consists of high quantity and low clarity of information.
[d]Group 4 Mixed consists of low quantity and low clarity of information.

REFERENCES

Abdolmohammadi and A. Wright. 1987. An Examination of the Effects of Experience and Task Complexity on Audit Judgments. *The Accounting Review*, 62:1-13.

Anderson, J.R. 1976. *Language, Memory, and Thought*. Hillsdale, N.J.: Erlbaum Associates.

____________1983. A Spreading Activation Theory of Memory. *Journal of Verbal Learning and Verbal Behavior*, 22: (261-295).

___________ 1982. Acquisition of Cognitive Skill. *Psychological Review*, 89(4): 369-406.

Asare, S. and L. McDaniel. 1996. The Effect of Familiarity with the Preparer and Task Complexity on the Effectiveness of the Audit Review Process. *The Accounting Review*, 1996, 71(2): 139-159.

Ashton, R. and S. Kramer. 1980. Students as Surrogates in Behavioral Accounting Research: Some Evidence. *Journal of Accounting Research*, (Spring): 1-15.

Baldwin-Morgan, A. 1995. Impacts of Accounting Expert Systems: A Review of Results and Methods. *Proceedings of the 1995 American Accounting Association Meeting, Orlando, Florida.*

Bonner, S. 1994. Model of the Effects of Task Complexity. *Accounting, Organizations and Society* 19 (April): 213-234.

Brown C. and D. Murphy. 1990. The use of Expert Systems in Public Accounting. Stillwater, OK: Oklahoma State University, unpublished working paper.

Butler, S. 1985. Application of a Decision Aid in the Judgmental Evaluation of Substantive Test of Details Sample. *Journal of accounting Research* 23 (2):513-526.

Calder, B., L. Phillips, and A. Tybout. 1981. Designing Research for Application. *Journal of Consumer Research* (September): 197-207.

Campbell, D. 1988. Task Complexity: A Review and Analysis. *Academy of Management Review* 13 (January): 40-52.

Caplan, L. and C. Schooler. 1990. Problem Solving by Reference to Rules or Previous Episodes: The Effects of Organized Training, Analogical Models, and Subsequent Complexity of Experience. *Memory & Cognition* 18(2): 215-227.

Cook, T.D. and D. T. Campbell 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.

Eining, M. and P. Dorr. 1991. The Impact of Expert System Usage on Experiential Learning in an Auditing Setting. *Journal of Information Systems* (Spring): 1-16.

Foos, P. 1992. Test Performance as a Function of Expected Form and Difficulty. *The Journal of Experimental Education* 60 (Spring): 205-211.

Hornick, S. and B. Ruf. (1997). Expert System Usage and Knowledge Acquisition: An Empirical Assessment of Analogical Reasoning in the Evaluation of Internal Controls. *Journal of Information Systems*.

Kachelmeier, S. J. and W. F. Messier, Jr. 1990. An Investigation of the Influence of a Nonstatistical Decision Aid on Auditor Sample Size Decisions. *The Accounting Review* 65 (1): 209-226.

Kerlinger, F. 1989. *Foundations of Behavioral Research*, 3rd. edition; Fort Worth, TX: Harcourt Brace Jovanovich.

Laughlin, P. R., and A. L. Ellis. 1986. Demonstrability and Social Combination Processes on Mathematical Intellective Tasks. *Journal of Experimental Social Psychology* 22: 177-189.

Levene, H. 1960. Robust tests for the equality of variance. In I. Olkin (Ed.), *Contributions to probability and statistics.* Palo Alto, CA: Stanford University Press.

Lewis, B., M. D. Shields, and S. M. Young. 1983. Evaluating Human Judgments and Decision Aids. *Journal of Accounting Research* 21 (1): 271-285.

Libby, R. 1981. *Judgment and Decision Making* New Jersey: Prentice Hall.

________ and P. Libby. 1989. Expert Measurement and Mechanical Combination in Control Reliance Decisions. *The Accounting Review* 64 (4): 729-747.

Messier, W. and J. Hansen. 1987. Expert Systems in Auditing: The State of the Art. *Auditing: A Journal of Practice & Theory* (Fall): 44-74.

Murphy, D.S. 1990. Expert System Use and the Development of Expertise in Auditing: A Preliminary Investigation. *Journal of Information Systems* (Fall): 18-35.

Odom, M. and P. Dorr. 1995. Epistemological Issues of Expert System Use: An Experimental Investigation of Knowledge Acquisition and Retention. *Journal of Information Systems (Spring).*

Simon, H.A. 1973. The Structure of Ill Structures Problems. *Artificial Intelligence* 181-201.

Steinbart, P. and W. Accola. 1996. The Relative Effectiveness of Alternative Explanation Formats and User Involvement on Knowledge Transfer from Expert Systems. *Journal of Information Systems (Spring)*.

Swieringa, R. J. and K. E. Weick 1982. An Assessment of Laboratory Experiments in Accounting. *Journal of Accounting Research* 20 (Supplement): 56-101.

Trottman, K. and R. Wood. 1991. A Meta-Analysis of Studies on Internal Control Judgments. *Journal of Accounting Research* 29 (1): 180-192.