Auditing:
A Journal of Practice
& Theory
Vol. 16, No. 2
Fall 1997

Reliance on Decision Aids: An Examination of Auditors' Assessment of Management Fraud

Martha M. Eining, Donald R. Jones and James K. Loebbecke

SUMMARY

The assessment of management fraud risk is a complex decision process, but one with which few auditors have had experience. As a result, decision aids have been suggested to support this process. Unfortunately, many times users do not rely on decision aids even when doing so would improve the quality of the decision. This paper reports on an experiment that examines the use of an expert system decision aid created to enhance the engagement of the user and increase reliance on the aid. Auditors using the expert system exhibited the ability to better discriminate between situations with varying levels of management fraud risk and selected more consistent subsequent decisions regarding appropriate audit actions than did users of a decision aid that provided only a suggested assessment, a logit statistical model. The logit model users did discriminate better than both checklist users and unaided decision makers.

Key Words: Management fraud, Decision aid, Expert system, Decision aid reliance, Red flags.

Data Availability: Contact the authors.

Statement on Auditing Standards (SAS) No. 82[1] (AICPA 1997) requires that the auditor assess the overall risk of material financial misstatement due to fraudulent financial reporting (management fraud) and that the auditor respond to this risk by designing an audit to provide reasonable assurance of detection. The presence or absence of red flags, over 30 of which are identified in SAS No. 82, can influence the auditor in assessing the likelihood of management fraud and have predictive value for assessing the likelihood of management fraud (Bell et al. 1993; Loebbecke et al. 1989). The risk assessment and audit planning response are important since too weak a response could jeopardize the effectiveness of the audit (i.e., could fail to uncover existing fraud) and too strong a response could result in an inefficient audit (McDaniel and Kinney 1995).

Prior research and documented audit failures indicate that auditors have difficulty assessing the likelihood of management fraud. SAS No. 82 gives little guidance on how the information in the red flags should be used to form a judgment of management fraud risk.[2] Auditors, most of whom have never encountered management fraud, rarely have the proper

[1] SAS No. 53 (AICPA 1988), which preceded SAS No. 82, had similar requirements.

[2] The AICPA plans to issue a publication, titled *Considering Fraud in a Financial Statement Audit: Practical Guidance for Applying SAS No. 82*, to supplement SAS No. 82.

Martha M. Eining and James K. Loebbecke are Professors at the University of Utah and Donald R. Jones is an Assistant Professor at Georgia State University.

We gratefully acknowledge the KPMG Peat Marwick Foundation Research Opportunities in Auditing for funding and other assistance. An earlier version of this paper was presented at the University of Southern California Audit Judgment Symposium, 1994. We would like to thank the participants of workshops at the University of Utah, the University of Texas at Arlington and Georgia State University for their comments and suggestions.

background that would lead to an ability to detect management fraud (Johnson et al. 1991; Loebbecke et al. 1989). In addition, multi-cue judgments of this type are inherently difficult for unaided human judges, even those considered to have a high level of expertise in the matter being judged (see Kleinmuntz (1990) for a review). Decision aids have been shown to outperform experts for multi-cue judgements (Kleinmuntz 1990; Libby and Libby 1989) so decision aids that make the firm's knowledge available to all auditors have the potential of improving auditor judgments of management fraud risk (Hackenbrack 1993).

The purpose of this study is to examine how three distinct types of decision aids influence auditors in assessing the risk of management fraud and in determining subsequent audit actions. The decision aids, which have been proposed for management fraud assessment, are: checklists (Pincus 1989), statistical models (e.g., a cascaded logit model) (Bell et al. 1993), and expert systems (Eining et al. 1990). All three of these decision aids use red flag cues and two of them, statistical models and expert systems, provide specific assessments of the risk of management fraud. While these assessments are likely to be superior to unaided human judgment based upon the same set of cues, recent studies suggest that decision makers are often hesitant to rely on decision aids, apparently placing more confidence in their own unaided judgment than in the aid (Ashton 1990, 1992; Arkes et al. 1986; Peterson and Pitz 1986). In those studies, the non-reliance on decision aids led to worse performance.

We conducted a laboratory study with auditors using their respective decision aid to assess the likelihood of management fraud for three situations with varying degrees of risk. Once each assessment was made, the auditors were asked to identify the appropriate additional audit steps. Our research focused on characteristics that would potentially increase reliance on the aid. These characteristics, referred to as constructive dialogue, were incorporated into the expert systems but not into the other two decision aids. All three decision aids were modeled after those used in practice, but presented in a common computerized form to enhance experimental control.

The results from our study provide strong support for the positive influence of the *constructive dialogue* in the expert system. The only difference between the expert system and the logit model was the inclusion of the *constructive dialogue* in the expert system, as both provided the same risk assessment for each case. Due to the influence of this *constructive dialogue*, the auditors who used the expert system were able to differentiate among three situations with varying risk of management fraud significantly better and agreed with the assessment from the system significantly more often than those using the logit model. Subjects receiving only a suggested assessment (logit model users) discriminated among the cases significantly better than those with no assessment (checklist and unaided control group). Furthermore, the use of the expert system resulted in subsequent audit planning decisions that were more consistent with risk assessments than did the use of the other decision aids.

This study contributes to our understanding of the use of auditing decision aids in two important ways. First, the study examines the use of a combination of characteristics derived from the psychology and decision-aid literatures designed to improve reliance on the decision aid. This study incorporated these characteristics into a realistic management fraud judgment-task setting, and found that when implemented into a computerized decision aid, they led to an increased reliance on the decision aid. Second, the study is one of the first to examine a subsequent decision after the use of a decision aid. Many auditing decisions are related sequentially, but typically not all are supported by a decision aid. Therefore, it is an important finding that the use of the expert system not only affected the original assessment, but also improved the subsequent decision related to audit planning.

The remainder of the paper proceeds as follows. The next section reviews the management fraud judgment process and relevant

psychology literature in terms of how the three types of decision aids for management fraud relate to reliance and bias. The section also develops a set of hypotheses used in the study. The third section discusses the methodology used to conduct our experiment, and fourth section presents the results from the study. Finally, the last section discusses our conclusions and the limitations of the study.

MANAGEMENT FRAUD AND DECISION AIDS

The Management Fraud Assessment

Auditors sometimes appear to have difficulty assessing management fraud risk. SAS No. 82 specifies more than 30 red flags that might affect the risk of financial misstatement but gives no specific guidance on how they might be combined into an overall assessment. Johnson et al. (1991) found that auditors who had never encountered management fraud were less able to detect it than auditors who had encountered it, despite industry-specific experience. In addition, few auditors have encountered management fraud. Loebbecke et al. (1989) found that less than 5 percent of the audit partners of one of the Big 6 firms had ever encountered material management fraud. Audit firms that failed to detect management fraud have experienced multimillion dollar lawsuits and negative press (see, for example, Palmrose 1987; and Arthur Andersen et al. 1992).

In an attempt to improve understanding of the relationship between red flags and risk, Loebbecke and Willingham (1988) conceptualized the management fraud risk judgment as having three components: *conditions* that were conducive to management fraud; *motivation* for management to commit management fraud; and *attitudes* of managers consistent with the commission of management fraud. Loebbecke et al. (1989) surveyed audit partners who had encountered management fraud and categorized red flags from SAS No. 53 and other literature according the Loebbecke/Willingham model. Bell et al. (1993) prepared a cascaded logit model using data from the Loebbecke et al. (1989) fraud cases and 305 non-fraud cases. These studies show that red flags can have significant predictive capability, raising the prospect that helpful decision aids can be developed.

Reluctance to Rely on Decision Aids

Despite all the effort that has gone into the development of decision aids, research indicates that decision makers are often reluctant to rely on them. Arkes et al. (1986) reported an experiment in which 11 treatment groups, including seven that were shown a decision rule, all performed substantially worse than if they had relied on the rule exclusively. A twelfth group, which had received a stern warning against non-reliance on the rule, matched the performance of the rule. In an experiment, Boatsman et al. (1995) found pervasive non-reliance by auditors on a decision aid that predicted the presence or absence of management fraud. In other similar studies where the decision makers were supplied with the output of a decision aid, they tended to do better than those without the aid, but still demonstrated substantial non-reliance, resulting in worse performance than relying on the aid would have given (Ashton 1990, 1992; Peterson and Pitz 1986; Whitecotton 1996a). Paradoxically, increasing performance pressure led to even less reliance on decision aids and worse performance (Ashton 1990).

Reliance on a decision aid is dependent upon the decision strategies adopted by the decision maker. In process tracing studies, Todd and Benbasat (1991, 1992, 1994) found that different decision aid features led to the use of different decision strategies. Arkes et al. (1986) found that the presence of outcome feedback in a probabilistic task led decision makers to adopt a strategy that relied on the decision rule until it failed. Ashton (1990) argued that increasing performance pressure could lead the decision maker to attempt to outperform the decision aid by developing and executing some strategies or heuristics for task performance instead of accepting the aid's recommendations. Ashton (1990, 153) points out, "For many cognitive tasks, however, several strategies will be available, the optimal strategy will not be readily identifiable, and little opportunity will exist to evaluate alternative strategies" (Earley et al. 1989; Earley et al. 1987).

Ashton (1990) suggested that individuals may not rely heavily on decision aids because they may not realize they need to do so. This lack of reliance may stem from the decision maker's opinion of her own ability or the ability of the decision aid. Many studies indicate that decision makers are generally overconfident concerning their own judgment (Lichtenstein et al. 1982; Mladenovic and Simnett 1994; Plous 1993). Arkes et al. (1986), Ashton (1990) and Whitecotton (1996b) found measurable overconfidence associated with nonreliance on a decision aid. The decision of whether (or how much) to use a decision aid may also be influenced by the level of confidence the decision maker places in the decision aid itself. Ashton (1990), for example, found that a decision aid with lower face validity is less relied upon. Davis and Kottermann (1995) found that describing a decision rule's benefits and providing cumulative feedback over a series of decisions led to greater reliance on the decision rule. However, case-by-case negative outcome feedback highlighted imperfect predictability of decision aids and was associated with lower reliance (Arkes et al. 1986; Ashton 1990). Thus, the tendency for the decision maker to discount the advice of the aid can be precipitated by overconfidence in the decision maker's own judgment, underconfidence in the advice of the aid, or both.

These observations suggest that the degree of influence of a decision aid may depend upon how the dialogue between the auditor and the decision aid affects the choice of decision strategies. If the decision maker enters the dialogue with an attitude of overconfidence in self and/or underconfidence in the aid, and the dialogue does nothing to combat this bias, then the decision maker is likely to select decision strategies that allow little influence from the aid. Alternatively, if the dialogue with the aid influences the decision maker toward selecting decision strategies that incorporate the aid's advice or processes to a greater degree, then the outcome will be decisions more in line with the advice of the decision aid.

The following section discusses the decision aids of interest to the current study— checklist, logit statistical model and expert system with *constructive dialogue*. The discussion identifies and discusses the effect the different aids may have on the overall decision process.

Decision Aids for Management Fraud

Checklist

Pincus (1989) has proposed management fraud checklists, consisting of lists of potential red flags, to serve as memory aids to ensure that auditors have identified all existing red flags on a particular audit. Checklists can successfully serve this function, although it is important that checklists are complete, since the items included on checklists may tend to reduce the ability to remember excluded items (Fischhoff et al. 1978). Pincus (1989), for example, found that unaided auditors outperformed checklist users, apparently because of their ability to think of some factors that were not on the checklist. After checklists have performed the function of helping identify red flags on an audit, they do not appear to offer additional benefit to their users. Checklists give no mechanical assistance for weighting and combining the red flag cues into an overall assessment.

Logit Statistical Model

Statistical models can be used to mechanically weigh and combine cues into a judgment. When used as decision aids, statistical models are generally more consistent and accurate than human judges (Einhorn 1972; Goldberg 1970; Kleinmuntz 1990). This is especially true when both the human judge and the decision aid have access to the same set of cues (Blattberg and Hoch 1990). Moreover, statistical models do not have to be perfect; bootstrapped models of an expert's decisions (Dawes 1971; Dawes and Corrigan 1974) and linear models with equally weighted factors typically outperform expert judges (Dawes 1979). Even statistical models with randomly weighted factors have outperformed human judgment (Dawes and Corrigan 1974).

Bell et al. (1993) developed a cascaded logit statistical model capable of converting identified red flag cues into an assessment of the likelihood of management fraud. The model

consists of two stages, with the first stage utilizing the red flag cues to reach assessments of the three components of the Loebbecke/Willingham model (conditions, motivation, attitude) and the second stage using these three assessments to arrive at an overall assessment of management fraud risk. Their model was derived using the 77 fraud cases from Loebbecke et al. (1989) and 305 non-fraud cases partitioned into an estimation sample and a holdout sample. The logit model has been implemented into a computerized decision aid that solicits the red flags in a manner very similar to that of a checklist and then reports its assessment to the user on a five-point scale from very low to very high. While the computerized decision aid generates the assessment based on the process described above, it provides the auditor with only an overall assessment and provides no information about the internal workings of the model.

The research on decision aid reliance indicates that the presence of a decision aid has a positive, if partial, influence on the user (Ashton 1990, 1992; Peterson and Pitz 1986; Whitecotton 1996a). Like the decision aids studied, the logit model is a statistically derived model of a probabilistic task. The logit model would likely have a high degree of face validity for auditors accustomed to using similar types of computerized decision aids. Note, however, that in prior research even the most favorably received decision aids gained only moderate reliance (Arkes et al. 1986; Ashton 1990, 1992; Peterson and Pitz 1986). Thus, if users tend to be overconfident in their own judgment (making them prone to discount or ignore the decision aid recommendation), the logit decision aid has no features that might avert them from selecting decision strategies of unknown quality, as described in Ashton (1990). Thus, we expect the logit model to be relied upon at a rate comparable to decision aids in the other studies; its recommendation will have some influence, but only to a moderate extent.

Expert System

Advances in technology currently provide the ability to develop decision aids, such as expert systems, capable of supporting the user's decision process well beyond merely suggesting an outcome. Expert systems differ from more traditional decision aids in two fundamental ways. First, they place emphasis on knowledge, typically generated as rules, rather than algorithmic solutions. Second, they provide access to this knowledge base to the user of the decision aid. In addition, sophisticated expert system software gives numerous capabilities for enhancing the dialogue between the user and the system. Using these advantages, expert systems have successfully assisted auditors in complex decision processes, resulting in more accurate and consistent decisions (Boritz 1985; Eining and Dorr 1991; Kelly et al. 1986; O'Leary and Watkins 1989).

Eining et al. (1990) developed a prototype expert system based on the Loebbecke/Willingham model. This expert system used its rules to combine red flag cues into an assessment of management fraud risk. The expert system was validated using the same data as the logit model (Bell et al. 1993), and was found to have a level of accuracy that is comparable to the logit model. The authors obtained the rule base built by Eining et al. (1990) to design and implement a *constructive dialogue*[3] for this rule base. The *constructive dialogue* was intended to exploit the findings of psychological and decision aid research to reduce overconfidence and improve the decision process. Exhibit 1 lists the steps followed in using the expert system and highlights features of the *constructive dialogue.*

The *constructive dialogue* has the following five features designed specifically to influence the auditor toward the use of the decision aid.

1) *Judgment Decomposition.* The *constructive dialogue* decomposes the risk judgment into judgments of the three components of the Loebbecke/Willingham model (conditions, motivations and attitudes) and then provides assistance in the combination of these assessments. Decomposition of global judgments into

[3] We use the phrase *constructive dialogue* as a global term to identify all the features of the expert system that were intended to increase reliance on the decision aid and/or improve the decision process.

EXHIBIT 1
Constructive Dialogue in Expert System

***Constructive Dialogue* Feature**

Judgment Decomposition	Decomposition of global judgment into three component subjudgments					
	Prior Judgment	System requires auditor to give unbiased judgment prior to viewing expert systems recommendations				
		Rule Presentation	System displays rules it used for this particular judgment			
			Reassessment Opportunity	System provides auditor with opportunity to change original assessment		
				Deviation Justification	System requires auditor to justify assessments that differ from the system's assessments	
					Steps in Use of Expert System	
					For each component (conditions, motivations, attitude):	
					1. Enter red flags that have been identified	
X	X				2. Enter auditor's assessment for individual component	
		X			3. View system's assessment for individual component View rules used by system to obtain that assessment	
			X		4. Enter auditor's final assessment for individual component	
				X	5. Enter auditor's justification if auditor's final assessment differed from system's assessment	
					After completion of all three components:	
					1a. View system's recommendation based on subject's assessment of each component	
					1b. View system's recommendation based on system's assessment of each component	
			X		2. Enter auditor's final overall assessment of risk of management fraud	

smaller subjudgments in decision aids was recommended by Beach et al. (1976). Einhorn (1972) and Libby and Libby (1989) found that experts were good at providing component judgments, but that a global combination of these judgments was better done mechanically.

2) *Prior Judgment.* The system asks the auditor to enter an assessment for each of the three components before revealing its own assessment of the component. This method captures a judgment that is not biased by the system's answer and reduces the likelihood of game playing with strategies to "outsmart the aid."

3) *Rule Presentation*. The system presents the rules used in determining its assessment to the user. Ye and Johnson (1995) found that the presence of a rule-based explanation facility led to a more favorable impression of a system; however, they did not test to see if this impression led to increased usage or reliance. If the auditor had not considered the particular relationships expressed in the rules, then the rule might enlighten or persuade the auditor of the correctness of the decision aid, thereby potentially increasing the auditor's confidence in the system. The rules could also influence the auditor's strategy construction process; having the auditor focus on specific rules could replace easy (but inaccurate) heuristics with more difficult (but more carefully considered) strategies.

4) *Reassessment Opportunity*. The system takes advantage of opportunities to give the auditor another chance to enter a decision more in agreement with the decision aid. Thus, if the dialogue enables the auditor to build a more carefully considered judgment, the system will prompt the auditor to enter that judgment.

5) *Deviation Justification*. The system requires the auditor to enter a justification for any deviation from the system's assessment. This selective justification requirement would avoid the negative aspect of a justification requirement found in Ashton (1990). Ashton (1990) surmised that subjects would have been reluctant to give "relied on the decision aid" as a justification. Our *constructive dialogue*, which requires justification only when the user deviates from the aid, removes the negative stigma of relying on the aid and adds a degree of effort for deviation.

All five of the factors included in the *constructive dialogue* are expected to influence the auditor toward relying on the aid. Thus, the auditors using the expert system would be expected to rely on their aid to a greater degree than the logit model users, even when the two decision aids provide the same suggested assessment for management fraud. The hypotheses stemming from the above discussion are presented below.

Hypotheses About Risk Assessment

In this section we express hypotheses concerning the influence of the three types of decision aids discussed above. To summarize, once the red flag cues have been correctly identified, (1) the checklist gives no help in combining the cues and therefore offers little that could be relied upon, (2) the logit model gives recommendations that are likely to discriminate better than most unaided auditors, but that may be relied upon only at a moderate level due to lack of interaction or understanding, and (3) the expert system gives equivalent recommendations, but engages the user in a *constructive dialogue* designed to increase reliance.

In management fraud risk assessment, it is difficult to verify the correctness of any particular judgment. However, two measures are available for ascertaining the quality of these judgments: (1) ability of a series of judgments to discriminate among red flag sets representing varying levels of risk, and (2) consensus among auditors. The first measure, ability to discriminate, is desirable since it does not depend upon the particular level of risk assessed (which is unknown), but merely acknowledges the ability to tell that some red flag sets are riskier than others. The second measure, consensus, provides an indication of the similarity of the judgments within one treatment. This measure has often been used in auditing research when no accuracy measure is available (Ashton 1985). The examination of consensus provides another measure of the influence of the decision aids on the auditor's assessment, allowing a comparison with the discrimination results.

The above discussions lead us to the following hypotheses.

H1A: Auditors using checklists will discriminate no differently than unaided auditors among red flag sets representing varying levels of management fraud risk.

H1B: Auditors using checklists will show no difference in consensus than unaided auditors in assessments of management fraud risk.

H2A: Auditors using the logit model will discriminate better than checklist users or unaided auditors among red flag sets representing varying levels of management fraud risk.

H2B: Auditors using the logit model will show a greater consensus than checklist users or unaided auditors in assessments of management fraud risk.

H3A: Auditors using the expert system will discriminate better than logit model users among red flag sets representing varying levels of management fraud risk.

H3B: Auditors using the expert system will show a greater consensus than logit model users in assessments of management fraud risk.

The expert system and logit model both provide a suggested assessment to the user. Based on the influences from the *constructive dialogue* of the expert system, we would expect to see the auditors agree with the recommendation from the expert system more often than they agree with the recommendation from the logit model. This leads to the following hypothesis.

H3C: The proportion of auditors agreeing with the final recommendation of the expert system will be greater than the proportion of auditors agreeing with the final recommendation of the logit model.

Hypothesis About Audit Planning

SAS No. 82 specifies that the auditor should respond to an assessment of higher engagement risk, such as high risk of management fraud, with changes in the planning and conduct of the audit. SAS No. 82 and other relevant standards give a variety of possible responses to high risk. These responses are summarized in exhibit 2 and classified according to their strength. While the exact relationships among red flags, assessment of management fraud, and audit design are not specified in the standards, decisions about them must be made and are important.

During a typical audit, the two decisions (assessment and response) are contiguous in time, implying that the knowledge and understanding residing in the auditor's working memory at the conclusion of the assessment of the likelihood of management fraud would still be present for the decision concerning the audit plan modification. Therefore, the auditor's mental representation of the second decision problem evolves from the first decision's problem representation. The auditor's problem representation for audit planning would likely contain the specific red flags identified, other information obtained through experience, and the outcome of the fraud assessment (Christ 1993). As

EXHIBIT 2
Potential Additional Actions Presented by Level of Strength

Ranking of Category Based on Strength of Response	Category of Response	Specific Response
1	No change	• None (Make no change in audit plan)
2	Increased vigilance	• Critically challenge accounting principles • Assign more experienced persons • Move audit work to year end • Require closer supervision of staff
3	Increased evidence collected	• Use tighter planning materiality amount (increase sample size) • Use more audit procedures
4	Exit engagement	• Consider withdrawing from engagement

part of the auditor's problem representation, the assessment of the likelihood of management fraud becomes a potential input to the decision of how to change the audit plan. A range of decision strategies is available at this point, centering in varying degrees around the fraud assessment.

Strategies for changing the audit plan that would most reflect the risk assessment would be selected when the decision maker has confidence in the assessment made or when the decision maker has a firmer comprehension of the relationship between the red flags identified and the risk assessment. The use of an expert system with a *constructive dialogue* should increase the confidence in the assessment made as well as the comprehension of the red flags identified. We would, therefore, expect to see greater correspondence between the assessment and the level of change in the audit plan for those auditors using the expert system. This leads to the following hypothesis:

H4: The degree of correspondence between the assessment of the management fraud risk and the level of change in the audit plan will be higher for expert system users than for other groups.

METHOD

A laboratory experiment was conducted to examine the influence of the use of different decision aids on the assessment and evaluation of management fraud. Auditors from one of the Big 6 accounting firms used their respective decision aid to assess the likelihood of management fraud for three separate cases representing situations with three different levels of risk of management fraud.

The experiment consisted of a between-subjects decision aid treatment and within-subjects repeated measures related to the assessment of the three cases. The decision aid treatment consisted of four levels: computerized checklist, computerized logit model, computerized expert system and an unaided control group whose members entered their responses on the computer. Subjects were randomly assigned to treatment groups. The order of the cases was counterbalanced across treatment groups.

Subjects and Procedures

Ninety-six auditors attending national manager auditing training sessions served as the subjects for the study. Three subjects had problems either with diskettes or directions, resulting in 93 usable responses. The auditors had an average of 3.2 years of audit experience with an average of 23 audit engagements. The mean subject age was 26 years and 44 percent were female. Discussions with firm representatives and reference to firm materials indicated that this group had sufficient background and experience to make the management fraud assessment. In addition, the instrument was pretested on a group of auditors from the same firm and at a comparable level. They were able to make the necessary assessments.

All subjects were presented with their respective decision aid on a Macintosh computer, similar to the ones used in their audit practice. The instructions for the experiment, booklets for the three cases, and a post-experiment questionnaire were in an envelope in front of the computer when the subjects arrived. Subjects were guided by both written instructions and the computer to complete each case before proceeding to the next one.

The instructions for the experiment provided information about the assigned decision aid, the experimental procedures, and the description of the scale for the assessment of the likelihood of management fraud. Each case consisted of a memo from the engagement's audit senior to the audit manager indicating that preliminary planning had been completed and that the red flags listed had been encountered. The subject, assuming the role of the audit manager, was asked to use his or her respective decision aid to assess the likelihood of management fraud being present in the company. After entering the assessment of each case on the computer, the subject answered questions in the case booklet regarding other aspects of the overall

evaluation for that case. Upon completion of all three cases, the subject filled out the post-experiment booklet, which asked for demographics and other information.

Development of Cases

The cases were developed to simulate realistic audit situations with low, medium and high potential risk of management fraud. A review of firm manuals and discussions with practicing auditors indicated that it is typical for the auditor to make an assessment of high, medium or low for the risk of management fraud. One of the authors, with considerable experience in both auditing and management fraud situations, selected the preliminary list of red flags for each category from real life situations found in Loebbecke et al. (1989). The red flags for each case were entered into both the actual expert system and the logit model to ensure that the same outcome was generated by each system for each case. Both decision aids initially discriminated among the cases, but slight modifications were made to the initial list of red flags for one case to achieve uniformity in the final assessment. Exhibit 3 provides a list of the red flags for each case.

Computerized Decision Aids

The computerized logit model was developed by the same accounting firm that provided the auditors for the study and contained the red flags from the Bell et al. (1993) study. The working expert system was developed using Level 5, an expert system shell, from the

EXHIBIT 3
Red Flags in Cases

Case Level	Red Flags Present
Low Risk	• Most of the members of the management team are new and therefore inexperienced. • The management team appears to have a propensity to take undue risks. • Accounting personnel exhibit inexperience and a degree of laxity in performing their duties.
Medium Risk	• There has been high management turnover. The new CEO has replaced the top three members of his team this year. • The company's industry is in a downturn because of difficult economic conditions. Increased governmental environmental regulations also seem to be a problem. • The company is in relatively weak financial condition, with the possibility that its net working capital position could become negative in the near term. • This is a new client. • A significant law suit has been filed against the company concerning the acts of its prior management by a group of minority shareholders.
High Risk	• Management operating and financial decisions are dominated by the chief executive officer. • The chief executive officer places undue emphasis on meeting earnings projections. • The chief executive officer and the chief financial officer both have an aggressive attitude toward financial reporting. • The chief executive officer exhibits strong personality anomalies. • The chief financial officer can be highly unreasonable. • The company is in a period of rapid growth. • The company's profitability is inconsistent with most of the industry. • A substantial portion of management compensation depends on meeting quantified targets. • The company has a weak control environment. • The company has entered into transactions with related parties. • The company has a planned public offering.

prototype developed by Eining et al. (1990) using the red flags from the Bell et al. (1993) study.[4] Both the expert system and the logit model were validated using data from the Loebbecke et al. (1991) and Bell et al. (1993) studies.

For strong experimental control, all four groups (checklist, logit model, expert system and control group) viewed a common computer interface developed using the software HyperCard, with which the subjects were familiar from their audit work. This common interface eliminated potential time differences due to differences in software and enabled us to control the order of the screens and maintain uniformity except for the aspects varied in the study. The HyperCard versions of the decision aids were developed specifically for the cases being analyzed and were not complete versions of the decision aids. Pilot tests indicated that the subjects did view these as complete operational decision aids, which was the intent.

Both the computerized materials and the case booklets were subjected to rigorous pretesting prior to conducting the experiment. The pretesting was initially done using colleagues, then using Ph.D. students with a background in auditing, and finally using auditors at the same level and from the same accounting firm where the study was conducted. Appropriate adjustments were made after each round of testing.

Dependent Variables

Dependent variables were measured for each of the three cases presented. Exhibit 4 describes the measurement of the assessment of management fraud. The actual assessment of the likelihood of material management fraud and the time taken to reach that assessment were collected through the computerized decision aid. After completing the computerized decision aid for each case, each subject was asked to select all of the audit actions that they believed to be necessary from a list of potential future audit actions. The sequence of each list was randomly generated. The levels of strength that we assigned to these actions (presented in exhibit 2) were not revealed to the subjects. While the specific audit actions that should be recommended are subject to interpretation, it is clear that the strength level of the recommended actions should correspond to the assessed risk of each situation.

RESULTS

Assessment of Risk

Discrimination Among Risk Levels

The results of the auditors' assessments of risk for the three cases are summarized in table 1. These data were analyzed to examine the extent to which auditors in each of the experimental groups were able to discriminate among cases with three different levels of risk.

Table 2 shows the breakdown of subjects in each group classified by degree of discrimination. For a subject to attain the highest degree of discrimination the subject should have rated the high-risk case as riskier than the medium-risk case and the medium-risk case as riskier than the low-risk case. This and other degrees of discrimination are defined in table 2, panel A. They are listed in ranked order from best discrimination (assessments sequenced from high to low risk) to worst discrimination (assessments in reverse sequence).

The levels of discrimination of different decision aid treatments were compared using the Mann-Whitney U test. The results of these tests are shown in table 2, panel B. Hypothesis 1A, that checklist users would perform no differently than the unaided group, was supported. Hypothesis 2A, that logit model users would discriminate better than either the unaided group or checklist users, was supported. Hypothesis 3A, that expert system users would discriminate better than the logit model users, was supported as well.

Consensus

We measured consensus among users of a decision aid by computing the "pairwise

[4] An expanded expert system was also created and tested on an additional 24 subjects. This system contained all the red flags from the Loebbecke et al. (1989) study and was developed to test that full model. The model developed with the reduced set of red flags to match those in the logit model performed as well as the expanded system. Therefore, we present our discussion and analyses based on only the reduced model (i.e., the one containing the same red flags as the logit model) for ease of exposition.

EXHIBIT 4
Scale for Assessing the Likelihood of Material Management Fraud

The matter that you are considering condition, motivation, attitude, or overall assessment would:

Statement	...be a rare event in the circumstances described. You would be extremely surprised at its occurrence.	...be unusual, but not rare, in the circumstances described. You would not expect it, but would not be overly surprised at its occurrence.	...be fairly common in the circumstances described. You would not be particularly surprised by its presence or absence.	...occur most of the time in the circumstances described. However, you do not consider it a "near certainty, and understand that it would not occur in all such circumstances all the time.	...almost always occur in the circumstances described. You would be extremely surprised at its absence in these circumstances.
Level	Very Low (1)	Low (2)	Moderate (3)	High (4)	Very High (5)

TABLE 1
Assessment of Risk in Three Cases:
Frequencies of Subjects Choosing Each Risk Level by Decision Aid Type

Subject's Risk Assessment	Unaided	Checklist	Logit Model	Expert System
Panel A: Low Risk Case				
Very Low	3	4	2	
Low	9	14	15	19
Moderate	7	4	6	4
High	4	2		
Very High				
Panel B: Medium Risk Case				
Very Low		1		
Low	1	11	4	
Moderate	11	7	16	14
High	9	4	3	8
Very High	2	1		1
Panel C: High Risk Case				
Very Low	1			
Low	1			
Moderate	3	8	1	
High	6	7	11	
Very High	12	9	11	23

correlational consensus" as given in Ashton (1985). First, for each pair of subjects within a decision aid group, the correlation of their three assessments was calculated. Then each of these correlations was transformed using the Fisher r-to-z transformation. The z-scores for each subject were then averaged, providing a single score that indicated how closely that subject agreed with the rest of the group. One-way ANOVA[5] indicated that the decision aid treatment was significant ($p = .0001$). The results of the paired comparison tests are presented in Table 3, with the untransformed means presented. Hypothesis 1B predicted that unaided auditors and checklist users would not differ on consensus. The results of the analysis showed no significant difference between those two groups, thus supporting the hypothesis. Hypothesis 2B predicted that the decision makers in the logit group would exhibit higher consensus than either the checklist users or unaided decision makers. This hypothesis was also supported. Finally, hypothesis 3B predicted that expert systems users would exhibit higher consensus than the logit model users. This hypothesis was supported as well.

Agreement with Decision Aid

The frequency data from table 1 were used to test H3C, which predicted that the proportion of subjects agreeing with their respective decision aid would be greater for the expert system users than for the logit model. The results indicate that a significantly higher proportion of expert system users agreed with the system's assessment than did logit users (Chi-squared = 6.9, 1 DF, $p < .01$), thus providing support for H3C.

Evidence of Auditor Engagement

The results of our hypotheses tests support our contention that the users relied on the expert system to a greater extent than on the logit model because of increased engagement. To

[5] All statistical analyses were performed with appropriate adjustments for unequal sample sizes.

TABLE 2
Degree of Discrimination Among Risk Levels by Decision Aid

Panel A: Frequencies of Subjects by Decision Aid

Degree of Discrimination Among Three Cases (listed in order from best to worst discrimination)		Frequencies of Subjects			
Category	**Definition (">" means "rated riskier than")**	**Unaided**	**Checklist**	**Logit Model**	**Expert System**
Full	Low < Medium < High	10	5	15	22
Partial	Low = Medium < High or Low < Medium = High	7	13	7	1
None	All ranked equally				
1 pair reversed	Low > Medium or Medium > High	6	6	1	
All reversed	Low > Medium > High				

Panel B: Summary of Tests Comparing Decision Aids' Discrimination

Hypothesis	Hypothesized Comparison (">" means "discriminates better than")	Mann-Whitney U	p-Value
H1A	Checklist = Unaided	231.5	.3111*
H2A	Logit > Unaided	189.5	.0334**
	Logit > Checklist	139.0	.0007**
H3A	Expert System > Logit	183.5	.0049**

* Two-tailed test.
** One-tailed test.

TABLE 3
Consensus Measures by Decision Aid

Decision Aid	Consensus Mean Pairwise Correlations*
Unaided	.622a **
Checklist	.656a
Logit Model	.889b
Expert System	.960c

* The analysis was completed on z-scores obtained in the Fisher r-to-z transformation as described in Ashton (1985). The numbers presented in the table are the means of the z-scores transformed to correlations.
** Multiple comparison tests using the Duncan multiple comparison of means method indicated significant differences at the .05 level. Measures that have the same letter do not differ significantly.

provide further support for the concept of engagement, we performed additional analyses on specific components of the expert system. We recorded the time that each subject took to use their respective decision aid and make an assessment. A one-way ANOVA of decision aid treatment on time was significant at $p < .0001$. The group using the expert system (8.7 minutes per assessment) took significantly longer than those using the logit model (5.2 minutes) and the checklist (4.0 minutes). The unaided group (1.3 minutes) took significantly less time than any of the other groups. The mean times are consistent with the demands of the different interfaces.

While the desire is for auditors to use the decision aid to improve their decision making, it is not desirable to have auditors rely totally on the aid and make their decisions by rote. To explore this, we examined whether the subjects using the expert system ever differed from the aid and, if so, whether they provided a justification, and whether they ever adjusted their decisions. Of the original 207 subjudgments[6] required of the expert system users for the three cases, 114 were initially different from the system recommendation. After auditors reviewed the rules and the recommendation from the system, 37 subjudgments from 15 subjects still differed from the system. Of these, all but one subject provided justification for this difference. This provides support for the contention that the auditors were engaged and did not merely accept the output from the system without question.

Correspondence Between Risk Assessments and Audit Planning Response

In this section we examine the extent to which the audit planning responses correspond to the risk assessments, the concern of H4. These two decisions (over the three cases) corresponded when stronger audit planning responses[7] were associated with higher assessments of audit risk, and equal audit planning responses were associated with equal assessments of risk. Table 4 (panel A) shows a summary of these data. When these data were analyzed, we found that a higher proportion of expert system users had correspondence than (1) unaided auditors (Chi-square = 7.26, 1 DF, $p < .01$), (2) checklist users (Chi-square = 4.78, 1 DF, $p < .05$), and (3) logit system users (Chi-square = 5.66, 1 DF, $p < .01$). These findings support H4. Even though logit users received a suggested recommendation of audit risk, that recommendation led to no better correspondence between assessments and the audit planning response. The constructive dialogue in the expert system, however, did.

It is possible that subjects' audit plan responses were more sensitive to the risk levels of the cases than to the subject's own assessments of these risks. To examine this issue, we checked the correspondence between the audit planning responses and the risk levels of the cases. These are summarized in table 4 (panel B). Compared with the correspondences for risk assessment (table 4, panel A), these correspondences were worse for the unaided and checklist groups, not significantly better for the logit group (Chi-square = .89, 1 DF, $p > .30$), and the same for the expert system group. Audit planning was no more sensitive to the red flags in the cases than to the auditor's own assessment of risk.

SUMMARY AND CONCLUSIONS

The current study examined the influence of the use of decision aids on the auditor's assessment of the risk of management fraud and the selection of appropriate audit actions. We specifically developed a decision aid that incorporated a *constructive dialogue* based upon

[6] The 207 subjudgments came from 23 subjects × three cases × three subjudgments (condition, motivation and attitude) per case.

[7] For this analysis, we used the level of strength from table 2 and the number of actions within each to determine the overall level of the actions chosen. If two sets of actions were of the same strength category but one had more actions in that category than the other, then it was considered stronger. For example, if an auditor selected "assign more experienced persons" and "move audit work to year end" for Case A, but only "assign more experienced persons" to Case B, then Case A was considered to have the stronger response even though both cases were categorized as "increase vigilance."

TABLE 4
Audit Planning Responses
Degree of Correspondence with Assessments and Risk Levels of Cases:
Frequencies of Subjects by Decision Aid

Degree of Correspondence		Frequencies of Subjects			
Category	**Definition**	**Unaided**	**Checklist**	**Logit Model**	**Expert System**
Panel A: Correspondence with Auditor's Risk Assessments					
Corresponded	The relation between audit planning response and risk assessment corresponded	5	7	6	14
Did Not Correspond	The relation between audit planning response and risk assessment did not correspond	18	17	17	9
Panel B: Correspondence with Actual Risk Levels					
Corresponded	The relation between audit planning response and actual risk levels corresponded	4	3	9	14
Did Not Correspond	The relation between audit planning response and actual risk levels did not correspond	19	21	14	9

psychology research designed to increase the engagement of the auditor in the decision-making process, thereby leading to greater reliance on the system. All outcomes of the study were consistent with predictions that the use of the *constructive dialogue* would increase the reliance on the system and increase the correspondence of subsequent decisions. The use of a decision aid that provided only a suggested assessment, the logit model, improved performance over the use of the checklist and those in the control group, but not to the extent of the expert system with the *constructive dialogue.*

This study contributes to our understanding of the use and reliance on decision aids in two important ways. First, the study supports the contention that a decision aid's design can increase reliance on that aid. The expert system's *constructive dialogue* incorporated features identified in the psychology literature. These five features were combined into a realistic computer environment to provide an omnibus test of their overall effect. Since this overall *constructive dialogue* increased reliance, future research should be undertaken to examine the effect of the individual features.

Second, the study investigates a decision subsequent to the use of the aid. The importance of the management fraud assessment stems from its potential effect on subsequent audit planning decisions. The auditors who used the expert system were able to make decisions regarding the additional audit actions that were more consistent with their assessment of the risk than the auditors in the other groups. This finding provides justification for the importance of continued decision aid research since it indicates that decision aid use can have wider impact than just the initial decision. Further research will need to be undertaken to understand fully how such sequential decisions are related.

The inclusion of both a checklist and an unaided group allowed us to investigate their support for the decision-making process. The

Pincus (1986) study of a red flag checklist, which focused primarily on the data acquisition phase of decision making and, secondarily, on cue combination, found a decreased performance for checklist users. This study, which focused only on the cue combination phase, found no significant differences between checklist users and unaided decision makers. Pincus' (1986) study underscored the importance of having the correct red flags on the checklist; our study highlights the incremental potential of creating decision aids that mechanically combine those red flags in a manner that enhances the auditor's reliance on the aid.

This study provides some insights into the influence of certain decision aids; however, there are also limitations and areas for future research. The current study examined realistic red flag sets using realistic decision aids, but the red flags were presented to the subjects without the context of a specific audit. This approach enabled us to test the influence of the decision aids on cue weighting and combination without the possible confound from red flag identification, but also limited our understanding of a complete decision setting. Future research should investigate the reliance on decision aids in a more context-rich environment in which auditors are presumed to: (1) locate the red flags in a complex audit situation, (2) assess the likelihood of management fraud, and (3) adjust a specified or self-determined audit plan.

This study used auditors from one Big 6 firm considered to be a relatively structured firm, making it difficult to generalize the results to other, less structured firms. However, the use of subjects from this firm did increase our internal validity as we were able to present the decision aids in a form with which they had familiarity from practice. In addition, we examined decision use on a one-time basis. Since the influence of decision aids might vary with continued use, it would be interesting to examine reliance over an extended time period.

REFERENCES

American Institute of Certified Public Accountants (AICPA). 1988. *The Auditor's Responsibility to Detect and Report Errors and Irregularities.* Statement on Auditing Standards No. 53. New York, NY: AICPA.

———. 1997. *Consideration of Fraud in a Financial Statement Audit.* Statement on Auditing Standards No. 82. New York, NY: AICPA.

Arkes, H. R., R. M. Dawes, and C. Christensen. 1986. Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes* 37: 93–10.

Arthur Andersen, Coopers & Lybrand, Deloitte & Touche, Ernst & Young, KPMG Peat Marwick, and Price Waterhouse. 1992. The liability crisis in the United States: Impact on the accounting profession. *Statement of Position* (August 6).

Ashton, A. H. 1985. Does consensus imply accuracy in accounting studies of decision making? *The Accounting Review* (April): 173–185.

Ashton, R. H. 1990. Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback and justification. *Journal of Accounting Research* 28 (Supplement): 148–180.

———. 1992. Effects of justification and a mechanical aid on judgment performance. *Organizational Behavior and Human Decision Processes* 52: 292–306.

Beach, L. R., B. D. Townes, F. L. Campbell, and G. W. Keating. 1976. Developing and testing a decision aid for birth planning decisions. *Organizational Behavior and Human Decision Performance* 15: 99–116

Bell, T. B., S. Szykowny, and J. J. Willingham. 1993. Assessing the likelihood of fraudulent financial reporting: A cascaded logit approach. Working paper, KPMG Peat Marwick, Montvale, NJ.

Blattberg, R. C. and J. S. Hoch. 1990. Database models and managerial intuition: 50% model and 50% manager. *Management Science* 36: 887–899.

Boatsman, J. R., C. Moeckel, and B. K. Pei. 1995. The effects of decision consequence on auditors' reliance on decision aids in audit planning. Working paper, Arizona State University.

Boritz, J. E. 1985. The effect of information presentation structures on audit planning and review judgments. *Contemporary Accounting Research* (Spring): 193–218.

Christ, M. Y. 1993. Evidence on the nature of audit planning problem representations: An examination of auditor free recalls. *The Accounting Review* (April): 304–322.

Davis, F. D. and J. E. Kottermann. 1995. Determinants of decision rule use in a production planning task. *Organizational Behavior and Human Decision Processes* 63: 145–157.

Dawes, R. M. 1971. A case study of graduate admissions: Application of three principles of human decision making. *American Psychologist* 26: 180–188.

———, and B. Corrigan. 1974. Linear models in decision making. *Psychological Bulletin* 81: 95–106.

———. 1979. The robust beauty of improper linear models in decision making. *American Psychologist* 34: 571–582.

Earley, P. C., T. Connolly, and G. Ekegren. 1989. Goals, strategy development, and task performance: Some limits on the efficacy of goal-setting. *Journal of Applied Psychology* (February): 24–33.

———, P. Wojnaroski, and W. Prest. 1987. Task planning and energy expended: Exploration of how goals affect performance. *Journal of Applied Psychology* (February): 107–114.

Einhorn, H. J. 1972. Expert measurement and mechanical combination. *Organizational Behavior and Human Decision Performance* (February): 86–106.

Eining, M. M., and P. B. Dorr. 1991. The impact of expert system usage on experiential learning in an auditing setting. *The Journal of Information Systems* (Spring): 1–16.

———, J. K. Loebbecke and J. J. Willingham. 1990. Expert systems: Issues for decision making in an auditing setting. *Third International Symposium on Expert Systems in Business, Finance, and Accounting*, Marina del Ray, CA.

Fischhoff, B., P. Slovik, and S. Lichtenstein. 1978. Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance* 4: 330–344.

Goldberg, L. R. 1970. Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin:* 422–432.

Hackenbrack, K. 1993. The effect of expertise with different sized clients on auditor evaluations of fraudulent financial reporting indicators. *Auditing: A Journal of Practice & Theory* (Spring): 99–110.

Johnson, P. E., K. Jamal, and R. G. Berryman. 1991. Effects of framing on auditor decisions. *Organizational Behavior and Human Decision Processes* (October): 75–105.

Kelly, K. P., G. S. Ribar, and J. J. Willingham. 1986. Interim report on the development of an expert system for auditor's loan loss evaluation. In *Artificial Intelligence in Accounting and Auditing,* edited by M. A. Vasarhelyi. New York, NY: Markus Wiener.

Kleinmuntz, B. 1990. Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin* 107: 296–310.

Libby, R. and P. A. Libby. 1989. Expert measurement and mechanical combination in control reliance decisions. *The Accounting Review* (October): 729–747.

Lichtenstein, S., B. Fischhoff, and L. D. Phillips. 1982. Calibration of probabilities: The state of the art to 1980. In *Judgment Under Uncertainty: Heuristics and Biases,* edited by D. Kahneman, P. Slovic, and A. Tversky. Cambridge, U.K.: Cambridge University Press.

Loebbecke, J. K, and J. J. Willingham. 1988. Review of SEC accounting and auditing enforcement releases. Unpublished working paper, University of Utah.

———, M. M. Eining, and J. J. Willingham. 1989. Auditors' experience with material irregularities: Frequency, nature and delectability. *Auditing: A Journal of Practice and Theory* (Fall): 1–28.

———, ———, and ———. 1991. The auditor's assessment of management fraud. *Conference on Auditing and Accounting Regulation, Center for Auditing Research and Education, Copenhagen Business School.*

McDaniel, L. S. and W. R. Kinney, Jr. 1995. Expectation-formation guidance in the auditor's review of interim financial information. *Journal of Accounting Research* (Spring): 59–76.

Mladenovic, R. and R. Simnett. 1994. Examination of contextual effects and changes in task predictability on auditor calibration. *Behavioral Research in Accounting* (Fall): 18–35.

O'Leary, D. E. and P. R. Watkins. 1989. Review of expert systems in auditing. *Expert Systems Review for Business and Accounting* (Spring/Summer): 3–22.

Palmrose, Z. 1987. Litigation and independent auditors: The role of business failures and management fraud. *Auditing: A Journal of Practice & Theory* (Spring): 90–102.

Peterson, D. K., and G. F. Pitz. 1986. Effect of input from a mechanical model on clinical judgment. *Journal of Applied Psychology* (January): 163–167.

Pincus, K. V. 1989. The efficacy of a red flags questionnaire for assessing the possibility of fraud. *Accounting, Organizations and Society* 14: 153–163.

Plous, S. 1993. *The Psychology of Judgment and Choice.* New York, NY: McGraw-Hill.

Todd, P. and I. Benbasat. 1991. An experimental investigation of the impact of computer based decision aids on decision making strategies. *Information Systems Research* (June): 87–115.

———. 1992. The use of information in decision making: An experimental investigation of the impact of computer-based decision aids. *MIS Quarterly* (September): 373–393.

———. 1994. The influence of decision aids on choice strategies: An experimental analysis of the role of cognitive effort. *Organizational Behavior and Human Decision Processes* 60: 36–74.

Whitecotton, S .M. 1996a. The effects of experience and a decision aid on the slope, scatter, and bias of earnings forecasts. *Organizational Behavior and Human Decision Processes* 66: 111–121.

———. 1996b. The effects of experience and confidence on decision aid reliance: A causal model. *Behavioral Research in Accounting* 8: 194–216.

Ye, L. R. and P. E. Johnson. 1995. The impact of explanation facilities on user acceptance of expert systems advice. *MIS Quarterly* (June): 157–172.