An experimental investigation of the effects of artificial intelligence systems on the training of novice auditors

Nitaya Wongpinunwatana
Department of Management Information System, Faculty of Commerce and Accountancy, Thammasat University, Bangkok, Thailand,
Colin Ferguson
Commerce Department, The University of Queensland, Brisbane, Australia
Paul Bowen
Commerce Department, The University of Queensland, Brisbane, Australia

Keywords
Auditors, Training, Artificial intelligence, Expert systems

Abstract
The primary objective of this research is to investigate the impact of task-technology fit on users' performance when using artificial intelligence systems for auditing tasks. Four artificial intelligence auditing systems, two problem-solving programs, and four questionnaires were developed. A laboratory experiment was performed with 292 undergraduate auditing students. The results suggested that the effect of task-technology fit on accuracy in solving problems was marginal for case-based reasoning with unstructured tasks. No significant effect was found on problem-solving accuracy for rule-based reasoning with structured tasks. The task-technology fit, however, marginally increased users' certainty of the correctness of their solutions.

1. Introduction

Artificial intelligence techniques such as rule-based reasoning[1] and case-based reasoning[2] can be used to implement intelligent tutoring systems. These training systems enhance cognitive information processing by providing the knowledge and the problem-solving strategies of experts to novice users (Murphy, 1990; Leidner and Jarvenpaa, 1995). Recently, institutions such as the University of Massachusetts and London School of Economics have developed intelligent tutoring systems to improve the effectiveness of both classroom education and on-the-job-training (Garcia, 1990; Angelides and Doukidis, 1990; Soloway, 1991; Woolf, 1996). Top engineering students at the University of Massachusetts indicated that intelligent tutoring systems helped them to develop a deeper level of understanding and enhanced communication with engineers and technicians (Woolf, 1996). Intelligent tutoring systems are also currently employed in the field of auditing (see, e.g. Böer and Livnat, 1990; Eining and Dorr, 1991; Morris, 1992; Marcella and Rauff, 1994). These systems improve novices' problem-solving performance (Böer and Livnat, 1990; Eining and Dorr, 1991).

Researchers are currently debating the benefits of case-based reasoning versus rule-based reasoning. Riesbeck and Schank (1989) believe that people do not reason from prior cases when well-established rules are available. In contrast, Ross (1989) argues that novices solve problems by using earlier examples without examining explicit statements of the relevant principles, explanations, or procedures. People solve problems either by using prior cases or by using rules, depending on how well each method fits the characteristics of the task. Consequently, either case-based reasoning or rule-based reasoning is likely to be more suitable for particular auditing domains, i.e. rule-based reasoning is effective for auditing activities that use theory to solve problems (Chi and Kiang, 1993; Allen, 1994). Conversely, case-based reasoning is proficient for auditing tasks that use experience to solve problems. Therefore, artificial intelligence systems that are more consistent with the characteristics of the task domain are expected to be more successful as training tools. This research investigates the impact of task-technology fit on users' performance in problem-solving. It also determines whether task-technology fit plays a role in the users' certainty of the correctness of their solutions.

The remainder of this paper proceeds as follows. The first section presents and explains the cognitive fit model. The second section presents a task-technology fit framework and the impact of task-technology fit on users. The third section describes the research method. The fourth section discusses the results and the last section provides conclusions, limitations of the research, and directions for future study.

2. Theoretical issue

Various theories from cognitive psychology have been adopted by researchers to explain the effect of expert systems on users' judgmental performance in problem solving. Some studies have applied cognitive theories, such as memory theory[3] and cognitive learning theory to predict the effects of artificial intelligence systems on problem solving and learning (see e.g. Murphy, 1990; Eining and Dorr, 1991). Other studies have used mental models theory[4] to predict the effect of matching the structure of the artificial intelligence system with the users' previous knowledge structures on problem solving (see e.g. Pei and Reneau, 1990).

The current issue and full text archive of this journal is available at

theory (see e.g. Vessey and Galletta, 1991; Sinha and Vessey, 1992; Vessey, 1994).

Cognitive fit theory predicts that effective and efficient problem solving occurs when the problem representation (information represented in different ways such as graphs and tables) and the task to be solved are matched (Vessey and Galletta, 1991; Sinha and Vessey, 1992; Vessey, 1994). In addition, it suggests that when technology matches task, the user develops an appropriate mental model for performance effects to occur (Vessey and Galletta, 1991).

Matching the problem representation directly to task has significant effects on problem-solving performance. Vessey and Galletta (1991) found that users make faster and more accurate decisions on symbolic problems[5] when information is represented in the form of tables than in the form of graphs. Their results show that users with graphs solve problems faster than those with tables on spatial tasks[6] but are less accurate than those with tables. Similarly, Jarvenpaa (1989) found that matching between the demands of the task and the graphical format used affects decision time but not decision accuracy. Although much empirical evidence supports the theory of cognitive fit, the problem-solving tasks employed in most research are simple (i.e. table and graph). The stability of the theory for complex tasks still remains untested (Vessey, 1994).

This study proposes that when the task and technology are well matched, users will develop a better mental model than when the task and technology are less well matched. The mental model that is developed then will enhance the users' performance in problem solving. This study differs from previous studies on task-technology fit in several ways. First, this study employs auditing tasks that are more complex than tasks used in the previous task-technology fit research. The previous studies typically used graphs and tables to represent problem solving tasks (for example, Jarvenpaa, 1989; Vessey and Galletta, 1991). Second, many studies have been conducted by using rule-based reasoning. The experiment used in this research adds case-based reasoning. Third, this study emphasises the matching of characteristics of auditing tasks to characteristics of artificial intelligence systems.

3. Hypotheses development

3.1 Task-technology fit framework

The task-technology fit framework permits the distinct characteristics of rule-based versus case-based reasoning systems to be evaluated. Rule-based reasoning is suitable for tasks with well-known and highly structured knowledge (Denna *et al.*, 1991; Slade, 1991; Winston, 1992; Gupta, 1994; Kesh, 1995). Rule-based reasoning outperforms case-based reasoning for tasks which focus on rules (Gupta, 1994). In addition, if past cases do not contain relevant knowledge that can be learnt, rule-based reasoning systems are preferred over case-based reasoning (Gupta, 1994). Rule-based reasoning systems, however, perform poorly when confronted with situations outside their problem domain. They cannot find solutions for problems that do not match the rules in their knowledge base. Thus, their capabilities to solve unfamiliar problems are limited.

Case-based reasoning is useful for tasks having several types of characteristics (Kolodner, 1993; Gupta, 1994; Zeleznikow and Hunter, 1995; Kesh, 1995). First, it is suitable when knowledge is incomplete. It allows users to fill in incomplete knowledge and then gives proposed answers. Though the solutions might not be optimal, they provide users with guidelines to generate their own solutions. Second, case-based reasoning is useful when no algorithmic methods are available to solve a problem. Using previous similar cases can be helpful in finding solutions when there are unknown or vaguely defined methods for evaluating problems. Nevertheless, users need to evaluate solutions based on what worked in the past. Third, cases are useful for interpreting open-ended and ill-defined concepts. Using specific knowledge such as cases to reason can be more useful than using generalisation (Kolodner, 1993). Finally, case-based reasoning systems match past cases to the new problem. These cases are ranked from high to low based on their strength of match. Users can analyse and use past cases to derive their own conclusions for the current problem (Kesh, 1995).

The different characteristics between rule-based reasoning and case-based reasoning are used to derive four factors involved in the matching of tasks to artificial intelligence technologies. These factors are

1. well-established guidelines;
2. input variables;
3. availability of algorithms; and
4. predefined solutions.

Well-established guidelines are clear and complete knowledge from textbooks, organisational policies, or experts. Guidelines are required for rule-based reasoning systems to solve problems. In auditing tasks, well-established guidelines consist of AICPA guidelines, questionnaires for internal control evaluation, etc. (Abdolmohammadi and Wright, 1987; Karan *et al.*, 1995). To solve problems, deductive reasoning is appropriate in areas where well-established guidelines to solve problems can be represented as a set of production rules (Allen, 1994; Gupta, 1994; Zeleznikow and Hunter, 1995). That is, rule-based reasoning is appropriate for well-defined tasks. In contrast, expert decision makers often compare current problems to similar past experiences where prior knowledge is ill-defined (Wong, 1993). For example, auditors use analogical reasoning for unfamiliar situations (Marchant, 1989; Koonce, 1993). When situations are familiar, experienced auditors use procedural knowledge instead of analogy (Marchant, 1989). Consequently, ill-defined knowledge is more appropriately stored in the form of cases.

Input variables are the parameters used to reach a conclusion. The input variables can be quantitative or qualitative. Qualitative variables are open-ended or ill-defined variables. Open-ended or ill-defined variables are not suitable for rule-based reasoning because the fuzzy problems cannot be structured in the form of production rules (Zeleznikow and Hunter, 1995). Case-based reasoning is more appropriate for ill-defined input variables.

Availability of algorithms is the availability of methods used to solve problems. Algorithms are sets of rules. If well-defined algorithms are available, rule-based reasoning is suitable. Conversely, case-based reasoning is more appropriate where well-defined algorithms are not available.

Predefined solutions are the possible solutions for a given problem. The solutions for the problem can be a single solution, a best solution, or plausible solutions. For alternative solutions, case-based reasoning is a suitable technique. The system provides not only a solution but also a set of cases similar to the current problem. The users are likely to accept the system's solution or can reach their own conclusions by using information from matched cases (Garfinkel, 1995). In contrast, for a single or a best solution, rule-based reasoning is appropriate. The system provides a solution for the problem.

Table I presents the task-technology fit framework showing how the four characteristics of tasks relate to artificial intelligence techniques. When the framework evaluation responses do not all favour a single technique, the most important factors are likely to be used to (Sutton, 1990). The main factors in building an artificial intelligence system are knowledge and input variables. Consequently, well-established guidelines and input variables are the first factors to be considered.

3.2 Research hypotheses

Figure 1 illustrates the model of the impact of task-technology fit on users. Hypotheses developed from this model are discussed in the following sub-sections.

3.2.1 Effect of task-technology fit on problem solving performance

Various artificial intelligence methods have been developed to cope with a broad range of problem domains. Each technique is applicable to different kinds of problems (Yoon and Guimaraes, 1993). The theoretical knowledge used in rule-based reasoning is suitable for solving structured tasks[7]. Conversely, the experience knowledge used in case-based reasoning is suitable for solving unstructured tasks[8]. Previous research shows that the task-technology fit affects users' performance. For example, both task-technology fit and use of systems significantly affect the effectiveness, productivity, and performance of users' jobs (Goodhue and Thompson, 1995).

Technology is likely to have a positive impact on individual performance if that technology fits the task well (Goodhue and Thompson, 1995). In addition, novices' accuracy of problem solving increases when the users use an expert system with structured tasks (Lamberti and Wallace, 1990). As a result, matching technology with tasks implies the following hypotheses:

H1a: When presented with structured tasks, novices trained using rule-based reasoning solve problems more accurately than novices trained using case-based reasoning.

H1b: When presented with unstructured tasks, novices trained using case-based reasoning solve problems more accurately than novices trained using rule-based reasoning.

3.2.2 Effect of task-technology fit on users' certainty of the correctness of their solutions

Pei and Reneau (1990) found that students have a high degree of certainty of the correctness of their solutions when they use an expert system with a structured task (inventory control and purchasing) that matches their previous knowledge gained through the lectures provided. In addition, their solutions when using financial software than participants in visual instruction on the computer monitor (Gist *et al.*, 1989). Therefore, it is likely that novices' certainty of the correctness of their solutions will increase after experiencing a tutoring system with a good task-technology fit. *H2a* and *2b* reflect the influence of task-technology fit on the users' certainty of the correctness of their solutions.

H2a: When presented with structured tasks, novices trained using rule-based reasoning have higher certainty of the correctness of their solutions than novices trained using case-based reasoning.

H2b: When presented with unstructured tasks, novices trained using case-based reasoning have higher certainty of the correctness of their solutions than novices trained using rule-based reasoning.

4. Research method

4.1 Research design

A laboratory experiment was designed to gather the data to test the hypotheses. The experimental design consists of task type with two levels and artificial intelligence technique with two levels. The two levels of the task type are structured and unstructured, while the two artificial intelligence techniques are rule-based reasoning (RBR) and case-based reasoning (CBR). Participants were divided into four groups: participants in rule-based reasoning for structured task, participants in rule-based reasoning for unstructured task, participants in case-based reasoning for structured task, and participants in case-based reasoning for unstructured task.

4.2 Construction of the research instruments

4.2.1 Selection of the task type

The proposed task-technology fit framework in subsection 3.1 was applied to classify six audit tasks (as shown in Table II). Empirical data from Karan *et al.* (1995) were used to test the framework. Features that are relevant to well-established guidelines and predefined solutions were selected from Karan *et al.*'s questionnaire. The data about input variables and availability of algorithms were collected from four experts. The degree of auditors' agreement with each statement was averaged for a grand total score. A nine-point scale from 1 for "no" to 9 for "yes" was used for characteristics of tasks. Open-ended or ill-defined input variables were reverse coded (R). As a result, a score of nine indicates a greater degree of suitability of the task for rule-based reasoning. A score of one indicates that case-based reasoning is the most appropriate. The midpoint of the scale is used to divide the selection of artificial intelligence techniques. Case-based reasoning is more suitable if the grand total score is between 1 and 4.5. A grand total score between 4.6 and 9 indicates that rule-based reasoning is likely to be more appropriate.

Table II indicates that case-based reasoning is likely to be suitable for going concern judgments, inherent risk assessments, and for setting materiality levels. Rule-based reasoning is likely to be more appropriate for disclosure compliance evaluations, control risk assessments, and for adequacy of allowance account evaluations (e.g. allowance for doubtful accounts, loan loss reserves, etc).

Two common tasks, going concern decisions and internal control over purchases evaluation [9], were selected. These tasks were chosen because they possess distinct characteristics that can be used to differentiate them. In addition, there were an adequate number of previous studies which provided information for task assessment.

4.2.2 Development of rules and cases

Two types of knowledge bases, namely rules

Table I
Task-technology fit framework

Characteristics of tasks	Artificial intelligence techniques: Rule-based reasoning	Case-based reasoning
Well-established guidelines	Yes	No
Open-ended or ill-defined input variables	No	Yes
Availability of algorithms to derive solution	Yes	No
Predefined solutions	Yes	No

Figure 1
Model of the impact of task-technology fit on users

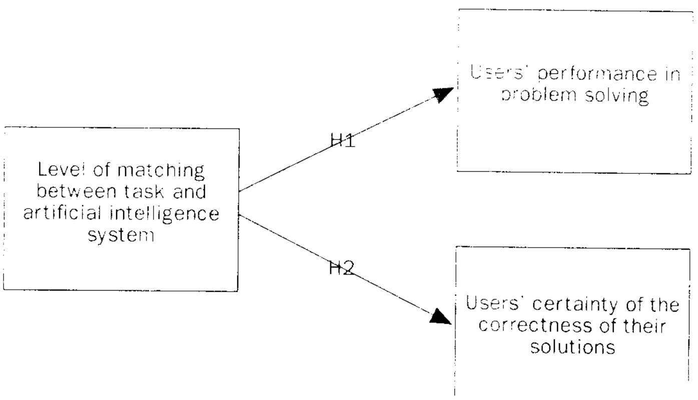


Table II

Classification [illegible] of auditing task by using task [illegible] framework

	GC	MAT	IR	CR	ALL	GA[illegible]
Well-established [illegible] defines:	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1. The knowledge required for the task is well defined by AICPA guidelines, [illegible], textbooks, firm policies, etc.	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2. The knowledge available for performing the task is incomplete (R)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3. The knowledge available for performing the task is unclear (R)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4. The task often presents unique situations that experts have never encountered before (R	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Input variables:						
1. The task requires input data that include open-ended or ill-defined concepts (R)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Availability of algorithms:						
1. The task has a good algorithm or method from which to derive solutions	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Predefined solutions						
1. The task involves finding either one solution, a best solution, or all plausible solutions (Enter 1 for one solution, 5 for best solution and 9 for all plausible solutions) (R)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2. Decisions for this task involve selecting between close alternatives	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3. The possible task solutions are pre-defined and the objective is to select from among this set of solutions	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Grand total score	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Notes: [illegible]

Sources: [illegible] predefined solutions [illegible] algorithms are from an auditing [illegible] analysed in Keen et al. [illegible]

in rule-based and case-based reasoning systems respectively. The four main factors in internal control over purchases evaluations and going concern decisions[10], and the information used to develop the help files were retrieved from:

- auditing text books (Gill et al., 199[illegible]; Trotman, 19[illegible]; Arens et al., 199[illegible]; Gill and Cosserat, 199[illegible]);
- uniform final examination reports (The Institute of Chartered Accountants of Canada, 1995); and
- articles in going concern evaluation (Boritz, 1991; Selfridge et al., 1993).

The rules were used to create cases so that both rules and cases have the same auditing factors. Each case was limited to one A4 page of text. The cases had to be short enough so that the participants would be able to learn how the auditor made judgement in the time allocated for the experiment. To simulate real world cases, the [illegible] cases were [illegible]

companies in Australia[11]. The internal control cases were developed using one of big six auditing internal control questionnaire [illegible] manuals.

Because the [illegible] had to provide [illegible] of the internal [illegible] over purchases [illegible] and the going concern evaluation [illegible] partners from three of the big six auditing firms in Thailand were asked to evaluate and give the rating for each rule (or factor) for [illegible] auditing tasks. A seven point scale from "extremely weak" [illegible] disagree" to "extremely strong" [illegible] was used for the rating of [illegible] internal control [illegible] purchase evaluation [illegible] going concern decision respectively. The mean score [illegible] rating of each rule (or factor) was calculated to represent the [illegible] judgement. Meanwhile, [illegible] of each case was derived from the [illegible] rules for the [illegible] ensure the [illegible]

The systems were [illegible] answers to questions.

The rule-based reasoning system provided procedural explanations[13]. Such explanations have been found to have positive effects on users' performance in problem solving by Lamberti and Wallace (1990). The case-based reasoning system provided details of cases and the reasoning used to derive their solutions. Unlike rule-based reasoning explanations, little research has been conducted on case-based reasoning explanations (Kolodner, [illegible]).

4.2.4 Development of problem-solving programs

As participants had to enter answers to questions on the six problem-solving cases using microcomputers, the two problem-solving programs (internal control over purchases evaluations and going concern decisions) were developed using Visual Basic. Each program has two screens: the instruction screen and problem-solving screen. The instruction screen contains the instruction to answer the case problems, the scale used to rate the internal control over purchases evaluation or going concern decision, and the degree of certainty of the correctness of the user's solution. The problem-solving screen has two windows: the case window and the question-answer window. The case window provides the problem-solving case (as explained in subsection 4.2.2). The question-answer window is for the questions about the case and their answers (the [illegible] of internal control or going concern, the certainty of the correctness of user's solution, and the brief explanation of the solution).

4.2.5 Test of the research instruments

To avoid bias and ambiguity in the research instruments, a pilot test was conducted with 12 postgraduate students studying an auditing course. Four out of the 12 students were also studying an information system subject. Three students were allocated to each group. Each student took the two-hour [illegible] university in Australia. [illegible] participated in this study. These students were selected because they represent novice auditors. [illegible] experiment was included in the course outline as an artificial intelligence exercise (AI) worth [illegible] per cent of the total course grade. Students were informed that the score depended on their participation in [illegible] AI procedures for the full two hours as well as the quality of their answers to the problem-solving cases.

4.4 Experimental procedures

The experiment was conducted in the computer rooms of the [illegible] of [illegible], Economics, and Law. The experiment [illegible] over a two-hour period and consisted of [illegible] sections: instructions (ten minutes), AI training (one hour), problem-solving [illegible] minutes), and commenting on the experiment (ten minutes). During AI training, participants [illegible] with rule-based reasoning (going concern decision or internal control over purchases evaluation) or case-based reasoning (going concern decision or internal control over purchases evaluation). During the problem-solving section, the participants answer the rating of going concern decision or internal control over purchases evaluation, the degree of certainty of the correctness of user's solution, and briefly explain the reason for the solution without consulting artificial intelligence systems.

5. Results

This section reports the results of assessing the variables with respect to the assumptions underlying the statistical tests and also reports the research results. The SAS statistical package, run under UNIX, was used to evaluate the assumptions before proceeding with a factorial multivariate analysis of covariance with repeated measures (factorial MANCOVA).

5.1 Evaluation of assumptions

Of students studying an auditing subject, 292 participated in this study. Responses from 60 of the 292 participants were not included in the analysis because of incomplete data, not following instructions etc. Deletion of this unusable data reduced the available sample size to 232 participants. Sample sizes of the remaining data were quite similar for the four groups: 60 participants in rule-based reasoning for internal control over purchases evaluation, 61 participants in case-based reasoning for internal control over purchases evaluation, 57 participants in rule-based reasoning for going concern decision, and 54 participants in case-based reasoning for going concern decision. Table III presents demographic data for the participants who provided usable results in each group. Results of the evaluation of assumptions of normality, homogeneity of variance, linearity, and multicollinearity of each dependent variable were satisfactory.

5.2 Analysis of the results and discussion

Factorial multivariate analysis of covariance with repeated measures (factorial MANCOVA) was used to analyse the effect of task-technology fit on the data of 232 participants. This statistic was used because each participant answered six problem-solving cases. The analysis was performed on dependent variables by cases. The independent variable is effects of two artificial intelligence systems within two auditing tasks. Because the experiment conducted by Murphy (1990) indicated that grade point average (GPA) was significant in his model, GPA was used as a covariate to adjust for the participants' performance differences in this study. A p-value of 0.05 was used as the significance level in the statistical tests. A p-value between 0.05 to 0.10 is deemed as marginally significant.

5.2.1 Effect of task-technology fit on level of accuracy in solving problems

H1a suggests that accuracy in solving structured tasks is higher for RBR than for CBR. *H1b* suggests that accuracy in solving unstructured tasks is higher for CBR than for RBR. The results of a factorial MANCOVA with repeated measures of the combined six problem-solving cases after adjustment for the covariate in Table IV indicate only marginal significance for the effect of task-technology fit on the level of accuracy in solving the problem with a p-value of 0.0815. The GPA covariate was found to be associated with the level of accuracy in solving the problem with a p-value of 0.0004.

Analysis of the individual six problem-solving cases after adjustment for the covariate in Table V (see also Figure 2) indicates that the marginal significance of task-technology fit was made by case 1. The task-technology fit appears to affect users' performance in problem solving in terms of accuracy with unstructured tasks in the directions predicted. The sign of the parameter estimates indicates that participants trained using case-based reasoning have fewer errors in solving going concern decisions than those trained using rule-based reasoning. Therefore, some support was obtained for *H1b*. However, only one of six problem-solving cases was significant ($p = 0.0001$). A few possible reasons may explain this finding. First, participants may experience fatigue with the experiment. Second, the amount of time spent in the experiment might not be enough for participants to develop appropriate knowledge from the systems. This result is consistent with those obtained by Borgman (1986), Murphy (1990), and Gregor (1996), who found that insufficient experiment time hinders the novice users' ability to develop cognitive learning of the systems. Furthermore, some participants made the following comments, which support the above assertions.

Table III
Demographic data for the participants who provided usable results

Items	Internal control over purchases evaluation RBR[a] (n = 60)	CBR[b] (n = 61)	Going concern decision RBR[a] (n = 57)	CBR[b] (n = 54)
Gender: Male	31	32	27	25
Female	29	29	30	29
Mean score (Std. Dev.) of age	20.47 (1.32)	20.38 (1.25)	20.72 (2.08)	21.65 (2.29)
Mean score (Std. Dev.) of grade point average	4.74 (0.78)	4.92 (0.74)	4.95 (0.72)	4.98 (0.69)

Table IV
Effect of task-technology fit on level of accuracy in solving problems

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SYSTEM(TASK)[a]	2	2.816	1.408	2.53	0.0815**
GPA	1	7.158	7.158	12.89	0.0004*
Error	227	12.087	0.555		

Notes: * $p < 0.05$; ** $p < 0.10$; [a] Effect of task-technology fit on level of accuracy in solving problems

There are too many problems to be solved. It is interesting to answer the first three questions but not the others. Maybe a longer time period could be allowed to answer the problems.

Table V
Effect of task-technology fit on level of accuracy in solving problems for cases 1 to 6

Case[a]	DF	Parameter estimate[b]	T for H_0: parameter = 0	Pr > |T|[c]	R^2
Case 1:	4, 227				0.1756
RBR – Internal control		–0.0352	–0.20	0.8393	
CBR – Internal control		0.0000			
CBR – Going concern		–0.7510	–4.17	0.0001*	
RBR – Going concern		0.0000			
Case 2:	4,227				0.1426
RBR – Internal control		0.1655	1.41	0.1607	
CBR – Internal control		0.0000			
CBR – Going concern		0.1942	1.59	0.1137	
RBR – Going concern		0.0000			
Case 3:	4,227				0.0970
RBR – Internal control		0.0994	0.87	0.3827	
CBR – Internal control		0.0000			
CBR – Going concern		–0.0728	–0.62	0.5386	
RBR – Going concern		0.0000			
Case4:	4,227				0.0224
RBR – Internal control		–0.0453	–0.33	0.7414	
CBR – Internal control		0.0000			
CBR – Going concern		–0.1596	–1.12	0.2638	
RBR – Going concern		0.0000			
Case 5:	4,227				0.0065
RBR – Internal control		0.0499	0.43	0.6641	
CBR – Internal control		0.0000			
CBR – Going concern		0.0359	0.30	0.7642	
RBR – Going concern		0.0000			
Case 6:	4,227				0.0475
RBR – Internal control		0.0115	0.10	0.9185	
CBR – Internal control		0.0000			
CBR – Going concern		0.0157	0.13	0.8930	
RBR – Going concern		0.0000			

Notes: * $p < 0.05$; ** $p < 0.10$; [a] Comparisons of the effect of task-technology fit on level of accuracy in solving problem: RBR = rule-based reasoning, CBR = case-based reasoning; Internal control = structured task, Going concern = unstructured task; [b] a negative sign of the first parameter estimate on each pair of independent variables means that the first user group (i.e. the user group that performs the first task in the pair of independent variables) has fewer errors (or more accuracy) in solving problems than the second user group. The positive sign of the first parameter estimate indicates that the first user group does not have fewer errors than the second user group; [c] a student's t probability (Pr > |T|) indicates whether mean differences of each pair of independent variables are significant

Third, detailed analysis revealed that the impact of task-technology fit showed only in the case of simple problems (case 1 in Table V). The reason may be the effect of immediate (or short-term) learning. Immediate learning, however, is facilitated by simple tasks (Bonner, 1994).

H1a is not supported. All p-values of the six problem-solving cases are greater than 0.10. In addition, some of the signs of the parameter estimates are not in the predicted direction. Contrary to expectations, participants who used rule-based reasoning with structured tasks solve problems less accurately. The results observed could be due to the nature of the structured task used in the experiment. The structured task might be easier to understand and learn in respect to its characteristics. In addition, the participants might have possessed adequate skills for this task prior to the experiment. Some aspects of internal control and the skills needed to solve the problems are taught in other courses. Consequently, it appears that prior knowledge may reduce the benefits of the system in training novices.

5.2.2 Effect of task-technology fit on degree of users' certainty of the correctness of their solutions

H2a suggests that degree of novices' certainty of the correctness of their solutions related to structured tasks is higher for RBR than for CBR. *H2b* suggests that degree of novices' certainty of the correctness of their solutions related to unstructured tasks is higher for CBR than for RBR. Some marginal support was obtained for the effect of task-technology fit on the degree of novices' certainty of the correctness of their solutions with a p-value of 0.1034 (see Table VI). The GPA covariate was also found to marginally affect novices' certainty of the correctness of their solutions with a p-value of 0.0889.

Analysis of the individual six problem-solving cases after adjustment (see also Figure 3) for the covariate in Table VII indicates that the degree of users' certainty of the correctness of their solutions of cases 1, 2, and 3 was affected by the task-technology fit. The participants trained using rule-based reasoning show more certainty of the correctness of their solutions of internal control over purchases evaluation than those trained using case-based reasoning, i.e. cases 1 and 2 in Table VII with *p*-values of 0.0387 and 0.0159 respectively. The participants trained using case-based reasoning indicate more certainty of the correctness of their solutions of going concern decisions than those trained using rule-based reasoning (cases 1 and 3 in Table VII with *p*-values of 0.0705 and 0.0087 respectively). Hence, the results provide some support for *H2a* and *2b*. All *p*-values of cases 4, 5, and 6 were greater

Figure 2

Comparison of mean score of errors in solving problems for cases 1 to 6

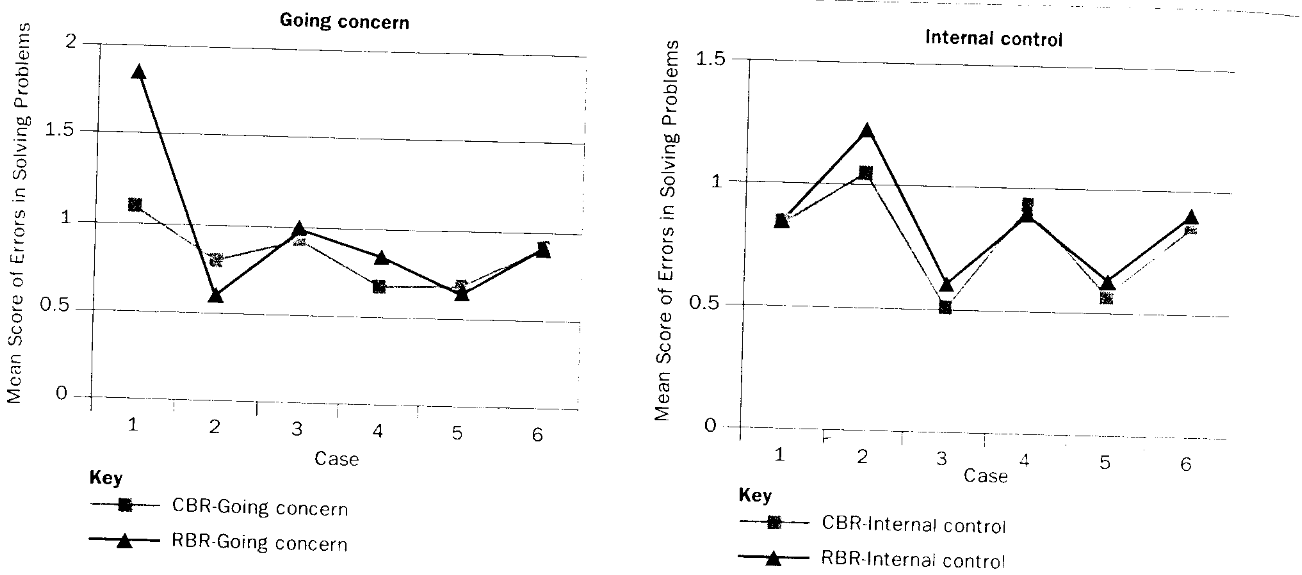


Table VI

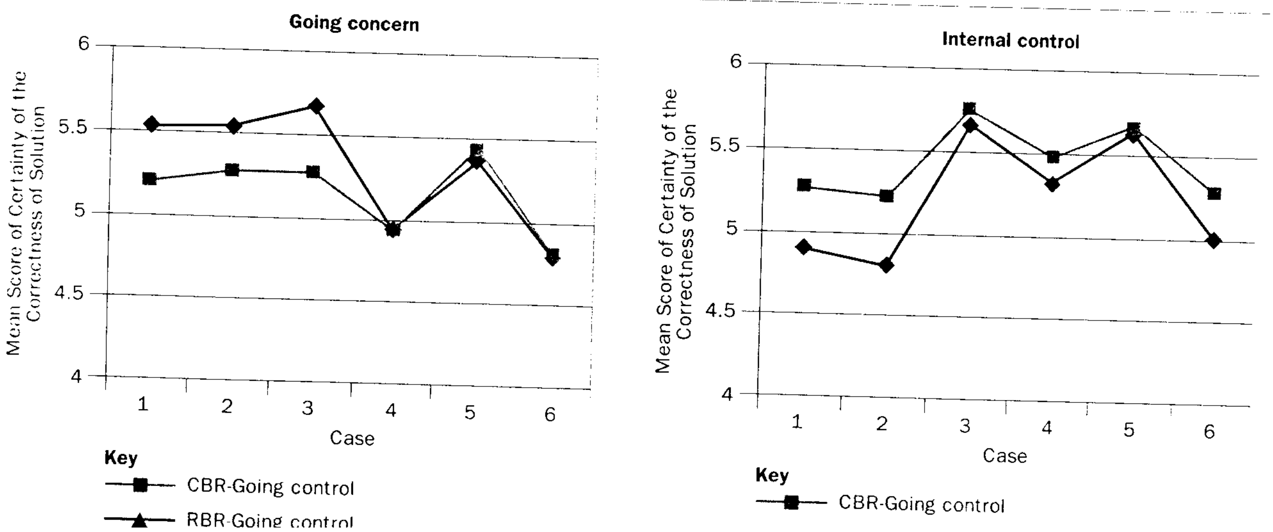
Effect of task-technology fit on degree of novices' certainty of the correctness of their solutions

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SYSTEM(TASK)[a]	2	11.701	5.850	2.29	0.1034***
GPA	1	7.450	7.450	2.92	0.0889**
Error	227	579.335	2.552		

Notes: ** $p < 0.10$; *** $p = 0.1034$; [a] Effect of task-technology fit on degree of novices' certainty of the correctness of their solutions

Figure 3

Comparison of mean score of certainty of the correctness of solutions for cases 1 to 6



than 0.10. The parameter estimates have both expected and unexpected signs.

It may be possible that the comparisons of mean differences of the four groups are significant by excluding some cases from the analysis. The cases excluded were mostly from the missing value and may be representative of the population. The findings of this research, however, revealed the significance of both combined effect of six problem solving cases and individual tests of each case within a particular main effect. Furthermore, factorial MANCOVA with repeated measures was run with all data (292 students). The statistical results for all data are quite similar to the results obtained after discarding some cases. Even though the results indicate some improvement after cleaning the data, the improvement is not significant. In addition, factorial MANCOVA with repeated measures will throw out incomplete data in its calculations. As a result, the significant results are more likely to be due to the matching of task to technique (or manipulations) rather than due to chance.

Table VII
Effect of task-technology fit on degree of users' certainty of the correctness of their solutions for cases 1 to 6

Source[a]	DF	Parameter estimate[b]	T for H_0: parameter = 0	Pr > |T|[c]	R^2
Case 1:	4,227				0.0557
RBR – Internal control		0.3693	2.08	0.0387*	
CBR – Internal control		0.0000			
CBR – Going concern		0.3356	1.82	0.0705**	
RBR – Going concern		0.0000			
Case 2:	4,227				0.0903
RBR – Internal control		0.3933	2.43	0.0159*	
CBR – Internal control		0.0000			
CBR – Going concern		0.2731	1.62	0.1062	
RBR – Going concern		0.0000			
Case 3:	4,227				0.0662
RBR – Internal control		0.0634	0.44	0.6616	
CBR – Internal control		0.0000			
CBR – Going concern		0.3977	2.64	0.0087*	
RBR – Going concern		0.0000			
Case 4:	4,227				0.0826
RBR – Internal control		0.1473	0.94	0.3467	
CBR – Internal control		0.0000			
CBR – Going concern		0.0097	0.06	0.9524	
RBR – Going concern		0.0000			
Case 5:	4,227				0.0226
RBR – Internal control		0.0285	0.18	0.8581	
CBR – Internal control		0.0000			
CBR – Going concern		–0.0776	–0.47	0.6391	
RBR – Going concern		0.0000			
Case 6:	4,227				0.0539
RBR – Internal control		0.2491	1.38	0.1702	
CBR – Internal control		0.0000			
CBR – Going concern		–0.0223	–0.12	0.9058	
RBR – Going concern		0.0000			

Notes: * $p < 0.05$; ** $p < 0.10$; [a] Comparisons of the effect of task-technology fit on users' certainty of the correctness of their solutions; RBR = rule-based reasoning, CBR = case-based reasoning; Internal control = structured task; Going concern = unstructured task; [b] a positive sign of the first parameter estimate on each pair of independent variables means that the first user group (i.e. the user group that performs the first task in the pair of independent variables) has more degree of users' certainty of the correctness of their solutions than the second user group. The negative sign of the first parameter estimate indicates that the first user group does not have more degree of users' certainty of the correctness of their solutions than the second user group; [c] a student's t probability (Pr > |T|) indicates whether mean differences of each pair of independent variables are significant

6. Conclusions and suggestions for future research

6.1 Conclusions

To conclude, this research has contributed to knowledge in the following ways. First, it developed a task-technology fit framework and a model of the impact of matching between a task and an artificial intelligence system on users. Second, it extended prior research on the impact of expert systems by comparing the effect of two different artificial intelligence systems with two different auditing tasks on performance in problem solving. Third, the results obtained from this research have partially provided support that the task-technology fit affects the users' performance in problem-solving. Finally, the research results also provide evidence that the task-technology fit affects the novices' performance in problem solving and certainty of the correctness of their solutions. The educational institutions or accounting firms (that want to use artificial intelligence systems as training tools) have to consider that different types of tasks affect the performance of users in solving problems differently.

6.2 Limitations of the research

The results of this study are limited to only the novice auditors so they may not be generalised beyond the novice auditors (i.e. the auditing experts). In addition, the research is applicable only to auditing novices studying the auditing course in Australia. The research may not be generalised to all auditing novices in countries where they may emphasise different auditing aspects. Moreover, this experiment concerns only a section of auditing. Therefore, the results of the research in this study should be generalised with care.

As the experiment was conducted in a short period of time (two hours), with the participants using the artificial intelligence system for the first time, the participants may not be able to learn because of a lack of time for learning. As a result, the findings of this research will not be generalised to situations where the participants used the system for a long period of time.

The use of four experts for data about input variables and availability of algorithms (see subsection 4.2.1) may not be a representative of auditors' agreement. Furthermore, the proposed method of selection of the task type may not be the best approach to match task to technology. Therefore, different methods could yield different results.

6.3 Directions for future research

The results of this study suggest several areas for future research. First, as this research required users to answer not only the rating of the problems but also to briefly explain the reasons for the solutions, the time measured may also include the time spent on explaining the solution. Future research could measure only the time required to solve the task. Furthermore, the replicated research should expand the experimental time (i.e. one week) to measure long-term learning which is not short-term recall (Eining and Dorr, 1991).

Second, the effect of task-technology fit on the confidence and perceptions of users needs to be investigated further. As this research used the retrieval-only case-based reasoning system which provides only similar cases to the problems, the users have the responsibility to derive their own solutions through using these retrieved cases. To reach the solution, the users, in particular the novices, may find that it is a difficult task. If the case-based reasoning system provides the answer to the problem automatically, users may show high confidence and favourable perceptions of the case-based reasoning system. Therefore, the studies could be conducted to address the effect of automatic solutions from the artificial intelligence system.

Third, this research did not allow the users to access the systems while they solved the problems. The research used the artificial intelligence systems as the training tools. The systems, however, can be used as decision aids. It is possible that the system may assist users by enhancing more efficient and effective performance in problem solving as the users can consult the system during solution time.

Notes

1 Rule-based reasoning systems are typified by expert systems that represent knowledge in the form of rules. These systems match

2 Case-based reasoning systems represent knowledge in the form of cases. These systems use previous experiences or cases to solve new problems.

3 Memory is an aid to problem solving (Glass et al., 1979). Problem-solving skills depend on how users encode sets of facts in their memory and later retrieve this knowledge to produce results or answers for particular tasks or problems (Glass *et al.*, 1979; McArthur, 1987; Murphy, 1990).

4 Mental models are conceptual representations of the software that the user builds in his or her mind when he or she interacts with a system (Borgman, 1986; Sein and Bostrom, 1989; Satzinger, 1994).

5 A symbolic task involves retrieving discrete or precise data value (Vessey, 1994).

6 A spatial task involves a comparison of trends (Vessey, 1994).

7 Structured tasks are routine or well-defined problems (Abdolmohammadi and Wright, 1987). The tasks demand only limited judgment or experience from a decision maker, e.g. the completion of standardised questionnaires with well-specified factors.

8 Unstructured tasks are undefined problems with no clear guidelines (Abdolmohammadi and Wright, 1987). To accomplish unstructured tasks requires the experience, knowledge, and insight of an expert decision maker.

9 The internal control over purchases evaluations is linked to control risk assessment. As internal control incorporates many applications (such as cash, account receivable, inventories, and purchases), one application was chosen. The internal control over purchases evaluations was selected because there are many sources of information (e.g. AICPA) which can be retrieved for this task.

10 The four factors in internal control over purchases evaluations are preparing purchase orders, receiving goods, preparing payment vouchers, and recording liabilities. The four financial factors in going concern are liquidity factors, operating factors, debt structure factors, and other financial factors. Although auditors also need to evaluate internal control over requisitioning of goods, and the implications of management, industry and external factors for the going concern decision, these aspects are not examined in this study. Covering all aspects of internal control and going concern decisions in the research would make the auditing tasks too complicated to use as training tools for novice auditors.

11 The activities of these companies were mining, property development, tourism, transportation, insurance, and high technology. The use of liquidating companies

going concern cases. The going concern cases in case-based reasoning system. however, were modified such that the going concern problems were equally spread over different levels of going concern factors.

12 This study examines rule-based reasoning and case-based reasoning systems. These artificial intelligence systems were chosen because both can provide explanations on how they reach their conclusions. Although some types of artificial intelligence systems such as neural networks can provide the same functions as the rule-based and case-based reasoning systems, neural networks are not the concern of this study. Neural networks do not yet contain the explanation facilities available in the systems used in this research (Gregor, 1996).

13 Procedural explanations include two parts: premise and action. The premise part (IF) states all necessary condition contexts. The action part (THEN) indicates one or more conclusions that can be drawn if the premises are satisfied.

References

Abdolmohammadi, M. and Wright, A. (1987), "An examination of the effects of experience and task complexity on audit judgments", *The Accounting Review*, Vol. LXII, No. 1 January, pp. 1-13.

Allen, B. (1994), "Case-based reasoning: business applications", *Communications of the ACM*, Vol. 37 No. 3, March, pp. 40-2.

Angelides, M.C. and Doukidis, G.I. (1990), "Is there a place in OR for intelligent tutoring systems?", *Journal of the Operational Research Society*, Vol. 41 No. 6, June, pp. 491-503.

Arens, A.A., Best, P.J., Shailer, G.E.P. and Loebbecke, J.K. (1996), *Auditing in Australia: An Integrated Approach*, 3rd ed., Prentice-Hall of Australia Pty Ltd.

Böer, G.B. and Livnat, J. (1990), "Using expert systems to teach complex accounting issue", *Issues in Accounting Education*, Vol. 5 No. 1, Spring, pp. 108-19.

Bonner, S.E. (1994), "A model of the effects of audit task complexity", *Accounting, Organizations and Society*, Vol. 19 No. 3, pp. 213-34.

Borgman, C.L. (1986), "The user's mental model of an information retrieval system: an experiment on a prototype online catalog", *International Journal of Man-Machine Studies*, Vol. 24, pp. 47-64.

Boritz, J.E. (1991), *The Going Concern Assumption: Accounting and Auditing Implications*, The Canadian Institute of Chartered Accountants.

Chi, R.T.H. and Kiang, M.Y. (1993), "Reasoning by coordination: an integration of case-based and rule-based reasoning systems", *Knowledge-Based Systems*, Vol. 6 No. 2, June, pp. 103-13.

Denna, E.L., Hansen, J.V. and Meservy, R.D. (1991), "Development and application of expert systems in audit service", *IEEE Transaction on Knowledge and Data Engineering*, Vol. 3 No. 2, pp. 172-84.

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

Garcia, A.L. (1990), "Intelligent training systems: smart tutors", *Bulletin of the American Society for Information Science*, Vol. 16 No. 6, August September, pp. 8-9.

Garfinkel, S. (1995), "AI as training tool", *Technology Review*, Vol. 98 No. 6, August/September, pp. 16-18.

Gill, G.S. and Cosserat, G.W. (1996), *Modern Auditing in Australia*, 4th ed., Jacaranda Wiley Ltd, Australia.

Gist, M.E., Schwoerer, C. and Rosen, B. (1989), "Effects of alternative training methods on self-efficacy and performance in computer software training", *Journal of Applied Psychology*, Vol. 7 No. 6, pp. 884-91.

Goodhue, D.L. and Thompson, R.L. (1995), "Task-technology fit individual performance", *MIS Quarterly*, June, pp. 213-36.

Gregor, S. (1996), "Explanations from knowledge-based systems for human learning and problem solving", unpublished PhD dissertation, University of Queensland, Brisbane, Australia.

Gul, F.A., Teoh, H.Y., Andrew, B.H. and Schelluch, P. (1994), *Theory and Practice of Australian Auditing*, 3rd ed., Nelson Australia Pty Limited, Australia.

Gupta, U.G. (1994), "How case-based reasoning solves new problems", *Interfaces*, Vol. 24 No. 6, pp. 110-19.

(The) Institutes of Chartered Accountants of Canada (1995), *Internal Control Questionnaire for Small Businesses*, The Canada Institute of Chartered Accountants.

Jarvenpaa, S.L. (1989), "Effect of task demand and graphical format on information processing strategies", *Management Science*, Vol. 35 No. 3, pp. 285-303.

Karan, V., Murthy, U.S. and Vinze, A.S. (1995), "Assessing the suitability of judgemental auditing tasks for expert systems development: an empirical approach", *Expert Systems With Applications*, Vol. 9 No. 4, pp. 441-55.

Kesh, S. (1995), "Case based reasoning", *Journal of Systems Management*, Vol. 46 No. 4, pp. 14-19.

Kolodner, J. (1993), *Case-Based Reasoning*, Morgan Kaufmann Publishers, Inc., San Mateo, CA.

Koonce, L. (1993), "A cognitive characterization of audit analytical review", *Auditing: A Journal of Practice & Theory*, Vol. 12, Supplement, pp. 57-76.

Lamberti, D.M. and Wallace, W.A. (1990), "Intelligent interface design: an empirical assessment of knowledge presentation in expert systems", *MIS Quarterly*, September, pp. 279-311.

Leidner, D.E. and Jarvenpaa, S.L. (1995), "The use of information technology to enhance management school education: a theoretical view", *MIS Quarterly*, September, pp. 265-91.

Marcella, A.J. Jr and Rauff, J.V. (1994), "Utilizing expert systems to evaluate disaster recovery planning", *Journal of Applied Business Research*, Vol. 11 No. 1, Winter, pp. 30-8.

Marchant, G. (1989), "Analogical reasoning and hypothesis generation in auditing", *The Accounting Review*, Vol. LXIV No. 3, July, pp. 500-13.

Morris, B.W. (1992), "Case-based reasoning in internal auditor evaluations of EDP controls", PhD dissertation, University of Pittsburgh, PA.

Murphy, D. (1990), "Expert system use and the development of expertise in auditing: a preliminary investigation", *Journal of Information Systems*, Fall, pp. 18-35.

Pei Buck, K.W. and Reneau, J.H. (1990), "The effects of memory structure on using rule-based expert systems for training: a framework and empirical test", *Decision Sciences*, Vol. 21, pp. 263-86.

Riesbeck, C.K. and Schank, R.C. (1989), *Inside Case-Based Reasoning*, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.

Ross, B.H. (1989), "Some psychological results on case-based reasoning", in *Proceedings of a Workshop on Case-Based Reasoning*, Pensacola Beach, FL.

Selfridge, M., Biggs, S.F. and Krupka, G.R. (1993), "A computational model of the auditor knowledge and reasoning processes in the going-concern judgment", *Auditing: A Journal of Practice & Theory*, Vol. 12, pp. 82-112.

Sinha, A.P. and Vessey, I. (1992), "Cognitive fit: an empirical study of recursion and iteration", *IEEE Transactions on Software Engineering*, Vol. 18 No. 5, May, pp. 368-79.

Slade, S. (1991), "Case-based reasoning: a research paradigm", *AI Magazine*, Spring, pp. 42-55.

Soloway, E. (1991), "Quick, where do the computers go?", *Communications of the ACM*, Vol. 34 No. 2, February, pp. 29-33.

Sutton, S.G. (1990), "Toward a model of alternative knowledge representation selection in accounting domains", *Journal of Information Systems*, Fall, pp. 73-85.

Trotman, K.T. (1995), *Cases Studies in Auditing*, Butterworth, Australia.

Vessey, I. and Galletta, D. (1991), "Cognitive fit: an empirical study of information acquisition", *Information Systems Research*, Vol. 2 No. 1, pp. 63-84.

Vessey, I. and Galletta, D. (1994), "The effect of information presentation on decision making: a cost-benefit analysis", *Information and Management*, Vol. 27, pp. 103-19.

Winston, P.H. (1992), *Artificial Intelligence*, 3rd ed., Addison-Wesley, Reading, MA.

Wong, E.D. (1993), "Understanding the generative capacity of analogies as a tool for explanation", *Journal of Research in Science Teaching*, Vol. 30 No. 10, pp. 1259-72.

Woolf, B.P. (1996), "Intelligent multimedia tutoring systems", *Communications of the ACM*, Vol. 39 No. 4, April, pp. 30-1.

Yoon, Y. and Guimaraes, T. (1993), "Selecting expert system development techniques", *Information & Management*, No. 4, April, pp. 209-23.

Yoon, Y., Guimaraes, T. and O'Neal, Q. (1995), "Exploring the factors associated with expert systems success", *MIS Quarterly*, March, pp. 83-106.

Zeleznikow, J. and Hunter, D. (1995), "Reasoning paradigms in legal decision support systems", *Artificial Intelligence Review*, Vol. 9, pp. 361-85.