Positive Impacts of an Intelligent System on Internal Control Problem Recognition

Chuleeporn Changchit
Management Sciences Dept.,
College of Business Adm.,
University of Iowa,
Iowa City, IA 52242
chuleeporn-
changchit@uiowa.edu

Clyde W. Holsapple
School of Management,
Carol M. Gatton College of Business Administration,
University of Kentucky,
Lexington, KY 40506-0034
cwhols@pop.uky.edu

Donald L. Madden
School of Accountancy,
Carol M. Gatton College of Business Administration,
University of Kentucky,
Lexington, KY 40506-0034
dmadden@pop.uky.edu

Abstract
Establishing and maintaining proper internal controls is an essential, ongoing, and complex process for today's organizations. Even though managers are ultimately responsible for ensuring internal control, they are typically not well versed in internal control principles. The literature contains no example of an intelligent system deployed to assist managers in recognizing potential problems in internal control systems. Yet, an intelligent system of this kind might prove to be highly beneficial. This study reports on the development of an intelligent system to assist managers, who do not have specific backgrounds in internal control concepts, evaluate potential weaknesses in internal control systems.

1. Introduction
Establishing and maintaining proper internal controls is an essential, ongoing, and complex process for today's organizations. Without the ability to ensure the accuracy and reliability of accounting information, a business organization could not survive in a competitive environment [4]. Internal Control issues are of significance to the entities in order to assure the accuracy, reliability and timeliness of the resulting financial reports [24]. Periodic checks by auditors evaluate a company's internal control system, probing it for weaknesses that need correction. Auditors use expert systems to assist in such work. However, it is managers, with their detailed day-to-day attention to company operations, who are in the best position to continually monitor for internal control problems. It is the obligation of management, not auditors, to establish and maintain a system of internal control [17]. There is one difficulty with this. Managers are typically not well versed in internal control principles, even though they are ultimately responsible for ensuring internal control. The literature contains no example of an expert system deployed to assist managers in recognizing potential problems in internal control systems. Yet, an intelligent system of this kind might prove to be highly beneficial.

Expert systems have emerged as an economically rewarding branch of artificial intelligence [21]. They also form a sub-group of decision support systems [6, 22]. Such system use captured human knowledge to solve problems that ordinarily require human experts [13]. In order to emulate a human expert’s behavior, the reasoning knowledge for a particular problem domain must be acquired and stored in the computer. Furthermore, there must be software that is able actively to process such knowledge in order to derive advice for a user.

Here, we report on the development of an expert system for supporting managers’ internal control evaluation. Because managers typically have neither specific background in internal control concepts nor hand-on experience with expert systems, such system might not provide positive outcomes. An experiment was, therefore, conducted to examine whether the use of such system could improve the effectiveness and efficiency of managers using it.

We begin with a background discussion to establish the context, motivation, and significance of this study. We then state the main objectives of this research in terms of hypotheses to be tested. Construction of the expert system for supporting managers’ internal control evaluation is briefly described, followed by the methodology used to conduct a laboratory experiment exploring the usefulness of this expert system, including subjects, independent and response variables, development of experimental materials, participants’
tasks, and statistical analysis procedures. The next section reports and discusses the research findings. The concluding discussion includes an identification of directions for future related research.

2. Background

The audit environment is a unique and highly complex decision-making environment [27]. Most audit problems are usually ill-structured or semi-structured. For instance, the problem of checking control weaknesses is a “nondeterministic polynomial” problem [18]. Such problems are often solved best by using the rules of thumb of experienced auditors. When such rules can be incorporated as reasoning knowledge in an expert system, there is a potential of developing such an expert system to aid in solving audit problems. Applications of expert system technology to auditing have become increasingly important. This may be due to factors such as increasing complexity of the auditing environment, growing specialization within the auditing profession, lower audit fees, and successful development of expert systems in other professions [5, 7, 12, 23, 25].

Among the expert system auditing applications, several have been developed for internal control evaluations, including ANSWER, CCR/36 Advisor, CFILE, C&L Control Risk Assessor, Internal Control Analyzer, and Internal Control Expert [3, 9, 10, 15, 16, 27, 30]. Internal control is defined in SAS No. 78 by the American Institute of Certified Public Accountants as a process - effected by an entity’s board of directors, management, and other personnel - designed to provide reasonable assurance regarding the achievement of objectives in the following categories: reliability of financial reporting, effectiveness and efficiency of operations, and compliance with applicable laws and regulation [1]. An accounting system supplemented by effective internal control systems can provide management with reasonable assurance that assets are safeguarded from unauthorized use or disposition and that financial records are sufficiently reliable to permit the preparation of financial reports [20].

It is noteworthy that the professional accounting literature clearly assigned responsibility for adopting sound accounting policies, maintaining an adequate internal control structure, and making fair representations in the financial statements to the management rather than the auditors [2, 8]. However, the focus to date of expert systems for internal control has been on helping auditors, not an organization’s management. These systems are generally designed to help auditors determine the extent of other tests they will perform in conducting an audit. If an auditor determines that the client’s internal controls are designed properly and are functioning as designed, he or she can reduce direct testing of account balances accordingly [14]. Although the literature does not report on expert systems that can assist managers evaluate internal control systems of their companies, there are several reasons that such systems deserve to be investigated:

- Establishment and supervision of internal control systems are the responsibility of management, not auditors [2, 8].
- Compared to auditors, organization’s managers should possess more knowledge about the company and thus have greater immediate insight into the operation of its internal control systems.
- The evaluation of internal control systems can be done more often if it is performed by the management because auditors are usually expensive and are not available upon request.

3. Objectives

The purpose of this research effort is to study the impact of using an expert system developed for helping managers who do not have specific backgrounds in internal control concepts evaluate internal control systems in their organizations. The six hypotheses presented in Table 1 are used to test the value of an expert system (ES) developed as a decision aid in detecting potential internal control weaknesses.

<table>
<thead>
<tr>
<th>Ha</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>Participants performing tasks with the expert system can better detect potential weaknesses of an internal control system than participants performing tasks without the expert system.</td>
</tr>
<tr>
<td>H2</td>
<td>Participants performing tasks with the expert system can accurately detect potential weaknesses of an internal control system more quickly than participants performing tasks without the expert system.</td>
</tr>
<tr>
<td>H3a</td>
<td>Participants performing tasks with the expert system perceive that tasks require less effort than participants performing tasks without the expert system perceive.</td>
</tr>
<tr>
<td>H3b</td>
<td>Participants performing tasks with the expert system are more satisfied with their accuracy than participants performing tasks without the expert system.</td>
</tr>
<tr>
<td>H3c</td>
<td>Participants performing tasks with the expert system are more satisfied with their speed than participants performing tasks without the expert system.</td>
</tr>
<tr>
<td>H3d</td>
<td>Participants performing tasks with the expert system perceive that performing task is more interesting than participants performing tasks without the expert system perceive.</td>
</tr>
</tbody>
</table>
This research concentrates only on the evaluation of controls typically found in the Sales and Collections cycle common to the merchandising industry. Knowledge acquisition for the expert system involved an auditor who is a partner of an international accounting firm. The resulting system closely emulates results of his cognitive processes. Thus, the system knowledge may be firm-specific or expert-specific.

4. Construction of an expert system for supporting managers’ internal control evaluations

In order to examine whether it is feasible to develop an expert system to help managers evaluate internal control systems in their organizations, an expert system was constructed that offers advice about weaknesses found in evaluating internal control systems in the sales and collection cycle of medium-size merchandising organizations. The prospective users are managers who are not experienced in auditing. The expert being modeled has more than ten years of experience in the area of internal control evaluation and demonstrated significant interest in the research project. The tool selected for developing the expert system was an integrated artificial intelligence environment called GURU [19]. The version 3.01 of GURU was used to implement this expert system.

Knowledge for the expert system was acquired via a six-month series of interviews with the expert. The expert was asked to identify all potential weaknesses that can occur in the sales and collection cycle of a medium-size merchandising organization. The result was a list of 126 internal control weaknesses. The expert was asked further to describe, in detail, the techniques and processes he used to discover each of these weaknesses in a client’s internal control system. The reasons for each decision-making heuristic were also acquired in an attempt to develop an expert system that would be able to emulate both the expert’s knowledge and his reasoning behavior.

Knowledge acquired from the expert was represented in sets of rules in the expert system. Upon processing these rules, an inference engine infers a potential internal control weakness recommendation. This recommendation identifies significant internal control weaknesses discovered in the situation being evaluated and indicates resulting exposures that could occur in a user’s organization. The expert was asked to critique the expertise captured in the rule sets in order to ensure the validity of the system. Validation of the system is often considered the cornerstone of expert system evaluation [3]. It is the process of analyzing the knowledge and decision-making capabilities of the expert system [26]. Test cases were generated from the manipulation of several cues for detecting the potential weaknesses in an internal control system over the sales and collection cycle. The expert was asked to evaluate each test case and detect its potential internal control weaknesses. Reasons for each potential weakness were also requested. Then, the prototype expert system was used to detect the potential weaknesses and offer reasons of such weaknesses as well. The results were then compared. It turned out that the expert and expert system identified similar weaknesses for each of the test cases. Where there were discrepancies, the expert reconsidered his responses and agreed that the expert system’s responses were indeed correct. Interestingly, this illustrates that an expert system can sometimes be useful even to an expert (e.g., to double check the expert’s reasoning).

5. Methodology

To examine the utility of the expert system, an experiment was conducted to assess the effectiveness and efficiency of managers using it. The experiment was conducted in a conference room and laboratory room of a college, providing an isolated and controlled environment for the study. The research design is a one-way treatment structure using a completely randomized design structure. Treatments had a grouping structure and fixed effects. Each participant was an experimental unit. There were 15 replications for each treatment.

5.1 Subjects

Three lists of potential subjects were obtained from three international accounting firms. From these lists, thirty practitioners voluntarily participated in this experiment. These participants served as novice internal control decision-makers. No participants had either specific background in internal control concepts or prior hands-on experience with an expert system. Tables 2 and 3 show the industries represented by the participants and the diversity of the participants’ roles, respectively.
Table 2. Industries of participants’ organizations

<table>
<thead>
<tr>
<th>Industries</th>
<th>No. of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consultant</td>
<td>5</td>
</tr>
<tr>
<td>Government</td>
<td>5</td>
</tr>
<tr>
<td>Legal</td>
<td>5</td>
</tr>
<tr>
<td>Health Care</td>
<td>4</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>3</td>
</tr>
<tr>
<td>Mining</td>
<td>3</td>
</tr>
<tr>
<td>Education</td>
<td>2</td>
</tr>
<tr>
<td>Real Estate</td>
<td>2</td>
</tr>
<tr>
<td>Transportation</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>30</strong></td>
</tr>
</tbody>
</table>

Table 3. Participants’ titles

<table>
<thead>
<tr>
<th>Title</th>
<th>No. of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Manager</td>
<td>5</td>
</tr>
<tr>
<td>Department Head</td>
<td>4</td>
</tr>
<tr>
<td>Administrative Manager</td>
<td>4</td>
</tr>
<tr>
<td>Supervisor</td>
<td>4</td>
</tr>
<tr>
<td>Assistant Manager</td>
<td>3</td>
</tr>
<tr>
<td>Finance and Accounting Manager</td>
<td>3</td>
</tr>
<tr>
<td>Director</td>
<td>2</td>
</tr>
<tr>
<td>Chief Accountant</td>
<td>1</td>
</tr>
<tr>
<td>Engineer</td>
<td>1</td>
</tr>
<tr>
<td>Lawyer</td>
<td>1</td>
</tr>
<tr>
<td>Marketing Manager</td>
<td>1</td>
</tr>
<tr>
<td>Tax Manager</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>30</strong></td>
</tr>
</tbody>
</table>

Participants had up to 10 years of experience with their organizations, with the average being 5.5 years. The percentage of their management duties ranged from 20 to 85 with an average of 52. The number of employees in their organizations ranged from 7 to 1,100 with an average of 175.

5.2. Independent variables

Two decision aid treatments were evaluated: No Decision aid (NDA) and an Expert System (ES). These two treatments were assigned to each group of the thirty participants (fifteen for each treatment) at the beginning of the study.

5.3. Response variables

Three response variables were measured as follows:
1. Effectiveness -- accuracy of decision-making was examined as a measure of the effectiveness of the system.
2. Efficiency -- the time used to accurately make decision was examined as a measure of the system’s efficiency.
3. Participants’ perceptions with the tasks -- post-experimental questionnaires were used to measure participants’ perceptions on the tasks in both groups.

5.4. Development of a list of internal control weaknesses

A list of internal control weaknesses was developed from a review of auditing texts, accounting texts, and inputs from accounting professors and experienced auditors. Two experienced auditors and two accounting professors were asked to evaluate each weakness and assign a score to each weakness in this list using a scale of 1 to 5 based on the degree of importance of each weakness (1-the least serious, 5-very serious). The average of the scores for each weakness (i.e., divided by four) was assigned as its importance. As described previously, importance scores for the weaknesses are used in assessing participants’ performances in internal control weakness recognition.

5.5. Development of case studies

Three case studies (A, B, and C) were generated from the manipulation of several cues for detecting the potential weaknesses in internal control systems. As with the test cases, these cues were obtained from a review of auditing texts, accounting texts, and input from accounting professors and experienced auditors. In each case, the scenario dealt with the adequacy of internal control over a company’s sales and collection cycle. These case studies also included the background information about the company and a partial organization chart that applied to each scenario. Each case study contained ten potential weaknesses in an internal control system. Three experienced auditors and three novice managers were asked to pilot test these case studies to ensure their similarity with respect to the degree of difficulty in detecting the potential internal control weaknesses. Ten MBA students were also asked to pretest these case studies in order to ensure their clarity. Revisions to these case studies were made based on the feedback they provided.

5.6. Development of post-experimental questionnaires

Questionnaires were developed to gather data about participants’ perceptions of their experiences in using the expert system, as well as their demographics. These questionnaires were pretested to ensure their clarity.
5.7 Experimental task

The experiment consisted of three sessions. Prior to the experiment, the three case studies were randomly assigned to the sessions, yielding the result of Case A being assigned to Session I, Case B to Session III, and Case C to Session II. The maximum time allowed for each session was two hours. Thirty participants were randomly divided into two groups: the NDA (No Decision Aid) Group and the ES (Expert System) Group. Each participant played the role of a manager trying to detect potential problems in the internal control systems described in a series of case studies.

In Session I, participants in both groups performed the case study (Case A) without any decision aid. Because no participant has any specific background in internal control concepts, a list of 126 possible internal control weaknesses was provided in this session. The time taken to make a decision about which weakness existed was measured and the accuracy score was calculated. In Sessions II and III, participants in the NDA Group performed the second and third case studies (Case C and Case B, respectively) without the expert system while participants in the ES Group were able to use the expert system as a decision aid in detecting internal control weaknesses in the assigned case studies. For each session; the time taken to make a decision was measured and the accuracy score was calculated. At the end of each of Sessions I, II, and III, the questionnaires were given to measure participants’ perceptions about the tasks they performed.

5.8. Task incentive

A performance-based reward mechanism is crucial for laboratory experiments because it creates a more realistic environment and gives subjects incentives to perform [29]. A performance-based reward system was used here to induce optimal participant behavior and reduce problems due to guessing. Participants were told that their performances across the series of cases would be used to calculate two scores: an accuracy score and a time per accuracy score. They were informed that prizes would be given to those participants attaining the highest accuracy scores and to those attaining the best time per accuracy scores.

6. Experimental findings

The experimental data were analyzed as a completely randomized design with a one-way treatment structure using the Analysis of Variance (ANOVA) technique. Treatments had a grouping structure and fixed effects. Each participant was an experimental unit. There were 15 replications for each treatment. The F-test was used to test hypotheses about group means for each response variable: (i) accuracy score (H1), (ii) time per accuracy score (H2), and (iii) participants’ perceptions (H3a, H3b, H3c, and H3d). In order to test the hypotheses, we compared the performances of each participant in Session III between the NDA Group and the ES Group. There were two major reasons why the performances in Session III were used instead of Session II. First, because no participants have hands-on experience in using the expert system, the accuracy scores in Session II might not reflect their actual competency. These scores might be reduced due to unfamiliarity with using the expert system. Second, the time measured in Session II could not be considered as the actual decision time because part of it might be considered as being used to learn about interacting with the expert system. Therefore, we felt that it was most appropriate to compare participants’ performances in Session III.

6.1. Calculation of accuracy score and time per accuracy score

Accuracy of decision-making was examined as a measure of the system’s effectiveness [11, 28]. Each weakness has a certain number of points associated with it, reflecting its importance. Some weaknesses have more points (i.e., more importance) than other weaknesses. In arriving at an accuracy score, all points related to correctly identified weaknesses were added. In order to prevent the subjects from trying to detect weaknesses by guessing, they were informed at the beginning of the experiment that one-third of all points for inaccurately identified weaknesses would be subtracted as the penalty for guessing. Non-response for a potential weakness results in neither addition nor subtraction.

The time used to make correct decisions (i.e., time per accuracy score) was examined as a measure of the system’s efficiency. The time per accuracy score was computed as follows:

\[
\text{Time Used to Perform Task} \quad \text{Accuracy Score}
\]

The following examples illustrate this calculation.

Example 1:

Mr. A correctly detects 5 weaknesses and incorrectly identifies 5 weaknesses. The time used to do the case is 40 minutes. Assuming that each weakness has 3 points associated with it, the accuracy score is 10 = (5 x 3) - (5 x 3)/3. The time per accuracy score is 4 = (40/10) minutes per accuracy score.
Example 2:
Mr. B detects 4 correct weaknesses, 3 incorrect weaknesses, and 3 non-responses. The time used to do the case is 40 minutes. Assuming that each weakness has 3 points associated with it, the accuracy score is $9 = ((4 \times 3) - (3 \times 3))/3$. The time per accuracy score is $4.4 = (40/9)$ minutes per accuracy score.

6.2. Accuracy of participants’ responses

Because participants were randomly assigned to each group, the ANOVA was performed on the first session in order to test for differences in accuracy score between the groups. At an alpha level of 0.05, no significant differences were found between the groups (p-value = 0.27).

Hypothesis H1 was designed to test the accuracy of decision making as a measurement of the system’s effectiveness. It was operationalized as follows:

$H_{10}$: $\mu_{DAC_{NDA \ Group}} \geq \mu_{DAC_{ES \ Group}}$

$H_{1a}$: $\mu_{DAC_{NDA \ Group}} < \mu_{DAC_{ES \ Group}}$

$\mu_{DAC} = \text{Mean of the Accuracy Score in Session III} = \frac{\sum_{i=1}^{15} ACs_{3i}}{15}$

$ACs_{3i} = \text{the accuracy score of participant i in Session III}$

We predicted that, on the average, participants in the ES Group could better detect potential weaknesses of an internal control system than participants in the NDA Group. This prediction is supported if the null hypothesis is statistically rejected. An F-Test was employed to test the hypothesis.

6.3. Time of participants’ responses

Because participants were randomly assigned to each group, the ANOVA was performed on the first session in order to test for differences in the time per accuracy score between the groups. Even though no significant differences were found between the groups at alpha level of 0.05, the differences were found at alpha level of 0.1 (p-value = 0.08). This indicates that participants randomly assigned to the NDA Group could detect internal control weaknesses more quickly than participants assigned to the ES Group. In order to reduce the effect of such bias between the groups, an F-Test was employed to examine the differences of participants’ performances in Session III minus Session I instead of on Session III only.

Hypothesis H2 was designed to test the time used by participants in accurately detecting the weaknesses in internal control systems as a measurement of the system’s efficiency. Thus, the hypothesis H2 was operationalized as follows:

$H_{20}$: $\mu_{DT_{NDA \ Group}} \leq \mu_{DT_{ES \ Group}}$

$H_{2a}$: $\mu_{DT_{NDA \ Group}} > \mu_{DT_{ES \ Group}}$

$\mu_{DT} = \text{Mean of the Deviation of the Time per Accuracy Score in Session III from the Time per Accuracy Score in Session I} = \frac{\sum_{i=1}^{15} (Ts_{3i}/ACs_{3i} - Ts_{1i}/ACs_{1i})}{15}$

$Ts_{3i} = \text{the time used by participant i in Session III}$

$ACs_{3i} = \text{the accuracy score of participant i in Session III}$

$Ts_{1i} = \text{the time used by participant i in Session I}$

$ACs_{1i} = \text{the accuracy score of participant i in Session I}$

We predicted that, on the average, participants in the ES Group could detect internal control weaknesses more quickly than participants in the NDA Group. This prediction is supported if the null hypothesis is statistically rejected.

6.4. Participants’ perceptions with the task

Hypothesis H3a, H3b, H3c, and H3d were designed to test whether there were differences in participants’ perceptions between the ES Group and the NDA Group. These hypotheses were operationalized as follows:

$H_{3a0}$: $\mu_{IIa_{NDA \ Group}} \geq \mu_{IIa_{ES \ Group}}$

$H_{3a1}$: $\mu_{IIa_{NDA \ Group}} < \mu_{IIa_{ES \ Group}}$

$H_{3b0}$: $\mu_{IIb_{NDA \ Group}} \geq \mu_{IIb_{ES \ Group}}$

$H_{3b1}$: $\mu_{IIb_{NDA \ Group}} < \mu_{IIb_{ES \ Group}}$

$H_{3c0}$: $\mu_{IIc_{NDA \ Group}} \geq \mu_{IIc_{ES \ Group}}$

$H_{3c1}$: $\mu_{IIc_{NDA \ Group}} < \mu_{IIc_{ES \ Group}}$

$H_{3d0}$: $\mu_{III_{NDA \ Group}} \geq \mu_{III_{ES \ Group}}$

$H_{3d1}$: $\mu_{III_{NDA \ Group}} < \mu_{III_{ES \ Group}}$
\(\mu_{\text{IIa}, \text{NDA Group}} = \) the mean of responses in Session III to the question “On a scale of 1 (very difficult) to 7 (very easy), how difficult was it to do the case study?” (NDA Group).

\(\mu_{\text{IIa}, \text{ES Group}} = \) the mean of responses in Session III to the question “On a scale of 1 (very difficult) to 7 (very easy), how difficult was it to do the case study?” (ES Group).

\(\mu_{\text{IIIa}, \text{NDA Group}} = \) the mean of responses in Session III to the question “On a scale of 1 (very unsatisfactory) to 7 (very satisfactory), how satisfied were you with your accuracy in answering the case study?” (NDA Group).

\(\mu_{\text{IIIa}, \text{ES Group}} = \) the mean of responses in Session III to the question “On a scale of 1 (very unsatisfactory) to 7 (very satisfactory), how satisfied were you with your accuracy in answering the case study?” (ES Group).

\(\mu_{\text{IIIb}, \text{NDA Group}} = \) the mean of responses in Session III to the question “On a scale of 1 (very unsatisfactory) to 7 (very satisfactory), how satisfied were you with your speed in answering the case study?” (NDA Group).

\(\mu_{\text{IIIb}, \text{ES Group}} = \) the mean of responses in Session III to the question “On a scale of 1 (very unsatisfactory) to 7 (very satisfactory), how satisfied were you with your speed in answering the case study?” (ES Group).

\(\mu_{\text{IIIc}, \text{NDA Group}} = \) the mean of responses in Session III to the question “On a scale of 1 (very unsatisfactory) to 7 (very satisfactory), how satisfied were you with your speed in answering the case study?” (NDA Group).

\(\mu_{\text{IIIc}, \text{ES Group}} = \) the mean of responses in Session III to the question “On a scale of 1 (very unsatisfactory) to 7 (very satisfactory), how satisfied were you with your speed in answering the case study?” (ES Group).

\(\mu_{\text{IIId}, \text{NDA Group}} = \) the mean of responses in Session III to the question “On a scale of 1 (very boring) to 7 (very interesting), how interesting was the task you performed?” (NDA Group).

\(\mu_{\text{IIId}, \text{ES Group}} = \) the mean of responses in Session III to the question “On a scale of 1 (very boring) to 7 (very interesting), how interesting was the task you performed?” (ES Group).

We predicted that, on the average, participants in the ES Group would:

- perceive that the task requires less effort than participants in the NDA Group would perceive.
- be more satisfied with their accuracy than participants in the NDA Group.
- be more satisfied with their speed than participants in the NDA Group.
- perceive that the task is more interesting than participants in the NDA Group perceive.

These predictions are confirmed if the null hypotheses are statistically rejected. F Tests were employed to test whether sample means of participants’ perceptions were distinct for hypotheses H3a, H3b, H3c, and H3d.

### 6.5. Results and discussion

F-tests were employed to test hypotheses about the group means for participants in the ES Group versus participants in the NDA Group for each response variable: (i) accuracy score (H1), (ii) time per accuracy score (H2), and (iii) participants’ perceptions (H3a, H3b, H3c, and H3d). Tables 4, 5, and 6 present the results of this analysis.

Hypothesis H1 examined whether there was an impact on participants’ accuracy in detecting potential weaknesses in internal control systems when using the expert system as a decision aid. The major finding in testing this hypothesis is summarized in Table 4. The results show that participants in the ES Group could detect internal control weaknesses significantly more accurately than participants in the NDA Group (\(\mu_{\text{DAC}_{\text{NDA Group}} = 3.95 \text{ vs. } \mu_{\text{DAC}_{\text{ES Group}} = 25.506}\)). The F-Test confirms that, at alpha level of 0.01, there was a significant difference in participants’ accuracy in detecting an internal control weakness between participants in the ES Group versus participants in the NDA Group (p-value = 4.18E-08).
Table 4. Analysis of variance (AC score)

<table>
<thead>
<tr>
<th>Source</th>
<th>Size</th>
<th>df</th>
<th>SS</th>
<th>F-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (ES)</td>
<td>30</td>
<td>1</td>
<td>3484.96</td>
<td>55.3848</td>
<td>4.18E-08*</td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td>28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Summary

<table>
<thead>
<tr>
<th>Source</th>
<th>Sample Size</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Decision Aid (NDA Group)</td>
<td>15</td>
<td>3.95</td>
</tr>
<tr>
<td>Expert System (ES Group)</td>
<td>15</td>
<td>25.506</td>
</tr>
</tbody>
</table>

* Significance of at least $\alpha = 0.01$.

Hypothesis H2 examined whether there was an impact on participants’ speed in accurately detecting potential weaknesses in internal control systems when using the expert system as a decision aid. The major finding in testing this hypothesis is summarized in Table 5. In the NDA Group, the result shows the positive deviation of time per accuracy score in Session III from Session I of 3.216, while in the NDA Group, the result shows the negative deviation of time per accuracy score in Session III from Session I of 21.865. This finding indicates that participants in the ES Group had significantly greater speed (i.e., the improvement in their performances in Session III vs. Session I) in accurately detecting internal control weaknesses than participants in the NDA Group. The F-Test result confirms that, at alpha level of 0.1, there was a significant difference in the time that participants in the ES Group used to accurately detect internal control weaknesses with the support of the expert system versus the time used by the participants in the NDA Group without any decision aid (p-value = 0.091413).

Table 5. Analysis of variance (Time per AC Score)

<table>
<thead>
<tr>
<th>Source</th>
<th>Size</th>
<th>df</th>
<th>SS</th>
<th>F-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Error</td>
<td>30</td>
<td>1</td>
<td>26979.35</td>
<td>0.091413*</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>28</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Summary

<table>
<thead>
<tr>
<th>Deviation of Time per AC Score in Session III from Session I (NDA Group)</th>
<th>Size</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
<td>3.216</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deviation of Time per AC Score in Session III from Session I (ES Group)</th>
<th>Size</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
<td>-21.865</td>
</tr>
</tbody>
</table>

* Significance of at least $\alpha = 0.1$.

Hypotheses H3a, H3b, H3c, and H3d examined the difference in participants’ perceptions between the ES Group and the NDA Group. The major findings for testing these hypotheses are summarized in Table 6.

Table 6. Results of testing hypotheses H3a, H3b, H3c, and H3d

<table>
<thead>
<tr>
<th>Participants’ responses</th>
<th>NDA</th>
<th>ES</th>
<th>F</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means of responses in S. III to the question “On a scale of 1 (very difficult) to 7 (very easy), how difficult was it to do the case study?” (H3a)</td>
<td>2.13</td>
<td>4.13</td>
<td>12.83</td>
<td>0.0013*</td>
</tr>
<tr>
<td>Means of responses in S. III to the question “On a scale of 1 (very unsatisfactory) to 7 (very satisfactory), how satisfied were you with your AC in answering the case study?” (H3b)</td>
<td>2.67</td>
<td>5.07</td>
<td>23.14</td>
<td>4.65E-05*</td>
</tr>
<tr>
<td>Means of response in S. III to the question “On a scale of 1 (very unsatisfactory) to 7 (very satisfactory), how satisfied were you with your speed in answering the case study?” (H3c)</td>
<td>2.40</td>
<td>4.60</td>
<td>19.11</td>
<td>0.0002*</td>
</tr>
<tr>
<td>Means of response in S. III to the question “On a scale of 1 (very boring) to 7 (very interesting), how interesting was the task you performed?” (H3d)</td>
<td>2.93</td>
<td>5.27</td>
<td>20.47</td>
<td>0.0001*</td>
</tr>
</tbody>
</table>

* Significance of at least $\alpha = 0.01$. 

* Significance of at least $\alpha = 0.01$. 

Proceedings of the 32nd Hawaii International Conference on System Sciences - 1999

0-7695-0001-3/99$10.00 (c) 1999 IEEE
The participants’ answers to the post-experiment questionnaires reveal that participants in the ES Group had better perceptions of the task than participants in the NDA Group. The results show the attitudes of participants in both group toward the task as follows:

- Participants in the ES Group perceived that the task was easier than participants in the NDA Group perceived (µES Group = 2.13 vs. µNDA Group = 2.93, p-value = 0.0001)
- Participants in the ES Group were more satisfied with their perceived accuracy than participants in the NDA Group (µES Group = 2.67 vs. µNDA Group = 5.07, p-value = 4.65E-06)
- Participants in the ES Group were more satisfied with their perceived speed than participants in the NDA Group (µES Group = 2.40 vs. µNDA Group = 2.13, p-value = 0.0002)
- Participants in the ES Group perceived that the task was more interesting than participants in the NDA Group perceived (µES Group = 2.93 vs. µNDA Group = 5.27, p-value = 0.0001).

7. Conclusion and directions of future research

The major finding of this study is that participants in the ES Group significantly outperformed participants in the NDA Group on a variety of dimensions. This indicates that it is feasible to build an expert system that can benefit managers, who are not auditing experts, in detecting potential internal control weaknesses. The system can help managers detect the weaknesses in their organizations’ internal control systems more effectively and more efficiently. In addition, the results show that the system can help improve the users’ perceptions of internal control evaluation tasks. On the average, the users reveal that they are more satisfied with their accuracy and speed when using the expert system as the decision aid. They also perceive that the task was easier and more interesting when allowed to perform the tasks with the expert system.

The major finding of this research suggests that installing such a system may help managers maintain an effective internal control system, thus providing more reliable accounting data and better safeguarding assets. It could let organizations save time and money by allowing weaknesses in internal control systems to be detected and solved more quickly. Such a system can also be beneficial for auditing firms by facilitating more reliable internal control in client firms, thereby reducing planned detection risk (i.e., less work and time).

This research lays the groundwork for several avenues of future research. Researchers could incorporate additional transaction cycles, industries, or other auditing functions into the expert system, and then study it in a manner similar to what has been reported here. Future research could investigate results of using the expert system over longer period of time (e.g., conducting additional sessions a week or month after the first three sessions). Another research avenue is to investigate the use of such an expert system as a training tool, instead of as a direct decision aid. Future research could develop additional expert systems by acquiring the expertise from multiple auditors and then conduct experiments similar to the one reported here. Finally, researchers could investigate the feasibility of integrating this kind of expert system with a company’s databases.

8. References


