Pergamon

International Journal of Accounting Information Systems xx (2003) xxx–xxx

INTERNATIONAL JOURNAL OF ACCOUNTING INFORMATION SYSTEMS

Auditor environmental assessments

Daniel E. O'Leary*

Leventhal School of Accounting, 3660 Trousdale Parkway, University of Southern California, Los Angeles, CA 90089-0441, USA

Received 15 May 2002; received in revised form 5 September 2003; accepted 8 September 2003

Keywords: Rule-based systems; Knowledge-based systems; Environmental assessments

1. Introduction

During the late 1980s and into the early 1990s, classic rule-based systems (also referred to as knowledge-based systems) appeared to have enormous potential for accounting and auditing problems (e.g., Kneale, 1988). The concept behind these classic expert systems was simple. Solicit knowledge from an expert or experts so that others may solve problems using that same knowledge. Embed that knowledge into a computer program in order to leverage the expert's knowledge. Then have users (other than the experts) be responsible for gathering data for the system (i.e., making "environmental assessments") and inputting that data to the system. Finally, have the system provide answers and have the users implement them. Such systems would allow delegating decision making down to the personnel who are responsible for gathering the data (Kelly et al., 1987; Trewin, 1995; Molay, 2002). As a result, the person using the system would be different than the person on which the expertise is based.

Unfortunately, there has been limited adoption of classic expert systems in accounting and auditing (e.g., Rennie and Gibbins, 1993). Why haven't expert systems lived up to their promise? There are multiple potential explanations for the lack of their success. One explanation is that it is difficult or impossible to truly capture human expertise in a computer program. However, another explanation is that somehow users are doing something that is inhibiting the ability of the program to help solve problems. For example, either the users

* Tel.: +1-213-740-4856; fax: +1-213-747-2815.
E-mail address: oleary@usc.edu (D.E. O'Leary).

1467-0895/$ – see front matter
doi:10.1016/j.accinf.2003.09.001

may not be using the system correctly or there may be something going wrong with the data-gathering and input processes.

Structurally, within a system knowledge base, different environmental assessments by the user lead the system to different parts of its knowledge and provide different recommended solutions. Unfortunately, in auditing and accounting judgments, there often are multiple potential environmental assessments to the same situations. As an illustration, in an example given below, the knowledge-based system asks their users to determine if documentation is "adequate," or if software is "complete and sufficient." Those terms "adequate," "complete," and "sufficient" are open to judgment because there is no single interpretation for all situations. This can cause different users to make *different environmental assessments, under the same conditions*, leading the knowledge-based system to make *different recommendations to different users*.

The purpose of this paper is to examine the extent to which environmental assessments, can vary based on who is doing the assessment. This paper provides evidence that different users, in particular (1) partners/managers and (2) staff (others), each generate different distributions of environmental assessments for situations representative of concepts used by a knowledge-based system. Based on the Libby and Luft (1993) model we would expect that differential experience and knowledge between those two groups would lead to differences in environmental assessments, which are substantiated by the findings.

I find that there are statistically significantly smaller variances for partner/manager distributions of environmental assessments. I find that staff distributions of environmental assessments are likely to cover the entire range of possible assessments for a given situation, whereas the partner/manager group distribution takes a smaller subset. I find that the distributions of assessments by partners and managers for two of four sets of situations are statistically different than the staff. Furthermore, I find that for one set of questions, the means between the partners/managers and staff are significantly different.

These findings are critical because knowledge-based systems are typically based on expert knowledge but used by the other nonexperts in the firm. For example, although Expertax was developed based on experts, those gathering the information were often supervisors and staff (Trewin, 1995). As a result, there is a potential "mismatch" between the views of those on whom the system is based and used. Solutions that would be derived by the two groups of users, for the same situations, are different.

Differential environmental assessments impacts more than just expert systems. Because expert systems are based on audit work, environmental assessments are part of auditor activity. As a result, environmental assessment differences also affect other venues, whether paper or computer based.

In addition, the purpose of this paper also is to understand what can be driving such different assessments between the two groups. In particular, for those environmental assessments where simple rules (explicit knowledge) exist for facilitating the assessments, there are no statistically significant differences in the means, between the two groups. However, for those with no such algorithms (tacit knowledge) there are statistically significant differences in the means, between the two groups. This suggests that tacit knowledge differences between the two groups are driving the findings.

Finally, I speculate as to what is the impact of differential assessments on the use of expert systems. If use of a system provides dissimilar results, in similar situations, that lack of consistency could ultimately drive out the use of such systems.

This paper proceeds as follows. Section 2 discusses the widespread nature of environmental assessments that appear in virtually all expert systems. Section 3 reviews some of the auditor decision-making literature and couches this problem in that literature. Section 4 briefly reviews the impact of organizational position on assessments and establishes the expectations for the relationship between environmental assessments of partners and managers, in contrast to the environmental assessments of those lower in the organization hierarchy. Section 5 briefly reviews some of the rules from one expert system that require environment assessment. Section 6 outlines the test instrument, the methodology used and statistical analysis used to investigate the results. The instruments focused on the judgments discussed in Section 5. Section 7 presents and discusses the results. Section 8 investigates some alternatives to user environmental assessments and their limitations. Section 9 briefly summarizes the paper.

2. Knowledge-based systems require environmental assessments

Auditors make environmental assessments as part of general audit activity. Typically, staff and senior auditors gather audit information and managers and partners review that information. As a result, virtually all-classic knowledge-based systems discussed in the literature exhibit the need for users to make environmental assessments. I review three accounting and audit systems, but the need to have their users make environmental assessments is not unique to these systems. In addition, I review a rule from a so-called emerging "business rule" system, illustrating the potential sets of issues goes beyond expert systems to include business rule systems.

While describing a knowledge-based system designed to support decision making of audits of advanced computer systems, Hansen and Messier (1986, p. 374) use the following "rule" (the most frequently used approach to represent knowledge) that requires the user to provide environmental assessments of completeness and sufficiency of controls, adequacy of recovery measures and documentation and other issues.

If

(1) Message control software is *complete* and *sufficient*, and
(2) Recovery measures are *adequate*, and
(3) *Adequate* documentation is generated to form a *complete* audit trail

Then there is *strong suggestive* evidence (.8) that controls over data loss are *adequate*.

Another auditing system was designed to assist the auditor in determining the going concern judgment. For example, Dillard and Mutchler (1987) used the following frame (another form of representing knowledge in knowledge-based systems) to capture some of the knowledge used by the system. This frame requires that the user assess the realism of the

estimate of productive capacity and the consistency of the capacity with target objective. In addition, it asks the user to rate that consistency over a 5-point scale.

Frame:
DIRS Node: production capacity, Next: Productive ability

Given a *realistic estimate* of the entity's productive capacity, is it *consistent* with the stated target market objectives of the firm? Rate the consistency on a 5-point scale with 1 being *very inconsistent* and 5 being *very consistent*.

A third auditing system was designed to assist in materiality assessments. Steinbart (1985, p. 60) and Steinbart (1987) used the following rule where the users were asked to assess liquidity and solvency.

If

1) the Client is a public entity, and
2) there is no *concern* about *liquidity* or *solvency*

Then it can be assumed that the users of the financial statement are primarily interested in the results of current operations.

Environmental assessment of concepts is typical of knowledge-based systems. These three systems illustrate some of the kinds of assessments that are expected to be made by accounting and auditing knowledge-based system users.

In addition, the recent emergence of "business rule" systems, such as Blaze Advisor, usher in a new generation of rule-based systems. For example, the following rule was provided by Molay (2002).

T1.1	IF	the *EnterpriseValue* of Customer is GOLD
T1.2	AND	the *Loyalty of Customer* is *LOW*
T1.3	THEN	the PersonalizationAction is *Call*

The terms "EnterpriseValue" and "Loyalty of the Customer" could all be based on environmental assessments.

3. Auditor decision making

Libby and Luft (1993) suggest that audit judgments are influenced by multiple factors, including individual ability, knowledge, accounting support environment factors and auditor motivation.

Ability relates to an individual's capacity to encode, retrieve and analyze information. As described by Libby and Luft (1993, p.428), "... abilities tend not to be specific to accounting settings."

Knowledge provides the individual with the information necessary to perform a particular auditing task or tasks. Typically, explicit knowledge is categorized as procedural and

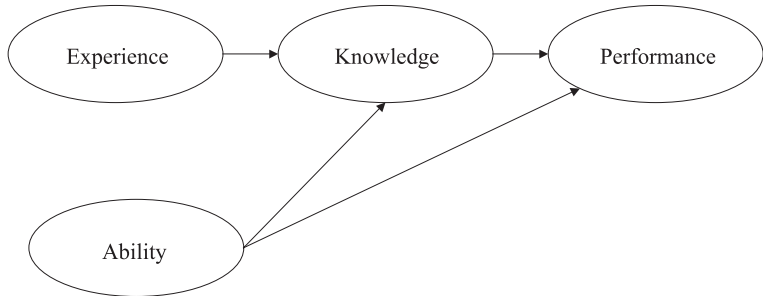

Fig. 1. Antecedents and consequences of knowledge. Source: Libby and Luft (1993, p. 433).

declarative. Procedural knowledge relates to how to perform a task, while declarative knowledge relates to the facts used to solve a particular problem. In contrast to explicit knowledge, tacit knowledge cannot be written down and typically is gathered through apprenticeship activities. For example, partners and managers would be expected to have different tacit knowledge than the staff because of their different experiences and apprenticeships.

The relationships between ability, experience, knowledge and performance are summarized in Fig. 1.

Accounting support environment factors include concern for authoritative guidance, technology support and accountability relationships. Support garnered from an expert system would fall into this category.

An individual's willingness to put forth effort captures the motivation factors. Such motivation is affected by organizational relationships, incentives and the desire to succeed.

In accounting and auditing, many judgment tasks are performed within the context of accounting standards, accounting firm policies and procedures and computer-based systems designed to support decision making. Since they are judgment tasks, these characteristics of accounting and auditing environments can impact judgment performance, typically by interacting with the factors in Fig. 1. Libby and Luft (1993) identify two principle classes of interactions. First, the environmental factor changes the task requirements (e.g., a spreadsheet eases computations). Second, the environmental factor changes the amount or allocation of effort, and changes motivation.

4. Impact of organization position on assessments

This section applies the audit decision-making model to the issue of environment assessments. This research does not directly examine issues of ability or motivation.

First, managers and partners are likely to have different explicit and tacit "knowledge" than staff.

- Managers and partners, and staff, all participate in continuing education programs designed to develop a common shared explicit knowledge. All are required to participate in continuing

education on a yearly basis. However, because they have been part of public accounting longer, partners and managers probably get exposure to more explicit knowledge. In terms of the Libby and Luft (1993), this relates to the experience and knowledge link.

- Managers and partners typically have different job responsibilities than the staff. Partners and managers direct the audit team. As a result, managers and partners are more likely to have greater explicit procedural knowledge than staff. These differences in job responsibility can impact both the experience and knowledge components of the Libby and Luft (1993) model.
- Managers and partners typically go through an "apprenticeship" as part of their job progression from staff to senior to manager and partner, working with other managers and partners in their everyday job activities. Apprenticeships generate tacit knowledge in their participants (e.g., Sveiby, 1999). For example, it may be that partners and managers in public accounting firms learn to disambiguate environmental assessments in a manner similar to other partners or managers. Apprenticeships would provide both differential experience and knowledge.
- Both partners and managers' positions have much more experience, as seen later in the paper. That experience is likely to provide them with both explicit and tacit knowledge. Again, Libby and Luft (1993) suggest that those differences in experience and knowledge will impact performance.

As a result, I expect that there are differences between the distribution of environmental assessments of the partners and managers, in contrast to staff. In particular, I would expect that partner/manager environmental assessments would vary less than staff assessments, over a smaller range of assessments. Finally, for those judgments where tacit knowledge is required, I would expect different assessments between the two groups.

Second, if there is a difference in the environmental assessments then the quality of the information generated by the computer support systems would vary, based on user group. In particular, different users will provide different assessments of the environment to the system,

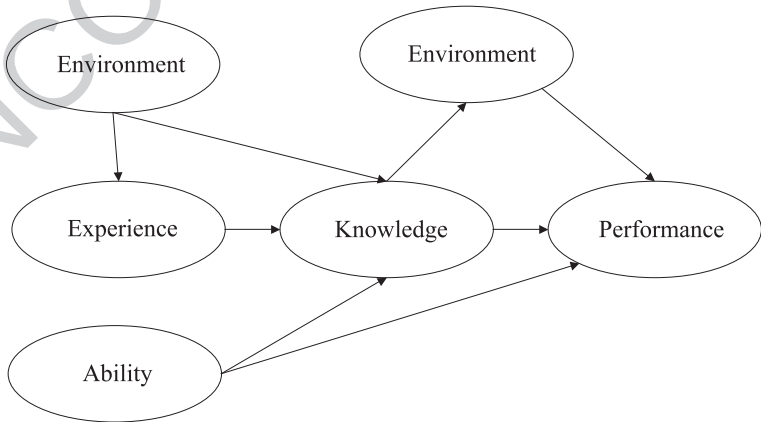

Fig. 2. Antecedents and consequences of knowledge. Source: Libby and Luft (1993, p. 433).

thus leading users to make different judgments. Accordingly, the computer tools would not be user neutral ("user not neutral hypothesis"): In general, users that have knowledge that provides a better environmental assessment, will get solutions that are more appropriate to the actual situation. Note that this is not the same as saying that so-called technology strong "power users" will generate better solutions. Instead, the organizational level group (partner and manager vs. staff) would drive the findings. Settings that had a manager or partner use the system could generate differential results from the system, when compared to those settings where staff used the system.

As seen in Fig. 2, the environment (hierarchical setting, accountability relationships, accounting regulations, computer programs, etc.) influences experience and knowledge. On the other hand, experience and knowledge influence the quality of the judgments generated by the individual's use of the computer programs. Different users will generate different quality results using the same programs. Use of computers is not user neutral. However, this paper does not examine this issue empirically.

5. Sample system: environmental assessment in "AUDITOR"

The empirical research discussed in this paper used rules from one of the first knowledge-based systems: AUDITOR (Dungan, 1983, p.1; Dungan and Chandler, 1983, 1985). AUDITOR (Dungan, 1983, p.1) was designed to be "... an operating model of one of the judgment processes carried out by professional auditors." AUDITOR was chosen because it is like many of the classic expert systems: its use requires multiple environmental assessments. In addition, the complete rule base is available (Dungan, 1983). The system was built with 36 rules. As far as the author knows there has been no change to the system since it was developed and is not currently used in practice.

A number of rules were found in AUDITOR that required situation assessments be provided by the user. For purposes of this paper, four rules from AUDITOR were selected for further investigation (Dungan, 1983, p. 184–189),

> The average age of the uncollected portions of this account is increasing.
>
> Collection effort being applied by your client to collect this account is less rigorous than is desirable.
>
> While small payments are being received on this account, the outstanding delinquent balance is growing larger.

In each case I have underlined the primary words that require environmental assessment. Although the concern is with providing assessments for four words ["increasing (uncollected balance)," "rigorous (collection effort)," "larger (balance)" and "materiality"], these rules contain other assessments that must be made, including issues such as "desirable" and "delinquent charge."

6. Empirical study

In order to test the extent to which environmental assessments differences are found in the same situations for different sets of users, an instrument was developed, tested and given to a sample of partners, managers and staff of an at the time "Big 5" accounting firm. The results were then analyzed using a range of statistical tests.

6.1. Subjects

A two-stage approach was used to study the ambiguity in the rules in AUDITOR. Accounting students were used as subjects for prepilot and pilot studies, and employees of a "Big 5" accounting firm were used as subjects with a slightly revised version of the instrument.

A prepilot study was used using two undergraduate students. Then, minor changes were made and the instrument was administered to 42 undergraduate accounting majors and 17 graduate accounting majors.

Using the feedback from the implementation of the pilot study and additional input from a member of the auditing faculty, some minor changes were made in the instrument. As a result, a slightly revised instrument was sent to 90 auditors at a "Big 5" auditing firm, of which 78 responded. The instrument was mailed to the firm's principle recruiter in the local office, who was responsible for distributing and collecting the instruments. The author requested that the instrument be given to auditors of different ranks. Accordingly, the sample included the entire range of audit personnel staff, senior, supervisor, managers and partners. Descriptive data are summarized below.

T2.1 Descriptive Data

T2.2 Level	Number	Average experience	Number with CPA
T2.3 Partner	6	13.75	6
T2.4 Manager	7	6.71	7
T2.5 Supervisor	5	3.70	3
T2.6 Senior	14	2.86	5
T2.7 Staff	37	0.98	1
T2.8 No data	7	1.42	1
T2.9 Total	76		23

6.2. Questions

Five rules from AUDITOR, discussed in the previous section, formed the basis of the instrument. A set of two or three questions was generated for each environmental assessment. Two question sets addressed environmental assessments that auditors must make as part of making delinquent account assessments, such as determining if the average age of an account is increasing or decreasing or staying the same (Table 1) or if a delinquent balance is growing smaller, larger or staying the same (Table 2). In order to assess the collectability of an

Table 1
Questions about "increasing"

Experts indicate that whether the average age of an customer's uncollected account is increasing is important in assessing the ultimate collectability of the account.							
a. For client A, in Month 1 the average age is 47 days, month 2 the average age is 42 and month 3 the average age is 45. Indicate whether the account average age is							
Increasing			Staying the Same			Decreasing	
1	2	3		4	5	6	7
b. For client B, in Month 1 the average age is 50 days, month 2 the average age is 54 and month 3 the average age is 49. Indicate whether the account average age is							
Increasing			Staying the Same			Decreasing	
1	2	3		4	5	6	7
c. For client C, in Month 1 the average age is 37.5 days, month 2 the average age is 39 and month 3 the average age is 37. Indicate whether the account average age is							
Increasing			Staying the Same			Decreasing	
1	2	3		4	5	6	7

account, auditors also have to make assessments regarding whether a balance was "large" (Table 3) or "material" (Table 4).

In order to test the similarity of environmental assessments, situations were constructed where there were multiple feasible assessments. In particular, in each environmental assessment subjects were asked to determine the extent to which they would assess the situation using a 7-point scale for each question. Subjects were asked to choose where on that scale they would rate their answer to the particular situation. A 7-point scale was chosen so that subjects could indicate the extent to which the relationship was true or false. Such a selection process is typical of knowledge-based systems, as noted earlier in the paper, where knowledge-based systems typically require a choice from a range of two or more alternatives.

Table 2
Questions on "rigorous"

Experts indicate that an important part of ensuring the collectability of those items whose collectability is questionable, is the rigor of the collection effort.						
a. Client H has an outstanding balance of $50,000. The collection efforts include having sent a letter to notify client H and a follow-up phone call to the accounts payable department of client H. The accounts have not yet been collected. How rigorous was the collection effort?						
Not rigorous						Rigorous
1	2	3	4	5	6	7
b. Client G has an outstanding balance of $50,000. The collection efforts include having sent a letter to notify client G, a follow-up phone call to the accounts payable department of G and the use of a collection agency. The account has not yet been collected. How rigorous was the collection effort?						
Not rigorous						Rigorous
1	2	3	4	5	6	7
c. Client J has an outstanding balance of $100. The collection efforts include having sent a letter to notify client J and a follow-up phone call to the accounts payable department of client J. The accounts have not yet been collected. How rigorous was the collection effort?						
Not rigorous						Rigorous
1	2	3	4	5	6	7

Table 3
Questions on "larger"

Experts indicate that one of the factors in determining whether an account will pay off the receivable is "whether the delinquent balance is growing larger."						
a. For account Z, over the course of the last 3 months, the total balance has the following history. Four months ago it was $1000, then $500, then $250 and then $1100. Was that balance getting . . .						
Smaller		Staying the same				Larger
1	2	3	4	5	6	7
b. For account Y, over the course of the last 3 months, the total balance has the following history. Four months ago it was $2000, then $3000, then $3500 and then $1900. Was that balance getting. . .						
Smaller		Staying the Same				Larger
1	2	3	4	5	6	7

6.3. Data analysis

Responses were broken into two different sets: (1) partner/managers and (2) staff, based on experience differences (e.g., Libby and Luft, 1993). Relationships between the test instrument responses of the two sets were summarized and analyzed using three basic approaches. First general observations about the data and the wide range of responses were made. Second, distributions of partner/manager data are compared to distributions of data derived from the other positions. Statistically, the two sets of distributions were compared using a χ^2 approach to test whether or not the distributions are independent of each other. If the distributions are independent, then that indicates that partner/manager evaluations of the judgments come from a different distribution than the staff. Third, distribution means and variances are investigated to determine which of the questions resulted in statistically significant differences in magnitude and variance between the partner/manager set and the staff set. Means were analyzed using a *t* test and variances were analyzed using an *F* test.

Table 4
Question on "materiality"

Experts indicate the write-off of an account that will require a material adjustment is a concern of auditors in the assessment of collectability of an account.						
a. Sales for the firm are $10,000,000 and accounts receivable at year-end are $1,000,000. The account under consideration is $37,500. How material is that account?						
Immaterial		Neither immaterial or material			Material	
1	2	3	4	5	6	7
b. Sales for the firm are $10,000,000 and accounts receivable at year-end are $1,000,000. The account under consideration is $50,000. How material is that account?						
Immaterial		Neither immaterial or material			Material	
1	2	3	4	5	6	7
c. Sales for the firm are $100,000,000 and accounts receivable at year-end are $10,000,000. The account under consideration is $250,000. How material is that account?						
Immaterial		Neither immaterial or material			Material	
1	2	3	4	5	6	7

7. Findings

Distributions of environmental assessments for each of the two sets and a χ^2 test are given in Tables 5–8. Variance and means are summarized in Tables 9 Tables 10.

7.1. Results summary

The results indicate that there was no general consensus of the subjects as to how each of the questions should be answered. In some of the situations posed to the subjects there were responses in each of the seven categories, with environmental assessments ranging from x to $\sim x$. There was substantial range of interpretation by both the partner/manager set and the staff set. The partner/manager set distribution of responses was statistically different than the staff set in two of the four environmental assessment settings posed to the subjects. Furthermore, for one of the question sets there was a statistically significant difference in the means between the two sets. Finally, for all four of the question sets there were some statistically significant differences in the variances, where partners/managers had less variance than the staff.

7.2. "Increasing (balance)"

For two of the three questions posed, the staff responses ranged across the entire spectrum of available answers, while the manager/partner responses were distributed across a smaller set of responses (e.g., for question c all the responses were a "4"). Each of the three questions resulted in statistically significant distributions between the partners/managers and staff, as seen in Table 5. In addition, two of the three questions resulted in statistically significant differences between the variances of the two groups (Table 9) and there was no difference in the mean responses between the two groups.

7.3. "Rigorous"

Responses to the questions on "rigorous" were distributed at each end of the spectrum for the staff, but not for the partner/manager group. None of the three questions generated distributions of responses that were statistically significantly different for the two groups (Table 6). However, the first two questions in the set resulted in statistically significantly different means between the two groups (Table 10), while the third resulted in statistically significant variances (Table 9) between the two groups.

7.4. "Growing larger (delinquent balance)"

Investigation of "larger" was similar to the investigation of "increasing." However, with the addition of a fourth point in the series to be estimated, the results became even less consistent. The range of responses by the staff ranged from 1 to 7 for one question in the set, but the partner/manager group had a tighter range of responses on each question. Both

Table 5
Summary of responses for questions on "increasing"

	Increasing					Decreasing		Total
	1	2	3	4	5	6	7	
Question a								
(PM)	0	0	0	7	5	1	0	13
(Other)	2	0	15	8	23	12	3	63
(Total)	2	0	15	15	28	13	3	76
$\chi^2 = 14.19$ ($df = 5$) (.025)								
Question b								
(PM)	0	0	1	10	1	0	1	13
(Other)	1	6	10	19	22	4	1	63
(Total)	1	6	11	29	23	4	2	76
$\chi^2 = 13.11$ ($df = 6$) (.05)								
Question c								
(PM)	0	0	0	13	0	0	0	13
(Other)	0	2	7	39	15	0	0	63
(Total)	0	2	7	52	15	0	0	76
$\chi^2 = 7.24$ ($df = 3$) (.10)								

(1) (PM) = number of partners and managers; (Other) = number of all other positions of respondents; (Total) = total number of respondents.

Table 6
Summary of responses for questions on "rigorous"

	Not rigorous						Rigorous	Total
	1	2	3	4	5	6	7	
Question a								
(PM)	1	3	4	4	1	0	0	13
(Other)	3	11	14	21	9	3	1	62
(Total)	4	14	18	25	10	3	1	76
$\chi^2 = 1.87$ ($df = 6$)								
Question b								
(PM)	0	0	1	0	4	7	1	13
(Other)	0	0	0	3	17	35	7	62
(Total)	0	0	1	3	21	42	8	76
$\chi^2 = 5.58$ ($df = 4$)								
Question c								
(PM)	1	1	0	3	1	5	2	13
(Other)	1	2	7	13	13	17	9	62
(Total)	2	3	7	16	14	22	11	76
$\chi^2 = 4.98$ ($df = 6$)								

(1) (PM) = number of partners and managers; (Other) = number of all other positions of respondents; (Total) = total number of respondents.

questions generated statistically significant differences between the two groups (Table 7). Furthermore, both questions generated statistically significant variances in the responses (Table 9), but no statistical difference in the means (Table 10).

7.5. "Material (account balance)"

The responses for the sets of questions on "material" for the staff group ranged from 1 to 7, but not for the partner/manager group (Table 8). None of the three questions in the set on "material" generated statistically significant differences between the distributions for the two groups as seen in Table 8. Two of the questions resulted in statistically significant differences between the variances of the distributions of responses of the two groups (Table 9). The means of the two groups were not statistically significantly different for any of the questions (Table 10).

7.6. Discussion of means and variances

The means were significantly different than each other for 2 of 11 questions. Only for the word "rigorous" were the means significantly different. Each of the words "increasing," "larger" and "material" could be based explicit knowledge, such as a computational rule or algorithm. For example, the slope of a linear regression coefficient could be used to determine "increasing" or "larger" as used in the discussion. "Material" might be computed using an algorithm, e.g., "5%." But "rigorous" has no computational approximation. This suggests that there is something requiring a different judgment in determining "rigorous," as compared to "increasing," "larger," or "material." As a result, tacit knowledge likely is driving the judgments for "rigorous." Accordingly, given the different experiences of the two groups it is not surprising to see the means be different.

Table 7
Summary of responses for questions on "larger"

	Smaller						Larger	Total
	1	2	3	4	5	6	7	
Question a								
(PM)	1	1	1	9	1	0	0	13
(Other)	4	10	16	16	9	6	0	61
(Total)	5	11	17	25	10	6	0	74
$\chi^2 = 9.71$ ($df = 5$) (.10)								
Question b								
(PM)	0	1	1	8	2	1	0	13
(Other)	3	6	12	8	20	9	3	61
(Total)	3	7	14	16	22	9	3	74
$\chi^2 = 15.31$ ($df = 6$) (.025)								

(1) (PM) = number of partners and managers; (Other) = number of all other positions of respondents; (Total) = total number of respondents.

Table 8
Summary of responses for questions on "material"

	Immaterial						Material	Total
	1	2	3	4	5	6	7	
Question a								
(PM)	0	2	5	2	2	0	0	11
(Other)	5	6	18	12	7	9	4	61
(Total)	5	8	23	14	9	9	4	72
$\chi^2 = 4.92$ ($df = 6$)								
Question b								
(PM)	0	0	2	4	3	2	0	11
(Other)	0	5	5	10	14	18	9	61
(Total)	0	5	7	14	17	20	9	72
$\chi^2 = 5.90$ ($df = 5$)								
Question c								
(PM)	0	2	2	4	0	0	3	11
(Other)	5	8	10	8	7	8	15	61
(Total)	5	8	10	12	7	8	18	72
$\chi^2 = 6.86$ ($df = 6$)								

(1) (PM) = number of partners and managers; (Other) = number of all other positions of respondents; (Total) = total number of respondents.

The variances were significantly different than each other for 7 of the 11 questions. This suggests that the partner manager group makes choices that are more consistent with each other. Although all have access to the same explicit knowledge rules, the partner manager group has been in quasi apprenticeships of being in a large professional services group. As a result, there is likely to be substantial shared tacit knowledge about such judgments.

Table 9
Variance of responses by question

Question	Partners and managers, S.D.	Others, S.D.	F test σ^2 ratio
1a	0.634	1.331	4.41***
1b	0.890	1.188	1.78
1c	0.000	0.744	∞***
2a	1.264	1.489	1.39
2b	1.273	1.560	1.50
2c	0.679	1.296	3.63***
3a	1.002	1.430	2.04*
3b	0.917	1.523	2.78***
4a	0.979	1.644	2.82**
4b	0.987	1.454	2.30*
4c	1.813	2.020	1.24

* Significant at the .10 level or better.
** Significant at the .05 level or better.
*** Significant at the .01 level or better.

t10.1 Table 10
t10.2 Means of responses by question

t10.3 Question	Partners and managers, mean	Staff, mean
t10.4 1a	4.538	4.555
t10.5 1b	4.231	4.127
t10.6 1c	4.000	4.047
t10.7 2a	2.692*	3.515*
t10.8 2b	2.615**	3.437**
t10.9 2c	2.000	2.406
t10.103a	3.615	3.560
t10.113b	4.077	4.229
t10.124a	3.363	3.869
t10.134b	4.454	5.016
t10.144c	4.273	4.442

t10.15 * Significantly different at the .05 level, $t = -1.859$, $df = 19.4$.
t10.16 ** Significantly different at the .05 level, $t = -1.780$, $df = 20.0$.

Furthermore, with greater explicit knowledge they are more likely to be familiar with similar rules or algorithms to facilitate judgements.

8. Alternatives to requiring environmental assessments

The results presented in this paper find that users provide a wide range of environmental assessments to the exact same situations confounding knowledge-based system use. Without a way of making environmental assessments of the same situations similar, that lack of consistency could drive out the use of expert systems, in spite of their promise. Unfortunately, there is no clear-cut solution to the problem of the possibility of multiple environmental assessments. However, there are some steps that can be taken to mitigate the impact of multiple potential environmental assessments.

8.1. Definitions

Requiring environmental assessments suggests supplementation of the system with a set of definitions and training. For example, Kelly et al. (1987, p. 172) present the user with the following.

> Based on your judgment, if a set of financial statements were to be generated as of today, do the comments provided by the loan officer suggest to you that the loan officer, based on his/her knowledge of the borrower's current condition, believe the borrower is in a strong, moderate or weak short-term liquidity condition?
> Enter
> Strong if the loan officer believes the short-term liquidity condition of the borrower is strong.

Moderate if the loan officer believes the short-term liquidity condition of the borrower is moderate.
Weak if the loan officer believes the short-term liquidity condition of the borrower is weak.

Thus, the user is provided with guidance on environmental assessments for "strong," "moderate," and "weak." Unfortunately, such definitions can be ambiguous or not used. Furthermore, it could be very difficult to design a system that would use all the definitions for all the environmental assessment situations and alternative data sets necessary for a term like "increasing." Theoretically, this change is one related to the audit support tool.

In addition, in the particular example, it is arguable if the descriptions provide any real support. For example, the explanation is defined using the term it is trying to define.

8.2. Append notes

Another alternative is to provide the ability of a user to append notes, in cases where there is uncertainty about their environmental assessments. User-appended notes could provide comments on why the user made the assessments that they made. However, there is not necessarily a relationship between environmental assessment and uncertainty. Furthermore, there are numerous costs associated with capturing such notes that can make their use infeasible. In addition, information that is ancillary is not likely to be examined or used.

8.3. Validation

Validation typically is used to examine the quality of the system (e.g., O'Leary, 1987). Unfortunately, not all scenarios can be tested, and not all possible rules and conditions can be examined. As a result, even if the system has a single rule requiring environmental assessment, that rule may not be examined during validation.

Furthermore, if that validation is concerned only with the computing artifact, assessment issues will not be addressed. If instead, the system also is tested with users then the validation findings are likely to vary based on the user groups examined. For example, as seen in this paper, if the users are from the partner and manager sets then the assessment are likely to be similar. However, validation focusing on nonpartners and nonmanagers may find variation in the analysis, leading to questioning of the system. This approach also is consistent with trying to change the audit support system to facilitate consistency.

8.4. Empirical tests with users

If there are different user groups or different groups using and reviewing results, empirical tests could be conducted to determine the existence of differences between user environmental assessments, as reported in this paper. Unfortunately, determining the existence of differences between groups does not provide a solution, only recognition of the problem. However, such tests could indicate the specific educational needs for use of a particular system.

8.5. Education

Education, part of the experience component of the Libby and Luft (1993) model, could be used to ensure that users view environmental situations in a similar manner. This approach is aimed at changing the users' knowledge. Heuristic rules for environmental assessments are likely to result from this process, such as the frequently used rule for materiality (''A quantity is material if it is 5% of some other quantity''). Education can facilitate generation of more uniform assessments. However, training cannot anticipate all possible situations.

8.6. Mechanistic estimation rather than user assessment

An alternative to having users provide environmental assessments is to have the system make those assessments. Data could be input to the system, without providing any human judgment.

For example, if materiality is the issue, then provide the system with a parameter (e.g., 5%) and an amount to multiply (e.g., sales or total assets) and let the system do the multiplication. Alternatively, statistical tests could be used to provide judgment. For example, a regression equation would be computed and the slope of that equation would be used to determine if the data were increasing or decreasing.

Unfortunately, there are some limitations to this approach. First, some assessments are not just a mathematical relationship. For example, ''rigorous'' may be more difficult to measure with such an equation. Second, if the company being audited realizes that simple relationships are being used to guide the analysis, then ''gaming'' behavior can ensue. Third, there may not be computer accessible data available for the assessments, e.g., rigorous. Fourth, the systems are modeled after expertise. It is unlikely that experts make decisions based on mechanistic models. As a result, if organizations want to leverage expert behavior, they will need to build real expertise into the system. Fifth, such an approach ignores the capabilities of people and their ability to monitor events around themselves. For example, ''rigorous'' may vary from situation to situation. Sixth, why have the auditor out there if the system is to make all of the assessments? Why not just hook the system up to the Internet and let it do all the work. The consistency achieved by the partners and managers in our sample likely occurred, in part, because of experiences of the participants in making these kinds of decisions. Seventh, this approach could ignore qualitative aspects of which the user may be aware but the system does not consider. Furthermore, some may suggest that this approach turns the user of the system into a data entry clerk, who does not ''think.'' As a result, there could be substantial negative reaction by the users and management, who want to be more than data entry personnel.

However, the organization ultimately needs to make a cost–benefit decision to determine if the computer or the person makes the assessments. Furthermore, the organization needs to determine if it wants to put its trust into a person or a machine or some combination of the two.

8.7. Summary

None of the alternatives examined in this section ensures a consistent response to environmental assessments except for the last one, mechanistic estimation. As a result, it is

easy to see that use of classic expert systems where the user is responsible for environmental assessment is likely to result in inconsistent audit decision making. As a result, if consistency is important then that would lead to a shift away from the use of classic expert systems.

Although mechanistic estimation also goes a long way to removing inconsistency from the process, it also removes the human from the process. With mechanistic estimation, both assessment and processing would be done by the system. Accordingly, if the concern were with consistent decision making, we would expect a shift to system-based analysis, such as that in continuous auditing systems where the systems contain expertise and environmental assessment of data streams.

9. Discussion and limitations

My analysis leads to three primary findings, each of which can influence development of accounting and auditing research and practice. First, I note that classic accounting knowledge-based systems typically solicit environmental assessments from their users, and the approach is the norm, not the exception. Systems are generally designed to work with the users, not without them. Second, the empirical results indicate that there are likely to be a (wide) range of environmental assessments to the same situation. One user may see a situation as material and another immaterial, each driving the system to make different recommendations. Third, the empirical results suggest that partners/managers and staff have different distributions for some environmental assessments. In particular, the variances for assessments that use explicit knowledge is smaller for the partner and manager group than for the staff group. For those judgments where tacit knowledge is required, it appears that the mean assessments themselves also are statistically significantly different for the two groups.

In addition, the empirical results suggest that an important component of accounting expertise, as represented in organizational position (e.g., partner) may be that experts make environmental assessments similar to other experts. It is not just the decisions that experts make or the knowledge they have, but also their interpretation of the data they gather. As a result when we think of models of auditor behavior, personnel with different experience or expertise or organizational level will likely be more effective in their use of models to generate good decisions than other users, simply based on their environmental assessments. This leads us to the "user not neutral hypothesis"—models will not generate the same solutions for different users for the same situations because their environmental assessments will differ. This hypothesis suggests that an additional class of interaction, beyond the two noted by Libby and Luft (1993). In particular, experience, knowledge and ability impact the use of the computer system, and thus the judgments recommended by the systems.

It raises a dilemma of sorts in the development and testing of knowledge-based systems. Typically such systems are built on knowledge derived from experts, who are typically partners and managers, and partners and managers test that knowledge. However, little attention is given to the impact of differing environmental assessments or impact of having different users provide different environmental assessments for inputs to knowledge-based systems.

Finally, it is arguable that the inability of different levels of personnel to produce consistent solutions to the same problem is a fundamental flaw of classic expert systems. If environmental assessments vary then the fundamental concept on which expert systems is based—solicit knowledge from experts so that others may solve problems using the same knowledge—is not feasible.

9.1. Limitations and extensions

As with all research, this paper has some limitations. First, it focuses primarily on the linkages between experience, knowledge and performance. This paper does not account for or study linkages to ability. Further research could examine the impact of ability on assessments. However, we can speculate that higher ability leads to better assessments. Second, this research examines the environmental assessments; however, it does not empirically examine the "user not neutral hypothesis." that hypothesis suggests that, in general, users that have experience and knowledge that provides them with a better environmental assessment, will get solutions that are more appropriate to the actual situation from their computer-based tools. Experience and knowledge influence the quality of the relationships that can be generated with computer-based tools. Future research could also examine this issue.

10. Uncited references

Hansen and Messier
O'Keefe and O'Leary, 1993
Willingham and Ribar, 1988

Acknowledgements

An earlier version of this paper was presented at the International Research Symposium on Accounting Information Systems, December 2002. I would like to thank Jim Hunton, the anonymous referees and participants at that meeting for their comments. I also would like to thank Eddy Vaassen and Jane Fedorowicz for their comments.

References

Dillard J, Mutchler J. A knowledge-based support system for the auditor's going concern judgment. Unpublished paper; 1987.

Dungan C. Analysis of audit judgment in the form of an expert system, Unpublished PhD dissertation, University of Illinois; 1983.

Dungan C, Chandler J. Analysis of audit judgment through an expert system. Faculty Working Paper no. 982, University of Illinois, School of Business; November 1983.

Dungan C, Chandler J. AUDITOR: a microcomputer-based expert system to support auditors in the field. Expert Syst October 1985;210–21.

Hansen J, Messier W. A preliminary investigation of EDP-XPERT. Auditing J Theory Pract 6.
Hansen J, Messier W. A knowledge-based expert system for auditing advanced computer systems. Eur J Oper Res 26:371–379.
Kelly K, Ribar G, Willingham J. Interim report on the development of an expert system for the auditor's loan loss evaluation. Proceedings of the Touche Ross/University of Kansas Symposium. 1987;167–88.
Kneale D. How Coopers and Lybrand put expertise into its computers. Wall Street J 1988.
Libby R, Luft J. Determinants of judgment, performance in accounting settings: ability, knowledge, motivation and environment. Account Organ Soc 1993;18(5):425–50.
Molay K. Developing real-world Java applications with Blaze Advisor from HNC. Web presentation, June 19, 2002.
O'Keefe R, O'Leary D. Expert system verification and validation. Artif Intell Rev 1993;7:3–42.
O'Leary D. Validation of expert systems: with applications to auditing and accounting expert systems. Decis Sci 1987;18:468–86.
Rennie M, Gibbins M. Expert beyond experience. Camagazine May 1993;40–6.
Steinbart P. The construction of an expert system to make materiality jugdments, Unpublished PhD dissertation, Michigan State University; 1985.
Steinbart P. The construction of a rule-based system as a method for studying materiality judgements. Account Rev January 1987;LXII(1):97–116.
Sveiby K. Tacit knowledge. In: Cortada J, Woods J, editors. The knowledge management yearbook. Boston: Butterworth Heinemann, 1999.
Trewin J. The impact of the introduction of an expert system on a public accounting organization. Int J Intell Syst Account Finance Manage 1995;5(3):185–97.
Willingham J, Ribar G. Development of an expert audit system for loan loss evaluation. Auditor productivity in the year 2000 Reston, VA: Arthur Young; 1988.