Spreadsheet Presentation and Error Detection: An Experimental Study

DENNIS F. GALLETTA, KATHLEEN S. HARTZEL,
SUSAN E. JOHNSON, JIMMIE L. JOSEPH, AND
SANDEEP RUSTAGI

DENNIS F. GALLETTA is an Associate Professor of Business Administration at the Katz Graduate School of Business, University of Pittsburgh. His current research interests include end-user computing and user training, particularly in the areas of end-user attitudes, behavior, and performance. His articles have been published in such journals as *Information Systems Research, Communications of the ACM, Decision Sciences, Information and Management, Data Base,* and *Accounting, Management, and Information Technologies.*

KATHLEEN S. HARTZEL is a doctoral candidate in management information systems at the Katz Graduate School of Business, University of Pittsburgh. Her current research interests include management of attitudes and behaviors during system development and use, effects of computer-mediated support on decision making and attitudes in both individuals and groups, and the impact of interorganizational information systems on cooperation and competition.

SUSAN E. JOHNSON is a doctoral candidate in artifical intelligence at the Katz Graduate School of Business, University of Pittsburgh, and is an information analyst with Bayer Corporation. Her current research interests include automated document generation and human–computer interaction.

JIMMIE L. JOSEPH is a doctoral candidate in management information systems at the Katz Graduate School of Business, University of Pittsburgh. His current research interests include human–computer interaction, effects of computers on organizational behavior, software utilization, and satisfaction measurement.

SANDEEP RUSTAGI is a doctoral candidate in management information systems at the Katz Graduate School of Business, University of Pittsburgh. His current research interests include information systems strategies in multinational organizations and international information systems.

ABSTRACT: The pervasiveness and impact of electronic spreadsheets have generated serious concerns about their integrity and validity when used in significant decision-making settings. Previous studies have shown that few of the errors that might exist in any given spreadsheet are found, even when the reviewer is explicitly looking for errors. It was hypothesized that differences in the spreadsheets' presentation and their formulas could affect the detection rate of these errors. A sample of 113 M.B.A. students volunteered to search for eight errors planted in a one-page spreadsheet. The spreadsheet was presented in five different formats. A 2×2 design specified that four

Journal of Management Information Systems / Winter 1996–97, Vol. 13, No. 3, pp. 45–63

ISSN: 1086–4415/1997 $9.50 + 0.00

groups were given apparently conventional spreadsheets for comparing paper and screen and the presence or absence of formulas. A fifth group received a special printed spreadsheet with formulas visibly integrated into the spreadsheet—printed in a small font directly under the resultant values. As in previous studies, only about 50 percent of the errors were found overall. Subjects with printed spreadsheets found more errors than their colleagues with screen-only spreadsheets but they took longer to do so. There was no discernible formula effect; subjects who were able to refer to formulas did not outperform subjects with access to only the final numbers. The special format did not facilitate error finding. Exploratory analysis uncovered some interesting results. The special experimental integrated paper format appeared to diminish the number of correct items falsely identified as errors. There also seemed to be differences in performance that were accounted for by the subjects' self-reported error-finding strategy. Researchers should focus on other factors that might facilitate error finding, and practitioners should be cautious about relying on spreadsheets' accuracy, even those that have been "audited."

KEY WORDS AND PHRASES: electronic spreadsheets, information system audits, spreadsheet validation.

SPREADSHEET SOFTWARE IS UNIQUE IN THAT THERE IS PERHAPS no other single software package that entails simultaneously liberal ease of use, widespread availability, and risk. The business community is becoming ever more aware of computing errors. In the last decade, reports of spreadsheet errors and erroneous business decisions have appeared from time to time in the popular press [42, 43]. The most striking example of public awareness of computer-related errors was the widespread publicity surrounding the Pentium[1] processor's floating point error.

Recently, some researchers have also addressed the serious nature of spreadsheet errors. They have learned about the conditions leading to errors, their probable extent, their likely effects, and general techniques for reducing their occurrence. Unfortunately, we are not yet able to predict or to prevent them completely. It is therefore important for us, for the foreseeable future, to focus on error *detection*. Toward this end, this study examines one aspect of the detection process, the effects of *spreadsheet presentation* on error detection.

Research on Spreadsheet Errors

MANAGING END-USER COMPUTING HAS BEEN FRAUGHT WITH DIFFICULTIES for many years [1, 19, 30], including a lack of planning [1, 18, 30], policies [8, 10, 15, 18, 41], and controls [19, 22, 40]. One possible explanation for these difficulties is the constancy of change: Just as organizations learn how to manage one technology, the next innovation invalidates the "old" rules.

It is easy, moreover, to use spreadsheet software incorrectly. One author notes that spreadsheet tools are harder than programming languages to use correctly [44]. While advancements in interfaces, help systems, and automated functions make spreadsheet software accessible to users with less and less experience [27], confusion and inadequate training have been reported [6, 28, 39]. And yet, spreadsheet authors have great confidence in their results [7, 12, 15], perhaps because of the ease of use and professional-looking results.

Perhaps their high confidence would diminish if spreadsheet authors were aware of the actual incidence of errors. Panko [37] notes that "every study that has looked for errors has found them and has found them in abundance" (p. 327). Although their numbers do not sound very high in terms of the percentage of cells with errors, they do sound alarmingly high in terms of the percentage of models with errors. Panko and Halverson found that subjects made errors in only about 1 percent of the spreadsheet cells, but the number of cells per model translated to an error rate of 57 percent of the spreadsheet models. According to Panko [34], even a tiny cell-error rate will be multiplied over a logic cascade into a high probability of an error in bottom-line values.

Both laboratory and field studies by other researchers indicate that between one-third and one-half of all spreadsheets contain errors [7, 10, 11, 12, 14, 22]. Field results indicate that the error severity can be quite high, from hundreds of thousands [26] to millions [12, 14] of dollars.

Researchers [3, 33] and practitioners [2, 38, 42] alike offer advice about how errors are made, what their effects are, and how they might be prevented. One of the most common prescriptions is to take advice from the programming literature and develop large spreadsheets in small chunks rather than all at once. Because modularizing a large spreadsheet requires linkages among the various chunks, it is unfortunate that spreadsheets' most error-prone areas have been in those linkages [25].

User training shows some promise as an approach for minimizing errors. Unfortunately, most courses and textbooks seem to focus more on how to use features of particular spreadsheet packages than on how to design the contents of the spreadsheets for error prevention. Also, many firms cut back on training expenditures when the specter of difficult economic times appears on the horizon.

The sheer size and complexity of spreadsheet models will make it very difficult, if not impossible, to eradicate errors completely. Hall's [22] field study found that 57 percent of sampled spreadsheets were linked to other spreadsheets or databases and that 55 percent contained macros. Floyd, Walls, and Marr [15] found that the typical spreadsheet constructed by seventy-two users contained 6,000 cells.

Three reasonable questions suggested by this situation are:

- How good are human reviewers at detecting spreadsheet errors?
- Can automatic error detection help eliminate errors?
- What factors facilitate manual detection?

Efficacy of Manual Detection

A previous laboratory study [16] showed that only about 56 percent of the seeded errors in independent, one-screen spreadsheets were detected. Even the fifteen subjects in the best cell of their study, filled with CPAs with extensive spreadsheet experience, found only 66 percent of the errors. When told of the actual error locations after the study, subjects generally considered the errors to be obvious after all, but simply overlooked.

Worse yet, in Panko's [34] study, only 16 percent of spreadsheet developers who were asked to debug their own creations could do so successfully. As Galletta et al. [16] stated, it is surprising that "even obvious, elementary errors in very simple, clearly documented spreadsheets are . . . difficult to find" (p. 91).

The small number of studies in spreadsheet error finding are supplemented by studies in the programming literature that describe error-finding rates of 33 percent to 46 percent for professional programmers [4, 31]. Because spreadsheets are so commonly used for important decisions, and the average spreadsheet includes linkages, macros, and hundreds or thousands of cells, there is indeed cause for concern.

Automatic Detection

Because humans have great difficulty finding errors, it is logical to turn to automated auditing tools for assistance. Unfortunately, spreadsheet auditing tools have limited effectiveness in error finding. These tools have been available for over a decade but they lag behind the variety of functions used in spreadsheets and cannot be expected to interpret the purpose of a spreadsheet for a bottom-line "seal of approval." Auditing tools built into the latest versions of Lotus 1–2–3[2] and Excel[3] simply point out, cell by cell, dependencies and linkages to other cells. Any discoveries of improper dependencies and linkages are left to the user to make, which requires detailed understanding of the *meaning* of the contents of each cell.

Factors that Facilitate Manual Detection

Given that automated tools are of limited utility, it is necessary to explore factors that will facilitate manual detection. A model by Galletta and colleagues [17] describes several factors that might contribute to the efficacy of error finding (see figure 1). The determinants of error-finding performance include individual factors, presentation factors, error factors, and external factors.

Individual factors can include the experience of the spreadsheet auditor in the domain of the spreadsheet problem. For example, the spreadsheet auditor might be very knowledgeable in printing-cost estimation, which would lead to more effective identification of errors in a spreadsheet for printing-cost estimation. Another spreadsheet auditor might be highly experienced in the domain of the spreadsheet tool used.

In a previous study in error detection, Galletta et al. [16] found that accounting expertise helped subjects find more errors, but that spreadsheet tool experience helped subjects find the errors more quickly. Subjects with both kinds of expertise were the best and fastest performers.

Skill in the problem of spreadsheet error finding itself might also be a valuable factor to study, especially when or if highly experienced spreadsheet auditors become an

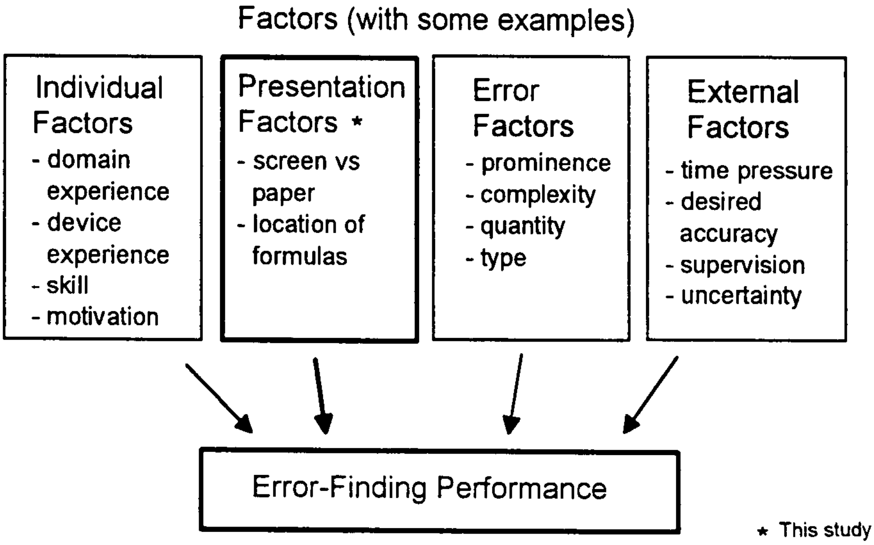

Figure 1. Conceptual Model (from [17])

abundant resource. At this point, relatively few people devote much time to auditing spreadsheets, especially compared with the millions who build them.

Motivation is also likely to be an important determinant of error-finding performance. It is easier to visualize a setting in which spreadsheets are examined as an afterthought and in haste than it is to imagine one in which a great deal of effort and planning is expended in error-finding incentive programs. Individuals sometimes even seem reluctant to find their own faults.

Presentation factors, examined in this study, have theoretical significance in error finding. Error finding is more of an art than a science. There are no definitive procedures for their examination. Some questions that we pose are: (1) Is error finding facilitated by printing spreadsheets or is it adequate to view them on the screen? (2) Does access to formulas help in the review of spreadsheets or do formulas require more reading, slowing down the reviewer? (3) If formulas are available, is it more helpful to place them contiguous to the results in each cell, in permanent, "spatially correct," full view of the reviewer?

The first of these questions can be answered with some confidence by reviewing the literature on reading from paper versus CRT screens. Gould and colleagues [20, 21] conducted a series of studies of why reading from paper was generally about 20 to 30 percent faster and more effective than reading from a typical CRT screen, as established in previous work [13, 29, 32]. The studies examined factors such as brightness, contrast, polarity, visual angle, and slope of the viewing surface, only to discover that the text quality was the important factor. Using a special "anti-aliasing" display that smoothed the jagged edges of computer characters, they were able to eliminate paper's incremental advantage.

What is not clear is the effect on error-finding performance of the location and availability of the formulas. Many spreadsheet users print all of the formulas in a spreadsheet for backup and audit purposes. Presumably, if a spreadsheet auditor can review the formulas underlying spreadsheet values, he or she can avoid entering the amounts into a calculator to check the totals of long lists of numbers. However, the availability of formulas also requires more reading, perhaps imposing greater cognitive load, which might impede the performance of spreadsheet auditors.

Error factors have been largely unexplored to date but are likely to be strong determinants of error-finding performance. The most obvious example is the prominence of the errors. Prominence includes the error's relative size and location on the spreadsheet. Another example of an error factor is the complexity of the spreadsheet or formula containing the error. Spreadsheet formulas range from simple linear sums to aggregations or manipulations of noncontiguous numbers, and even to linkages with data not visible or accessible to the spreadsheet auditor.

The sheer number of errors can also affect performance. The presence of many errors might mask any particular error. When confident reviewers find an error, they might conclude prematurely that their reviewing work is complete.

The final factor to consider is the nature of the error itself. Several authors [16, 36] offer a multitude of error "types." Panko and Halverson differentiate between *mechanical errors* (where an improper key is pressed or an improper cell is clicked in a formula) and *logic errors* (where the algorithm used is incorrect). Panko and Halverson proposed and found logic errors to be more common than mechanical errors because they are more difficult to find. Recently, Panko [34] has added the category of omission errors, where an important element is missing from a formula.

Another categorization scheme is that of *domain* versus *device* errors [16], mirroring the categorization of experience described earlier. Domain errors are those that result from misunderstanding or misapplication of the content area. For example, discovering an accounting error such as including prepaid expenses in an income statement or adding a debit to a liability account requires accounting knowledge. Device errors, on the other hand, include errors in carrying out calculations that are otherwise correct. Examples of device errors include omitting a cell at the end of a range or typing an incorrect cell address in a formula; discovery of such errors requires knowledge of spreadsheet functions.

External factors might also be important determinants of error-finding performance. Examples of promising external factors would be time pressure, desired accuracy, extent of supervision, and overall uncertainty in the spreadsheet. Time pressure and desired accuracy are likely to affect the amount of effort expended in error finding. The notion of desired accuracy has some empirical support [40], where the amount of risk determined the amount of review given the spreadsheets. Obviously, a spreadsheet devoted to an analysis of a merger should receive greater scrutiny than a spreadsheet that sums costs for a company picnic. The extent of supervision can also affect error finding for a variety of reasons. Little supervision is likely to lead to an absence of imposed review procedures. Little supervision can also fail to point out the importance of errors, serving as the breeding ground for a large number of errors and leading to a lack of understanding of their potential impact. Finally, uncertainty in the problem solved by the spreadsheet can lead to the absence of objective comparison information that would have been used to test the models in the spreadsheet. A model that sums costs can be proven to be correct, but formulas that take into account probabilities for simulating likely outcomes can be difficult to examine or test.

In this study, we focus on a single set of variables, the presentation factors. Other factors will be examined in future studies.

Hypotheses

BASED ON THE SCREEN-VERSUS-PAPER LITERATURE, REVIEWING A SPREADSHEET on paper should result in higher performance than reviewing a spreadsheet on a screen [13, 20, 21, 29, 32]. Performance can be expressed in terms of both the quantity of errors and time required to find the errors.

> *H1: Reviewing a spreadsheet presented on paper will result in higher error-detecting performance than reviewing a spreadsheet presented on a screen.*

As an "audit trail," formulas can allow users to review how sums are derived. As discussed earlier, formulas should obviate the need for tabulation. Although formulas present more reading material, introducing the possibility that reviewers will feel the need to read them when it might not be necessary, they are intended to be used as discretionary resources. Therefore, we propose that their availability will aid in error-finding performance:

> *H2: Reviewing a spreadsheet with formulas will result in higher error-detecting performance than when formulas are not available.*

A final hypothesis addresses the issue of accessibility of those formulas. Examination of formulas usually requires the user to move back and forth between different sheets of paper, or to move the cursor about on the screen because all but the current cell's formulas are hidden. The problem is particularly vexing on paper spreadsheets, where obscure lists of cell contents and results are listed in linear format on numerous supplementary pages. Cursor positioning and page turning would add enough steps to the task, and perhaps enough time [9], that we propose that error-finding performance would be aided by making all formulas visible, contiguous to the results as they are placed on the spreadsheet. To that end, we experimented with a special "integrated" fifth design on paper (see appendix B) that displays the formulas in small print directly beneath their results. This experimental design was an attempt to go beyond discovering factors leading to error-finding performance, and to provide actual *improvements* in that performance.

> *H3: Error-finding performance will be enhanced by integrating spreadsheet formulas in the printed spreadsheet than when they are hidden or attached in a linear list.*

Method

A ONE-PAGE SPREADSHEET WAS CREATED THAT REQUIRED LITTLE domain knowledge of subjects (see appendix A). The setting involved discovering the level of funding

Table 1. The Five Treatments

	Conventional formulas	No formulas	Integrated formulas
Paper	Printed spreadsheet, formulas attached (customary approach when audit trail desired)	Printed spreadsheet results only (customary approach for presentation)	Printed spreadsheet with formulas spatially integrated into the presentation
Screen	On-screen spreadsheet (customary approach)	On-screen spreadsheet with formulas removed	

necessary to provide enough resources for three terms of a student's M.B.A. program. This task was chosen because it was likely to be familiar to the student participants.

To test the hypotheses, five treatments were created (see Table 1). The first four represented a 2×2 matrix of conventional spreadsheets: screen versus paper and formulas versus no formulas were the two dimensions. The fifth treatment represented the experimental, integrated format presented on paper.

The table describes the treatments. The most common treatments found in practice are the paper treatment without formulas and the screen treatment with formulas. Perhaps less common is the printed listing of formulas provided by some spreadsheet packages (appendix C). Ours was printed automatically by Lotus 1–2–3. Although such a treatment is likely to be rare, a screen treatment without formulas was created simply to enable statistical analysis of both dimensions in Table 1. Finally, the integrated formulas treatment was created (appendix B). This spreadsheet allowed the reviewer to examine the results of formulas (in large print) as well as formula entries (in small print, directly under the large results) at one glance.

Subjects

M.B.A. students in five sections of a graduate MIS course at a large university in the northeastern United States were invited to volunteer for the study. Out of nearly 250 candidates, 115 volunteers were found. After disqualifying two subjects for failing to follow instructions, 113 subjects remained. Subjects were not paid, but as an incentive, a cash prize of five dollars was promised to the top third of participants in each condition.

Performance was assessed by examining the number of errors correctly identified and the amount of time required to complete the task. An additional, exploratory measure recorded the number of false positives—correct items identified as errors.

Materials

Subjects were asked to find as many errors as they could in the simple, one-page student budget worksheet. Although they were not told how many errors there were, eight errors were planted in the spreadsheet. The errors ranged from values improperly entered from the assumption sheet to errors in entering formulas. As in a previous study [16], the impact of each error was tightly constrained to enable scoring of the results with a minimum of confusion that would result from cascading errors. Pilot testing resulted in modifications to the instructions, planted errors, consent form, and exit questionnaire to maximize clarity.

An answer sheet was created that would not reveal that there were eight errors in total; subjects were told that while the sheet was filled with fifteen lines, "a great deal of extra space" was provided. On each line, subjects recorded the location of the error, the time the error was found, and a brief description of the error. Projected on the wall was a large timer that showed the elapsed time to the second.

A calculator was furnished to all students in case they wanted to do manual calculations. The subjects in the computer conditions used 486, 33 MHz networked AT&T PCs running Lotus 1–2–3 version 5 for Windows[4] and 14-inch color super-VGA monitors.

Experimental Procedures

To stimulate participation, arrangements were made with the professor teaching the classes to dismiss students thirty minutes early. The first author entered the classroom, and the professor introduced him, encouraged students to participate, and then left the room. In spite of the encouragement, about 20 percent of the students followed the professor out the door immediately. Some apologized on their way out, stating that a big assignment was due. The first author described the purpose of the study and asked for volunteers to be experimental subjects.

The remaining students were asked to read the informed consent form and portions of the instructions. Students were told that their performance would be scored by assessing both the number of errors correctly identified and the speed at which the errors were found. The subjects were then randomly assigned to one of the five conditions and the computer-assigned subjects were asked to follow the experimenter to a computer lab.

We learned about the need to provide continual encouragement on the way to the lab in the first of the five classes. While moving to the lab, several students (another 20 percent) left the line and fled downstairs to a nearby lounge. We provided a dramatic remedy for that problem in the remaining four sessions by supplying a second usher near the end of the line to smile and stand "innocently" in front of the staircase. This made it slightly less convenient to escape the experiment.

For all groups, a researcher instructed the participants to begin and then immediately started the LCD-projected timer. Subjects were asked to record the error location and current time displayed on the screen after each error was found. They were also asked to briefly (to save time) describe the error so that when the answer sheets were scored, the accuracy of the answer could be more precisely determined.

Subjects were told that they should find as many errors as they could and to use their own judgment to say when they were finished. After finding the last error, instructions pointed them to an exit questionnaire that asked them to record the final time for completing the task and for demographic (described below) and other data.

Coding of Outcomes

To provide some assurance of coding accuracy, two judges scored the results. Both assigned scores for each subject without knowledge of their experimental condition or the other judge's score. A third judge compared the findings of the two original coders and found that the raters only disagreed on six occasions from 904 judgments (8 for each of 113 subjects). The third judge, blind to the experimental condition, served as a "tie breaker" for each of the six discrepancies and determined the proper coding of each item. The judges therefore agreed on more than 99 percent of the judgments. Cohen's Kappa, a measure of interrater reliability unbiased by the distribution of the possible responses, was also over 0.99 [5].

Results

THE OVERALL RESULTS ARE SHOWN IN TABLE 2. THE FIRST STEP IN THE ANALYSIS of those data involved a preliminary ANOVA to detect any possible differences in demographic variables between the groups, which would have indicated a bias in assignment of subjects. Demographic variables included several measures of experience: entering data, using a computer, modeling data relationships, auditing for error finding, and training users. Other demographic variables included gender, age, foreign status, and English as a primary language. Subjects also signed permission forms that we used to obtain their total GMAT scores from the dean's office. Fortunately, there were no significant differences in any of the variables among the five groups, and our random assignment of subjects appears to have succeeded.

Table 2 summarizes, for all five conditions, the number of errors found and time taken (up to the last error found). The following discussion presents our findings for subsets of that table.

A two-way ANOVA[5] considering screen versus paper as one factor and formulas versus no formulas as another factor permitted examination of the first two hypotheses. Hypothesis 1 predicts higher performance for subjects using paper than for subjects using screens. The ANOVA found that the screen effect is indeed significant on both number of errors found [$F(1,109) = 7.62$; $p = 0.007$] and time taken [$F(1,108) = 6.23$; $p = 0.044$]. Interestingly, while the number of errors found behaved in the hypothesized direction, the time taken did not; subjects using paper actually took longer. The descriptive statistics are shown in Table 3.

Hypothesis 2 predicts higher performance for subjects able to refer to formulas than for those who are not. The ANOVA (and ANCOVA) failed to reveal any differences[6] in performance based on the presence or absence of formulas, and therefore fails to support hypothesis 2. Table 4 contains the descriptive statistics for formula and non-formula treatments.

For hypothesis 3, ANOVA was used to test the efficacy of our experimental

Table 2. Overall Results, Means (Standard Deviations)

	Conventional formulas	No formulas	Integrated formulas
Paper	$N = 23$ 4.2 errors found (1.6) 15.0 minutes taken (5.4)	$N = 23$ 4.6 errors found (1.6) 15.6 minutes taken (3.8)	$N = 22$ 4.4 errors found (1.8) 15.4 minutes taken (3.4)
Screen	$N = 23$ 3.9 errors found (2.0) 13.8 minutes taken (6.0)	$N = 22$ 3.2 errors found (1.3) 12.2 minutes taken (5.3)	

Table 3. Cell Means (Standard Deviations), Screen versus Paper

	Screen ($n = 45$)	Paper ($n = 68$)
Number of errors found	3.6 (1.7)	4.4 (1.7)
Time taken	13.0 (5.7)	15.4 (4.2)

Table 4. Cell Means (Standard Deviations), Formulas versus No Formulas

	Formulas ($n = 68$)	No formulas ($n = 45$)
Number of errors found	4.2 (1.8)	3.9 (1.6)
Time taken	14.5 (5.0)	13.9 (4.9)

integrated formula treatment; performance of subjects in that treatment was compared with performance of subjects in all other treatments (see Table 5). Again, no differences were found,[7], and hypothesis 3 is not supported.

Exploratory Analysis

Three exploratory analyses were also conducted. The first analysis investigated whether or not the false positives were affected by the treatment. These data were obtained by grading the subjects' submissions. Means for false positives by group appear in Table 6.

The second analysis divided the time taken by the number of errors found, yielding the number of minutes taken to detect each real error. This would allow us to determine if there was a trade-off effect between time and accuracy. Means for this new variable also appear in Table 6.

Two-way analysis of the means in Table 6 failed to reveal significant differences between the treatments. However, when the experimental integrated treatment is subjected to a one-tailed t-test comparing the number of false positives in that treatment with all others, the test reveals significance at the 0.05 level.[8] A one-tailed test is justified by the assumption that performance should be better in the integrated

Table 5. Cell Means (Standard Deviations), Integrated Formulas versus the Other Treatments

	Integrated formulas ($n = 22$)	Other cells ($n = 91$)
Number of errors found	4.5 (1.8)	4.0 (1.7)
Time taken	15.4 (3.4)	14.2 (5.3)

Table 6. Cell Means (Standard Deviations), for Exploratory Variables

	Conventional formulas	No formulas	Integrated formulas
Paper	2.0 false positives (3.3) 3.7 minutes per error (1.5)	1.7 false positives (1.6) 4.3 minutes per error (3.7)	1.2 false positives (1.5) 3.8 minutes per error (2.0)
Screen	2.0 false positives (2.8) 3.4 minutes per error (1.5)	2.8 false positives (2.6) 4.4 minutes per error (2.7)	

treatment. This suggests that the integrated paper treatment has some effect on reviewer performance.

The third analysis investigated the self-reported strategy of the subjects. Subjects in formula groups checked from a list of ten strategies. The choices assessed independently what subjects did first and the manner in which subjects resolved intermediate totals. The data on subjects' initial steps seemed to fall naturally into three groups: those who began their quest by checking addition by row (adding across rows), those who started by checking addition by column (adding down columns), or those who began by scanning for formulas to investigate. The data on intermediate totals fell into two groups. Intermediate totals could be evaluated immediately (that is, subjects would discover that school costs in row 4 are dependent on the total of rows 10–12, and immediately check to see if the total is correct), or could be evaluated later (that is, they could wait until they arrive at rows 10–12 to make sure the totals are correct). Subjects were subdivided according to how they evaluated intermediate totals. Table 7 summarizes the row-versus-column data and Table 8 summarizes the intermediate-total data.

The data in Tables 7 and 8 were subjected to ANOVA; differences were found to be significant in each case [$F(2,91) = 3.94$; $p < 0.023$ in Table 7; $F(1,75) = 4.60$; $p < 0.035$ in Table 8]. Furthermore, the data in Tables 7 and 8 were merged and subjected to ANOVA, and the analysis reveals the strategy differences more strongly and specifically. An ANOVA comparing all of the cells suggests that the strategies indeed led to significantly different performance [$F(4,88) = 3.67$; $p < 0.008$]. The best strategies seem to be either to focus first on rows and resolve all intermediate totals (mean errors found = 4.9) or to focus only on formulas (4.7). The worst strategies are the rest, averaging nearly identical results in the other cells (3.4).

Table 7. Cell Means (Standard Deviations), Row versus Column Strategy

	Number of errors found
Rows-first strategy	4.3 (1.8)
Columns-first strategy	3.4 (1.9)
Formulas-first strategy	4.6 (1.1)

Table 8. Cell Means (Standard Deviations), Intermediate Total Strategy

	Number of errors found
Immediate resolution of intermediate totals	4.3 (1.9)
Postpone resolution of intermediate totals	3.4 (1.8)

Discussion and Conclusions

OVERALL, IT APPEARS THAT SPREADSHEET REVIEWERS ARE NOT AIDED in their search for errors by having the formulas on hand but are aided in finding more errors by examining spreadsheets on paper rather than on screen. It is puzzling that paper spreadsheets seemed to take longer to debug. Interestingly, as in a previous study [16], only about half of all of the errors are found, suggesting that there is room (and need) for substantial improvement. Also, providing printed formulas in an integrated manner (contiguous with the spreadsheet values) appears to reduce the number of false positives (correct items flagged as errors) when compared with all other treatments. Finally, self-reported strategies for error detection appear to affect the number of errors found; the best advice seems to be to pay close attention to linkages between subtotals and their source data.

Speculation might help us interpret the unexpected results. The reason for the time results, which contradicted our paper-versus-screen hypothesis, might be that the paper treatments led subjects to try harder to find the errors. Extra effort needed for reading from the screen might have tired subjects and induced them to give up earlier. The lack of support for the formulas versus no-formulas hypothesis might indicate that spreadsheet auditors do not rely on formulas, perhaps merely because it takes as long and works as effectively to inspect formulas as it takes to check the math in each computed cell using a calculator or using quick mental calculations.

The integrated treatment seemed to furnish modest improvement in error checking by reducing identification of correct cells as errors. However, it did not provide significant aid in finding more errors. The new treatment should be studied further through an extensive set of studies using more subjects, different spreadsheets, and/or different types of errors. Such a study would be needed before we could draw well-founded conclusions.

The tiny size (by typical standards) of our spreadsheet might have made it impossible to discern any major facilitative effects of formula contiguity (because the formulas were already quite accessible, given the small size). Some caution is in order for enlarging the spreadsheet. If the overall difficulty of the problem were increased (a result of increasing its size), subjects could cluster at a very low level of performance, attenuating artificially any differences between cells. It would then be difficult to achieve significance in comparisons among cells.

Our exploratory strategy results are very encouraging for linking behavior and performance variables in predicting or facilitating error finding. Future research should focus on error-detection strategies, examine other parts of our conceptual model, and employ larger sample sizes to enable simultaneous evaluation of several independent variables. A better understanding of the spreadsheet error-finding process might help avoid further headlines reporting significant spreadsheet disasters.

NOTES

Please contact the first author at galletta@vms.cis.pitt.edu for a copy of the experimental materials.

1. Pentium is a registered trademark of Intel, Inc.
2. Lotus 1–2–3 is a registered trademark of Lotus Development Corporation.
3. Excel is a registered trademark of Microsoft, Inc.
4. Windows is a registered trademark of Microsoft, Inc.
5. Analysis of covariance for each was also run. The GMAT score was the most significant covariate. The GMAT score correlated strongly with the time required to perform the task ($r = -0.26, p < 0.01$) and with the number of errors found ($r = 0.17, p < 0.05$). One-tailed probabilities are shown. More important, the ANOVA results were not altered by introducing any of the covariates.
6. For number of errors found, $F(1,108) = 0.42; p = 0.517$; for time taken, $F(1,109) = 0.42; p = 0.520$.
7. For number of errors found, $F(1,111) = 1.54; p = 0.217$; for time taken, $F(1,110) = 1.04; p = 0.311$.
8. The equal variance test was marginal at $F(1,111) = 3.326; p = 0.071$. Assuming unequal variances, $t(56.7)=2.25$; one-tailed $p = 0.014$. Assuming equal variances, $t(111) = 1.62$; one-tailed $p = 0.053$.

REFERENCES

1. Alavi, M.; Nelson, R.R.; and Weiss, I.R. Strategies for end-user computing: an integrative perspective. *Journal of Management Information Systems, 4,* 3 (Winter 1987–88), 28–49.

2. Anderson, K., and Bernard, A. Spreading mistakes. *Journal of Accounting and EDP. 4,* 2 (Fall 1988), 42–45.

3. Ballou, D.P.; Pazer, H.L.; Belardo, S.; and Klein, B. Implications of data quality for spreadsheet analysis. *Data Base, 18,* 3 (Spring, 1987), 13–19.

4. Basili, V.R., and Selby, R.W., Jr. Four applications of a software data collection and analysis methodology. In J.K. Skwirzynski (ed.), *Software System Design Methods.* Berlin: Springer-Verlag, 1986.

5. Bishop, Y.M.M.; Fienberg, S.E.; and Holland, P.W. *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, MA: MIT Press, 1975.

6. Brancheau, J.C.; Vogel, D.R.; and Wetherbe, J.C. An investigation of the information center from the user's perspective. *Data Base, 17,* 1 (Fall 1985), 4–18.

7. Brown, P.S., and Gould, J.D. An experimental study of people creating spreadsheets. *ACM Transactions on Office Information Systems, 5,* 3 (May 1987), 258–272.

8. Cale, E.G., Jr. Quality issues for end-user developed software. *Journal of Systems Management, 45,* 1 (January 1994), 36–39.

9. Card, S.; Moran, T.; and Newell, A. *The Psychology of Human-Computer Interaction.* Hillsdale, NJ: Laurence Erlbaum, 1983.

10. Cragg, P.B., and King, M. Spreadsheet modeling abuse: an opportunity for OR? *Journal of the Operational Research Society, 44,* 8 (August 1993), 743–752.

11. Creeth, R. Microcomputer spreadsheets: their use and abuse. *Journal of Accountancy, 159,* 6 (June 1985), 92.

12. Davies, N., and Ikin, C. Auditing spreadsheets. *Australian Accountant* (December 1987), 54–56.

13. Dillon, A. Reading from paper versus screens: a critical review of the empirical literature. *Ergonomics, 35,* 10 (October 1992), 1297–1326.

14. Ditlea, S. Spreadsheets can be hazardous to your health. *Personal Computing, 11,* 1 (January 1987), 60–69.

15. Floyd, B.D.; Walls, J.; and Marr, K. Managing spreadsheet model development. *Journal of Systems Management, 46,* 1 (May–June 1995), 38–43, 68.

16. Galletta, D.F.; Abraham, D.; El Louadi, M.; Lekse, W.; Pollalis, Y.; and Sampler, J. An empirical study of spreadsheet error-finding performance. *Accounting, Management, and Information Technologies, 3,* 2 (February 1993), 79–95.

17. Galletta, D.F.; Hartzel, K.S.; Johnson, S.; Joseph, J.L.; and Rustagi, S. An experimental study of spreadsheet presentation and error detection. *Proceedings of the 29th Hawaii International Conference on System Sciences.* Maui: IEEE Computer Society Press, 1996, pp. 336–345.

18. Galletta, D.F., and Heckman, R. A role theory perspective on end-user development. *Information Systems Research, 1,* 2 (June 1990), 168–187.

19. Galletta, D.F., and Hufnagel, E. A model of end-user computing policy: determinants of organizational compliance. *Information and Management, 22,* 1 (January 1992), 1–18.

20. Gould, J.D.; Alfaro, L.; Barnes, V.; Finn, R.; Grischkowsky, N.; and Minuto, A. Reading is slower from CRT displays than from paper: attempts to isolate a single-variable explanation. *Human Factors, 29,* 3 (March 1987), 269–299.

21. Gould, J.D.; Alfaro, L.; Finn, R.; Haupt, B.; and Minuto, A. Reading from CRT displays can be as fast as reading from paper. *Human Factors, 29,* 5 (May 1987), 497–517.

22. Hall, M.J.J. A risk and control oriented study of the practices of spreadsheet application developers. *Proceedings of the Twenty-Ninth Hawaii International Conference on System Sciences.* Maui: IEEE Computer Society Press, 1996, pp. 364–373.

23. Hayen, R.L., and Peters, R.M. How to ensure spreadsheet integrity. *Management Accounting, 70,* 10 (October 1989), 30–33.

24. Hendry, D.G., and Green, T.R.G. Creating, comprehending, and explaining spreadsheets: a cognitive interpretation of what discretionary users think of the spreadsheet model. *International Journal of Human–Computer Studies, 40,* 6 (June 1994), 1033–1065.

25. Janvrin, D., and Morrison, J. Factors influencing risks and outcomes in end-user development. *Proceedings of the Twenty-Ninth Hawaii International Conference on System Sciences.* Maui: IEEE Computer Society Press, 1996, pp. 346–355.

26. Kee, R.C., and Mason, J.O., Jr. Preventing errors in spreadsheets. *Internal Auditor, 45,* 1 (February 1988), 42–47.

27. King, M.; Lee, R.A.; Piper, J.A.; and Whittaker, J. *Information Technology and the Management Accountant.* London: CIMA, 1991.

28. Lederer, A.L., and Sethi, V. Planning for end-user computing: top-down or bottom-up? *Information Management Review, 2,* 2 (Fall 1986), 25–34.

29. McGoldrick, J.A.; Martin, J.; Bergering, A.J.; and Symons, S. Locating discrete information in text: effects of computer presentation and menu formatting. *Journal of Reading Behavior, 24,* 1 (January 1992), 1–20.

30. Munro, M.C.; Huff, S.L.; and Moore, G. Expansion and control of end-user computing. *Journal of Management Information Systems, 4,* 3 (Winter 1987–88), 5–27.

31. Myers, G.J. A controlled experiment in program testing and code walkthroughs/inspections. *Communications of the ACM, 21,* 9 (September 1978), 760–768.

32. Oliver, R. Proof-reading on paper and screens: the influence of practice and experience on performance. *Journal of Computer-Based Instruction, 20,* 4 (April 1994), 118–124.

33. Olson, J.R., and Nilsen, E. Analysis of the cognition involved in spreadsheet software interaction. *Human–Computer Interaction, 3,* (1987–88), 309–349.

34. Panko, R.R. Hitting the wall: development and debugging errors in a "simple" spreadsheet problem. *Proceedings of the 29th Hawaii International Conference on System Sciences*. Maui: IEEE Computer Society Press, 1996, pp. 356–363. [URL: http://www.cba.hawaii.edu/PANKO/risks/h6wall.htm.]

35. Panko, R.R. Spreadsheets on trial: a survey of research on spreadsheet risks. *Proceedings of the 29th Hawaii International Conference on Systems Sciences*. Maui: IEEE Computer Society Press, pp. 326–335. [URL: http://www.cba.hawaii.edu/PANKO/research/risks/h6trial.htm.]

36. Panko, R.R., and Halverson, R.H., Jr. Patterns of errors in spreadsheet development I: quantitative errors. Working Paper, Department of Decision Sciences, College of Business Administration, University of Hawaii, March 1995. [URL: ftp://splicer2.cba.hawaii.edu/panko/sserrors/sswpmar.doc]

37. Panko, R.R., and Halverson, R.H. Are two heads better than one (at reducing spreadsheet errors)? *Proceedings of the 2nd AIS Americas Conference on Information Systems*. Phoenix: AIS, 1996, pp. 69–70.

38. Pearson, R. Lies, damned lies, and spreadsheets. *Byte, 13,* 13 (December 1988), 299–304.

39. Rockart, J.F., and Flannery, L.S. The management of end-user computing. *Communications of the ACM, 26,* 10 (October 1983), 776–784.

40. Schultheis, R., and Sumner, M. The relationship of application risks to application controls: a study of microcomputer-based spreadsheet applications. *Journal of End User Computing, 6,* 2 (Spring 1994), 11–18.

41. Speier, C., and Brown, C.V. Perceived risks and management actions: differences in end-user application development across functional groups. *Proceedings of the Twenty-Ninth Hawaii International Conference on System Sciences*. Maui: IEEE Computer Society Press, 1996, pp. 374–384.

42. Stang, D. Spreadsheet disasters I have known. *Information Center, 3,* 11 (November 1987), 26–30.

43. Stone, D.N., and Black, R.L. Using microcomputers: building structured spreadsheets. *Journal of Accountancy, 168,* 4 (October 1989), 130–142.

44. Xenakis, J.W. 1–2–3: a dangerous language. *Computerworld, 21,* 10 (March 9, 1987), 31, 34.

APPENDIX A: The Spreadsheet used in the Task

	A	B	C	D	E	F	G
1	Cash Budget		Assuming	2%	inflation/semester		
2				Fall	Spring	Summer	Overall
3	Cash, beginning			$1,000	$1,360	$673	
4	Outflows - School costs			$4,468	$4,474	$4,480	$13,422
5	Living costs			$4,172	$4,213	$4,485	$12,870
6	Inflows - Loans			$3,000	$3,000	$3,000	$9,000
7	Support from home			$6,000	$5,000	$6,000	$17,000
8	Cash, end			$1,360	$673	$980	$3,013
9				========	=======	=======	========
10	School (contractual): Tuition			$4,115	$4,115	$4,115	$12,345
11	Fees			$53	$53	$53	$159
12	School (other): books/supplies			$300	$306	$312	$918
13	Living (contractual)	MONTHLY	MONTHS				
14	Housing	$450	4	$1,800	$1,800	$1,800	$5,400
15	Insurance	$53	4	$212	$212	$212	$636
16	Living Costs (other)						
17	Food	$330	4	$1,320	$1,346	$1,373	$4,039
18	Entertainment	$150	4	$600	$612	$624	$1,836
19	Transportation	$40	4	$160	$161	$164	$485
20	Clothing	$21	4	$80	$82	$84	$250

APPENDIX B: New Formula-Integrated Printed Spreadsheet

	A	B	C	D	E	F	G	
1	**Cash Budget**		**Assume:**	**2%**	**inflation/semester**			1
2				**Fall**	**Spring**	**Summer**	**Overall**	2
3	**Cash, beginning**			**$1,000**	**$1,360** +D8	**$673** +E8		3
4	**Outflows - School costs**			**$4,468** @SUM(D10..D12)	**$4,474** @SUM(E10..E12)	**$4,480** @SUM(F10..F12)	**$13,422** @SUM(D4..F4)	4
5	**Living costs**			**$4,172** @SUM(D14..D20)	**$4,213** @SUM(E14..E20)	**$4,485** @SUM(F12..F19)	**$12,870** @SUM(D5..F5)	5
6	**Inflows - Loans**			**$3,000**	**$3,000**	**$3,000**	**$9,000** @SUM(D6..F6)	6
7	**Support from home**			**$6,000** @ROUND(1000 +D4+D5-D3-D6,-3)	**$5,000** @ROUND(1000 +E4+E5-E3-E6,-3)	**$6,000** @ROUND(1000 +F4+F5-F3-F6,-3)	**$17,000** @SUM(D7..F7)	7
8	**Cash, end**			**$1,360** +D3-D4-D5+D6+D7	**$673** +E3-E4-E5+E6+E7	**$980** +F3-F4-E5+F6+F7	**$3,013** @SUM(D8..F8)	8
9				=========	=========	=========	========	9
10	**School (contractual): Tuition**			**$4,115**	**$4,115**	**$4,115**	**$12,345** @SUM(D10..F10)	10
11	**Fees**			**$53**	**$53**	**$53**	**$159** @SUM(D11..F11)	11
12	**School (other): books/supplies**			**$300**	**$306** @ROUND (+D12*(1+$D$1),0)	**$312** @ROUND (+E12*(1+$D$1),0)	**$918** @SUM(D12..F12)	12
13	**Living (contractual)**	MONTHLY	MONTHS					13
14	**Housing**	**$450**	**4**	**$1,800** +C14*B14	**$1,800** +D14	**$1,800** +D14	**$5,400** @SUM(D14..F14)	14
15	**Insurance**	**$53**	**4**	**$212** +C15*B15	**$212** +D15	**$212** +D15	**$636** @SUM(D15..F15)	15
16	**Living Costs (other)**							16
17	**Food**	**$330** 11*30	**4**	**$1,320** +C17*B17	**$1,346** @ROUND (+D17*(1+$D$1),0)	**$1,373** @ROUND (+E17*(1+$D$1),0)	**$4,039** @SUM(D17..F17)	17
18	**Entertainment**	**$150**	**4**	**$600** +C18*B18	**$612** @ROUND (+D18*(1+$D$1),0)	**$624** @ROUND (+E18*(1+$D$1),0)	**$1,836** @SUM(D18..F18)	18
19	**Transportation**	**$40**	**4**	**$160** +C19*B19	**$161** @ROUND (+D19+(1+$D$1),0)	**$164** @ROUND (+E19*(1+$D$1),0)	**$485** @SUM(D19..F19)	19
20	**Clothing**	**$21**	**4**	**$80** +C20*20	**$82** @ROUND (+D20*(1+$D$1),0)	**$84** @ROUND (+E20*(1+$D$1),0)	**$250** @SUM(C20..F20)	20
	A	B	C	D	E	F	G	

APPENDIX C: Spreadsheet Formulas

A:A1: {Page 1/1} (G) [W21] 'Cash Budget
A:C1: {Page} (G) [W7] 'Assume:
A:D1: {Page} (P0) 0.02
A:E1: {Page} (G) ' inflation/semester
A:D2: "Fall
A:E2: (G) "Spring
A:F2: (G) "Summer
A:G2: (G) "Overall
A:A3: {Page} (G) [W21] 'Cash, beginning

```
A:D3: (C0) 1000
A:E3: (C0) +D8
A:F3: (C0) +E8
A:A4: {Page} (G) [W21] 'Outflows—School costs
A:D4: (C0) @SUM(D10..D12)
A:E4: (C0) @SUM(E10..E12)
A:F4: (C0) @SUM(F10..F12)
A:G4: (C0) @SUM(D4..F4)
A:A5: {Page} (G) [W21] ' Living costs
A:D5: (C0) @SUM(D14..D20)
A:E5: (C0) @SUM(E14..E20)
A:F5: (C0) @SUM(F12..F19)
A:G5: (C0) @SUM(D5..F5)
A:A6: {Page} (G) [W21] 'Inflows—Loans
A:D6: (C0) 3000
A:E6: (C0) 3000
A:F6: (C0) 3000
A:G6: (C0) @SUM(D6..F6)
A:A7: {Page} (G) [W21] ' Support from home
A:D7: (C0) @ROUND(1000+D4+D5-D3-D6,–3)
A:E7: (C0) @ROUND(1000+E4+E5-E3-E6,–3)
A:F7: (C0) @ROUND(1000+F4+F5-F3-F6,–3)
A:G7: (C0) @SUM(D7..F7)
A:A8: {Page} (G) [W21] 'Cash, end
A:D8: (C0) +D3-D4-D5+D6+D7
A:E8: (C0) +E3-E4-E5+E6+E7
A:F8: (C0) +F3-F4-E5+F6+F7
A:G8: (C0) @SUM(D8..F8)
A:D9: (C0) \=
A:E9: (C0) \=
A:F9: (C0) \=
A:G9: (C0) \=
A:A10: {Page} (G) [W21] 'School (contractual): Tuition
A:D10: (C0) 4115
A:E10: (C0) 4115
A:F10: (C0) 4115
A:G10: (C0) @SUM(D10..F10)
A:A11: {Page} (G) [W21] ' Fees
A:D11: (C0) 53
A:E11: (C0) 53
A:F11: (C0) 53
A:G11: (C0) @SUM(D11..F11)
A:A12: {Page} (G) [W21] 'School (other): books/supplies
```