Pergamon

Accting., Mgmt. & Info. Tech. 9 (1999) 141–160

ACCOUNTING MANAGEMENT AND INFORMATION TECHNOLOGIES

www.elsevier.com/locate/amit

Spreadsheet development and 'what-if' analysis: quantitative versus qualitative errors

Thompson S.H. Teo *, Margaret Tan

Department of Decision Sciences, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore

Received 1 January 1997; received in revised form 1 March 1999

Abstract

Past research has shown that errors are relatively common in all types of spreadsheets. As spreadsheets are used widely by executives in analyzing and supporting their decision making, especially in financial analysis, budgeting and forecasting applications, it is important for spreadsheets to be accurate. Errors undetected in spreadsheets may have undesirable consequences. For example, errors may adversely impact the firm's competitiveness or profitability when the costing of projects is prone to incorrect computation. For this purpose, we investigate the types of errors that may occur even for simple domain-free spreadsheet problems. In addition, we also show that spreadsheet errors are difficult to detect during 'what-if' analysis (i.e. when some design parameters are changed) when spreadsheets are not properly designed. The results show that most students do not take due care in designing spreadsheets. It appears that the techniques in teaching spreadsheets should really focus on how to design a comprehensive spreadsheet that is both easy to maintain and debug rather than just demonstrating the many features of spreadsheets.

Keywords: End user computing; Spreadsheet; Modeling; Error; 'What-if' analysis

1. Introduction

The use of spreadsheets is prevalent among end-users as a fundamental tool in business analyses. The American Management Association found that spreadsheet

* Corresponding author. Tel.: +65-874-3036; fax: +65-779-2621.
E-mail addresses: fbateosh@nus.edu.sg (T.S.H. Teo), fbatanm@nus.edu.sg (M. Tan)

PII: S0959-8022(99)00006-5

applications were used by an overwhelming 91% of a sample of end-users, thereby making it the most common computer application in business (American Management Association, 1988). End-users usually use spreadsheet models to assist them in making decisions. As errors in spreadsheets can result in incorrect or 'less than optimal' decisions being made, such decisions can have undesirable consequences in terms of the inappropriate choice of business strategies or may be very critical in positioning the company's strategy. It can also adversely impact the firm's bottom-line. Further, spreadsheet errors can result in losses ranging from hundreds of thousands (Kee & Mason, 1988) to millions (Ditlea, 1987) of dollars. It is, therefore, important to ensure that spreadsheets are built correctly.

Numerous researchers and practitioners in recent years (Brown & Gould, 1987; Cragg & King, 1993; Crawford, 1994; Panko & Halverson, 1997) have consistently highlighted the problem of spreadsheet errors. This growing concern over spreadsheet errors can be attributed to the increasing popularity of spreadsheets to support important financial analysis, budgeting, and forecasting applications. Also, there appears to be a tendency for end-users to view spreadsheets as simple tools and to be overconfident about the error-free nature of their spreadsheets (Floyd, Walls & Marr, 1995). In fact, Panko and Sprague (1996) emphasized that previous studies on spreadsheet errors have always found error rates that would be unacceptable in practice. Indeed, the growing concern over errors in spreadsheets should be seriously addressed, as most executives do not really check or verify the accuracy or validity of the spreadsheets before they use the solutions. Compounding the problem is the fact that inadequate care is taken to design spreadsheets, which in turn makes error detection and correction even more difficult.

2. Research objectives

Considering the seriousness of how errors in spreadsheets could impact decision making, the main thrust of this paper is to examine the different types of errors found in spreadsheets. Thus, this paper describes a study that was conducted to investigate the different types of spreadsheet errors in solving a simple domain-free spreadsheet problem. The use of a domain-free spreadsheet problem would help to rule out threats to validity caused by differences in task domain knowledge. The first objective of the study is to confirm Panko's (Panko, 1996) results that spreadsheet errors are still relatively common even for spreadsheet exercises that require little domain knowledge. This confirmation is necessary to validate further the scientific investigation, as Panko's study appears to be the first study (or at least one of the few studies) that uses a domain-free spreadsheet.

The second objective is to examine the different types and nature of errors in spreadsheet design. Examining the different types and nature of errors will provide us with a better understanding of the common errors when using spreadsheets. Also, the different types of errors are important as they may require different strategies to mitigate them during spreadsheet designs.

The third objective of the study is to examine in-depth the changes in error rates (if any) during 'what-if analysis' (i.e. changing some of the design parameters of the initial model). As 'what-if' analyses are relatively common in practice, examining such changes in error rates will provide some insights as to the types of errors that may be made during changes to the spreadsheet model. Hence, this investigation will provide us with a more informed approach to understand the circumstances in which the errors occurred.

Finally, this study also aims to examine the intercorrelations between different types of errors. This is to enable us to determine which types of errors are associated with each other, for example, whether jamming errors tend to be positively associated with logic errors. By understanding these errors as well as their relationships or associations with each other, techniques in teaching spreadsheets as well as emphasis on various aspects of design can be better revised and improved accordingly.

3. Literature review

Panko and Halverson (1996) carried out an extensive review of research on spreadsheet risks and errors, and proposed that issues in spreadsheet risks can be better understood in terms of the life cycle stage, the research issues, and the methodology. Based on their reviews of existing literature, prior research on spreadsheet errors can be summarized in three main streams.

The first research stream seeks to examine spreadsheet errors at different life cycle stages. Intuitively, spreadsheet error rates can be expected to vary across the life cycle since more errors are usually detected as the model moves from initial requirements through to operational use. Various field audits and experiments have been carried out in the different life cycle stages, i.e. the cell-entry stage (Olson & Nilsen, 1987–1988), the draft stage (Janvrin & Morrison, 1996; Panko, 1996), the debugging stage (Galletta, Abraham, El Louadi, Leske, Pollalis & Sampler, 1993; Panko & Sprague, 1996) and the operational stage (Davies & Ikin, 1987; Hall, 1996). Generally, these studies have found that spreadsheet errors are indeed common and are often not detected.

The second stream of research examines policies on spreadsheet development and implementation issues. Generally, such research has shown that only a very small percentage of companies have explicit, written policies on spreadsheet development, implementation and testing (Cale, 1994; Cragg & King, 1993; Hall, 1996). Given the important nature of spreadsheets in facilitating better business decision-making, these findings are disturbing. One plausible explanation is that many companies view spreadsheets as simple tools and believe that spreadsheet errors are usually minor or can be easily detected. As such, corporate controls are usually informal or nonexistent. Another plausible explanation may be that many executives tend to view errors in spreadsheets as merely careless typographical errors rather than errors caused by inadequate policies on spreadsheet development and implementation.

The third stream of research examines user work practices and seeks to better understand the processes involved in creating spreadsheets. For example, the extent of planning prior to building the spreadsheets, and the extent of testing the models may vary among different organizations. Generally, researchers have found that pre-planning, documentation and post-development debugging are not being widely practised (Cragg & King, 1993; Hall, 1996; Schultheis & Sumner, 1994) in building spreadsheets. Also, user controls in spreadsheets development are usually less detailed than other applications developed within the organization. Again, the reason may be that spreadsheets are usually viewed as simple tools, thus detailed development procedures are unnecessary.

4. Types of errors

Different classification schemes can be used for the types of errors found in spreadsheets. The distinction between the different types of errors is important as they may require different types of error reduction or correction methods. In general, they can be classified in terms of quantitative and qualitative errors. Quantitative errors usually result in numerical errors that may result in incorrect bottom-line values while qualitative errors usually take the form of poor spreadsheet design in terms of duplicating information or in terms of placing the values of more than one variable in a single cell.

Panko and Halverson (1995) classified quantitative errors in term of three main types, namely, mechanical, logic, and omission errors. Mechanical errors are simple slips such as mistyping a number, pointing to a wrong cell address, or selecting an incorrect range of values or cells. These errors may be due to carelessness, mental overload or distractions. Logic errors are more complex and they are due to incorrect formulas caused by incorrect algorithms to solve a problem. For example, adding profit margin only to variable costs (instead of variable plus fixed costs) to compute sales needed for a target profit is a logic error. Logic errors are generally more difficult to detect than mechanical errors as they often require an understanding of the cognitive processes of the spreadsheet developer. For example, if there are several ways to arrive at the solution, it is pertinent that one needs to understand the cognitive processes of the spreadsheet developer in order to determine whether the various steps used to arrive at the solution are appropriate or correct. Omission errors tend to result from leaving something out of the model that should have been there. For example, failure to include overhead costs in total costs is an omission error. Similar to logic errors, omission errors can be difficult to detect.

Qualitative errors may not immediately result in incorrect bottom-line values but they can make debugging difficult. They can also lead to quantitative errors being made during the 'what-if' analysis (Panko & Halverson, 1996). Furthermore, the presence of qualitative errors may make maintenance of the spreadsheets difficult, and this is especially problematic for the creation of alternative solutions or scenarios of the problem.

Panko (1988) suggested two laws that govern good spreadsheet modeling. The first law specifies that a variable should be defined in only one of two ways: either the variable is defined by a number or it is defined by a formula not containing any numbers. Violation of the first law results in 'jamming' where values of more than one variable are placed in a single cell, for example, values of length and height are placed in a single cell instead of two separate cells.

The second law of modeling specifies that information of the variable must never be duplicated or be repeated in a model, for example, the values of length and breadth are repeated in adjacent cells. Violation of the second law may result in data inconsistency. We will use violations of these two laws as measures of qualitative errors.

5. Research method

A field study in terms of an experiment conducted in a classroom setting was used to collect data to investigate the research objectives.

5.1. Sample and procedure

The sample consisted of 176 undergraduate business students taking a second year course in information systems at a large university in Singapore. All these students have the following courses in their first year: mathematics and quantitative analysis, accounting and an introductory computer course. In the introductory computer course, the students were taught the fundamentals of spreadsheet design using Microsoft Excel, and the use of computation and problem solving methods. Students were not taught how to look for errors in spreadsheets.

To collect data for the study, the two exercises were given as part of the students' class or term assignments for the information systems course that they were taking. A survey packet containing the first spreadsheet exercise together with a post-experiment questionnaire was distributed to students. Students were briefed that their participation would provide valuable feedback to assist IS lecturers teaching spreadsheets in the first year introductory IS course to improve on their course design. The students completed the exercise on their own at home. They then returned the questionnaire together with a diskette containing their spreadsheet exercise 2 weeks later.

A second spreadsheet exercise was given to the same students 2 weeks after they had completed the first exercise. The purpose was to allow the students to carry out 'what-if' analysis in order to determine whether existing errors were corrected and/or new errors were made. Diskettes containing the first exercise were returned to each of these students and they were asked to work on the second exercise during class time. The students were again asked to complete a similar post-experiment questionnaire.

The rationale for completing the first exercise at home was that we wanted to give the students sufficient time to read and understand the spreadsheet exercise, as well as to be able to work at their own pace. The second exercise was carried out in the classroom since we wanted to observe them as they made changes to their spreadsheets as well as limit the time taken. This would perhaps simulate, to a certain extent, situations in the real world where 'what-if' analyses are often carried out under time constraints.

5.2. The questionnaire

The descriptions of the two spreadsheet exercises are given in Appendix A. The first spreadsheet exercise was a close adaptation of the wall problem used by Panko (1996). Instead of naming the two types of walls as brick and lava rock, we named them brick and concrete in order to be closer to common terminology for building materials used in Singapore. Other aspects of the spreadsheet exercise, for example, the numerical values of the variables, were the same as Panko's. The spreadsheet exercise was relatively simple since it did not require any specific domain knowledge (e.g. accounting expertise). Only basic knowledge such as computation of volume of wall (given length, breadth and height), labor costs (given the amount of manpower and wage rate) and bid amount (given profit margin) were required.

Two questionnaires were used in conjunction with the spreadsheet experiment. The pre-experiment questionnaire captured details of students' demographic characteristics as well as their perceptions of their own level of spreadsheet expertise. The post-experiment questionnaire captured the following information on their perceptions of the spreadsheet tasks on a seven-point Likert scale: whether the exercise was difficult, whether the exercise was stressful, whether the exercise was interesting, whether their background knowledge was sufficient, and whether time passed by slowly during the exercise.

The second spreadsheet exercise required students to perform 'what-if' analysis by changing some of the design parameters in their spreadsheet model from the first exercise. A total of seven design parameters were changed during 'what-if' analysis. The post-experiment questionnaire was similar to that used in the first exercise. Students were also asked to record the time taken for each of the spreadsheet exercises.

5.3. Error determination

5.3.1. Quantitative errors

Panko's (Panko, 1996) model solution for the spreadsheet exercise on the wall problem was used as the basis for assessing errors (see Appendix B). Similar to Panko's methodology, we compared the subject's solution to the model solution. If there was an error in the bottom-line value, we reviewed the formula codes in the spreadsheet and changed them until the correct bottom-line values were obtained. We then recorded the different types of quantitative errors.

5.3.2. Qualitative errors

In addition to determining quantitative errors, we also counted the number of violations of Panko's (Panko, 1988) two laws of spreadsheet modeling. This gave us an indication of the number of qualitative errors due to poor spreadsheet design. To be consistent with quantitative errors, we selected only those spreadsheets with incorrect bottom-line values in the first and/or second exercises. For jamming errors (law 1 violation), we counted the additional number of variables that were jammed into a single cell. For example, if length, height and thickness of the wall were entered into a single cell as 20*6*2, the number of jamming errors was recorded as two since there were two additional variables in the cell. For duplication errors (law 2 violation), we counted the number of cells with duplicated or repeated information.

6. Results and discussion

The results are organized as follows. First, the demographic characteristics of respondents are presented. Second, spreadsheet errors in terms of both quantitative and qualitative errors, cell error rate, details of quantitative errors, and intercorrelations between different types of errors are examined. Third, the results of the post-experiment questionnaire are discussed.

6.1. Demographic characteristics

Five spreadsheet models were rejected as the students failed to complete both exercises. Another three responses were rejected as the students failed to complete both sets of post-experiment questionnaires. Hence, a total of 168 valid responses were used in the analyses.

Table 1 presents the demographic characteristics of the students. There are more female than male students, and this trend tends to reflect the approximate distribution of gender among business students in the university. The mean age is 20.8 years, which is typical of second year students. The number of years of spreadsheet experience is about 1.35 years, which is expected as our sample consists of second year business administration students who learned spreadsheets as part of their first year studies. The mean level of spreadsheet expertise measured in terms of a seven-point Likert scale (1='Poor', 7='Excellent') is 3.61, which shows that respondents are generally neutral about their spreadsheet ability. One possible reason is that students learned about spreadsheets in their first year and may have used spreadsheets only in a few other courses. Another possible reason could be that students are generally modest about their spreadsheet ability, which may be a typical response for those of Asian culture.

Table 1
Students' demographic profile

	Number	Percent
Sex		
Male	58	34.52
Female	110	65.48
	Mean	**SD**
Age	20.78	1.18
Experience with Excel (years)	1.35	0.92
Expertise (7-pt scale)	3.61	1.11

6.2. Spreadsheet errors

Table 2 presents the different types of spreadsheet errors found in this study. Of the 168 spreadsheets developed, 70 spreadsheets had errors in the first exercise and 84 had errors in the second exercise, thereby giving an error rate of 41.7% and 50%, respectively (compared with Panko's 38%, and Panko and Sprague's 35%). These error rates are generally consistent with previous studies on spreadsheet errors.

Note that in the first exercise, we had 70 spreadsheets with errors. When we did the second exercise, some errors were corrected but new errors were also introduced. In other words, there were 13 corrected errors, meaning that these spreadsheets had incorrect answers in the first exercise but correct answers in the second exercise. This also means that 57 (that is, $70-13=57$) had incorrect values in both exercises.

On the other hand, spreadsheets with correct answers in the first exercise but incorrect answers in the second exercise totaled 27. This figure takes into account what is reflected as the difference between the total number of spreadsheets with errors in the first and second exercise plus corrected errors ($84-70+13=27$). In other words, the total of 27 takes into account the fact that some errors were corrected and some new errors were made.

Although the fraction of spreadsheets with errors was high, there were actually very few errors per spreadsheet, i.e. only 0.5 (86 errors out of 168 models) and 0.8 (128 errors out of 168 models) errors per model for the first and second spreadsheet exercises, respectively. The per-spreadsheet error rate for the first exercise is slightly higher than that obtained by Panko (0.4) and Panko and Sprague (0.42). Hence, it appears that the problem with spreadsheet development is not the absolute number of errors. Rather, it is that minor errors can cascade down into errors in bottom-line values. This can be quite serious as managers do not verify or validate the values and often rely on bottom-line values in the model to make decisions.

6.2.1. Quantitative errors

As shown in Table 2, the error rates for mechanical and logic errors were not significantly different in the two exercises. In contrast, the number of omission errors for the second exercise was significantly different from the first exercise. Note that the percentages for corrected and new errors for quantitative errors do not tally with those found in the row 'spreadsheets with errors' because the former refers to individual errors while the latter refers to errors in bottom-line values. Since spreadsheets with incorrect bottom-line values can have different numbers and types of errors, the percentages found under quantitative errors are not directly comparable with those shown in the row 'spreadsheet with errors' in Table 2.

The results show that mechanical errors are most easily detected, with about 40% of these errors made in the first exercise being corrected in the second exercise. In contrast, logic and omission errors are less easily detected, with about 30% of these errors made in the first exercise being corrected in the second exercise. It appears that very few students (13 out of 70 students with incorrect bottom-line values) actually found and corrected their errors, which were made in the first spreadsheet exercise. It seems that the majority assumed that their first spreadsheet model was

Table 2
Results of spreadsheet errors

	1st exercise		2nd exercise		t	Corrected errors		New errors	
	No. (1)	Percent	No. (2)	Percent		No. (3)	Percent (3)/(1)	No. (4)	Percent (4)/(2)
Spreadsheets without errors	98	58.33	84	50.00					
Spreadsheets with errors	70	41.67	84	50.00	2.24*[c]	13	18.57	27	32.14
Total	168	100.00	168	100.00					
Quantitative errors									
Mechanical errors	23	26.74	27	21.09	0.82	9	39.13	13	48.75
Logic errors	23	26.74	23	17.97	0.00	7	30.43	7	30.43
Omission errors	40	46.51	78	60.93	3.92***	13	32.50	51	65.38
Total quantitative errors	86	100.00	128	100.00		29		71	
Qualitative errors									
Jamming errors	549	71.02	465	65.59	−3.15	111	20.22	27	5.81
Duplication errors	224	28.98	244	34.41	2.35	5	2.23	25	10.25
Total qualitative errors	773	100.00	709	100.00		116		52	
Cell error rates (CER)									
CER for mechanical errors		0.55		0.64					
CER for logic errors[a]		0.98		0.98					
CER for omission errors		0.95		1.86					
CER (quantitative errors)		2.05		3.05					
CER for jamming errors[b]	390	9.29	337	8.02					
CER for duplication errors		5.33		5.81					
CER (qualitative errors)		14.62		13.83					

[a] Computed on the basis of formula cells.
[b] Based on the number of error cells.
[c] $*p<0.05$; $***p<0.001$ (two-tailed).

correct and thus, error-free. Consequently, students concentrated on changing the new parameters in the 'what-if' analysis, rather than first checking and confirming whether the original model was correct. In other words, code inspection was not usually performed prior to 'what-if' analysis. Note that we deliberately did not tell students to perform code inspection on their spreadsheet models prior to 'what-if' analysis, as we wanted to see what they would normally do during 'what-if' analysis.

There is an increase in the number of spreadsheets with errors in the second exercise, as shown in Table 2. The new quantitative errors (as percentage of their respective error categories) comprised about 50% of mechanical errors, 30% of logic errors, and 65% of omission errors. Since the second exercise involved changing a few design parameters, it was perhaps expected that the number of new logic errors might be lower than that of mechanical or omission errors. One plausible explanation for the relatively high number of new mechanical and omission errors in the second exercise may be due to the students performing the 'what-if' analysis under time constraints. Consequently, carelessness resulting in mechanical slips and omissions may be more common.

6.2.2. Qualitative errors

Table 2 also shows the number of qualitative errors. There are about two-thirds jamming errors and one-third duplication errors. It is interesting to note that jamming errors occur more frequently than duplication errors for both exercises. Also, jamming errors are more easily corrected during 'what-if' analysis compared to duplication errors. This is evident in Table 2 that shows that 111 (20.22%) jamming errors were corrected compared to 5 (2.23%) duplication errors. One likely explanation is that students making jamming errors in the first spreadsheet exercise may find it harder to perform 'what-if' analysis because of the difficulty in distinguishing the various variables 'jammed' into a single cell. Consequently, in the second exercise, they may take appropriate measures to reduce the number of jamming errors.

The results also show that the percentage of new duplication errors (10.25%) is greater than new jamming errors (5.81%). One likely explanation is that during 'what-if' analysis, students may deliberately duplicate information in the spreadsheet model in order to make it easier for them to locate and change design parameters. Hence, it appears that there may be a trade-off between jamming and duplication errors where the reduction of one type of error may lead to an increase in the other type of error.

6.3. Cell error rate

Panko (1996) proposed that the cell error rate (CER) — the number of errors per hundred model cells — would be a more valuable indicator than the absolute number of errors itself. This is because the CER may help professional spreadsheet developers anticipate the number of errors that the spreadsheet is likely to contain at the end of the draft stage. The CER for the first and second exercises are 2.05% (compared to 1.7% by Panko, and Panko and Sprague) and 3.05%, respectively. These values are close to the error rates of 3–7% found in code inspection studies of programming errors (Panko & Halverson, 1995).

In terms of qualitative errors, the CER for the first and second exercises are 9.29% and 8.02%, respectively. These values are much higher than that of quantitative errors, thereby implying that good spreadsheet design is often lacking in the building of spreadsheet models.

6.4. Details of quantitative errors

Table 3 shows the categories of errors according to the three types of quantitative errors.

From Table 3, mechanical errors in terms of incorrect hours per laborer per day and incorrect cell address/range are the most common. It seems that these errors are caused mainly by typing wrongly either the values of the variables or the cell addresses and ranges. Logic errors mainly occur in the form of computing labor costs twice while omission errors mainly involve the failure to include the number of laborers per team. Although it is plausible that this omission error may actually be a logic error, we felt that since the number of laborers per team was not even mentioned in the spreadsheet model, it is more an omission error than a logic error. Also, the failure to change various parameters in the second exercise is likely to be an omission error. Interestingly, forgetting to change the number of laborers per team and the wall thickness appears to be common.

6.5. Intercorrelations of the different types of errors

Table 4 presents the intercorrelations of the different types of errors. Jamming errors are significantly positively correlated with logic errors for both exercises. This implies that spreadsheet design with jamming errors may tend to increase the number of logic errors. One plausible explanation is that when more than one variable is jammed into a single cell, it may be harder to visualize the correct algorithm to solve the problem, thereby resulting in more logic errors.

In contrast, duplication errors for both exercises are significantly negatively correlated with mechanical errors in the first exercise only. This is an unexpected but interesting result as it implies that the duplication of information on a spreadsheet may actually decrease the number of mechanical errors. Intuitively, we would expect that the additional typing caused by duplicating information would lead to additional errors. One possible reason for the opposite effect is that when information is duplicated, it may result in greater ease of finding information in the model. Another explanation can be derived from Lerch (1988) who found that pointing to cells in a different column produces higher error rates compared to pointing to cells in the same column. Intuitively, it is likely for duplication of information on the spreadsheet to occur more frequently in the same column to facilitate ease of computation. This result appears to suggest that, with duplication, there may be more opportunities for pointing to cells in the same column (rather than in different columns), thereby resulting in fewer mechanical errors.

Omission errors do not appear to have any relationship with either jamming or duplication errors. This is quite surprising since for spreadsheets with jamming

Table 3
Details of quantitative errors

Type of error	1st exercise		2nd exercise	
	No.	Percent	No.	Percent
Mechanical errors	*23*	*26.74*	*27*	*21.09*
Incorrect hours per laborer per day	9	10.47	8	6.25
Incorrect cell address/range	6	6.98	7	5.47
Incorrect number of laborers per team	2	2.33	2	1.56
Incorrect thickness of wall	2	2.33	1	0.78
Incorrect overhead percentage	2	2.33	1	0.78
Incorrect cost per cubic foot	1	1.16	4	3.13
Incorrect profit margin	1	1.16	2	1.56
Others (occurring once each)	0	0.00	2	1.56
Logic errors	*23*	*26.74*	*23*	*13.69*
Added labor costs twice	9	10.47	7	5.47
Added profit to material cost not total cost	3	3.49	3	2.34
Added brick and concrete into single bid	3	3.49	2	1.56
Divide instead of multiply to get material cost	3	3.49	1	0.78
Multiplied twice by crew size	2	2.33	2	1.56
Bid price as profit margin	1	1.16	3	2.34
Others (occurring once each)	2	2.33	5	3.91
Omission errors	*40*	*46.51*	*78*	*46.43*
Omitted # laborers per team	21	24.42	10	7.81
Forgot to add total to get bid	5	5.81	5	3.91
Omitted overheads	4	4.65	4	3.13
Omitted labor costs	3	3.49	2	1.56
Omitted hours per laborer	2	2.33	1	0.78
Forgot to change wall thickness	NA	NA	12	9.38
Forgot to change # laborers per team	NA	NA	10	7.81
Forgot to change unit material cost	NA	NA	7	5.47
Forgot to change hours per laborer	NA	NA	7	5.47
Forgot to change profit margin	NA	NA	7	5.47
Forgot to change overheads	NA	NA	5	3.91
Forgot to change # days	NA	NA	3	2.34
Others (occurring once each)	5	5.81	5	3.91

Table 4
Intercorrelation matrix of errors

Quantitative errors	Qualitative errors			
	First exercise		Second exercise	
	Jamming	Duplication	Jamming	Duplication
First exercise				
Mechanical error	0.137	−0.267**[a]	0.028	−0.264**
Logic errors	0.275**	−0.040	0.265**	0.001
Omission errors	−0.190	0.019	−0.105	0.016
Second exercise				
Mechanical errors	0.149	−0.172	−0.071	−0.185
Logic errors	0.228*	−0.014	0.239*	0.036
Omission errors	−0.078	0.036	0.032	0.054

[a] *$p<0.05$; **$p<0.01$ (two-tailed).

errors, it may be difficult to find and change design parameters which have been combined into a single cell, thereby perhaps resulting in omission errors. Similarly, it is possible that duplication of information may clutter up the spreadsheet, thereby leading to a greater tendency for respondents to forget to change some information. Further research is needed to examine in greater detail the relationship between qualitative and quantitative errors.

6.6. Profile of perceptions of the spreadsheet exercises

Table 5 presents the post-experiment results in terms of capturing the students' perceptions of the spreadsheet tasks in this study. We asked the students to record the time taken to perform each spreadsheet exercise. The time taken to complete the first spreadsheet exercise is generally greater than that for the second exercise. A paired sample *t*-test revealed significant time differences between the two spreadsheet exercises. This result is expected since the second exercise is carried out in the computer laboratory under time constraints. Furthermore, the second exercise only requires students to change some of the design parameters in their initial spreadsheet model.

In terms of the questions in the post-experiment questionnaires, we also carried out paired sample *t*-tests to compare differences in perceptions between the two exercises. These perceptive measures enable us to obtain some feedback about how students felt about the exercises (which may be useful for determining if modifications to the exercises are needed in future experiments). In addition, the *t*-tests help us to determine whether students felt differently about the two exercises. Significant results would prompt us to probe for possible explanations. The results showed that students found both exercises to be easy, relaxing, slightly boring, and that time passes quickly. In addition, the students felt that their background was sufficient as evident from the mean scores of below 4 for both exercises (1='sufficient', 7='insufficient'). Despite this, the paired sample *t*-test showed significant differences in this item in that students were less sure of whether their background was sufficient in the second exercise compared to the first exercise. One possible reason could be that the second exercise was carried out under time constraints, whereas the first exercise was carried out in their free time without any time constraint. Another reason could be that since the second exercise requires some basic knowledge of 'what-if' analysis, students may have been uncertain of their skills in this area.

7. Limitations

The limitation of the study is the use of the undergraduate students in the spreadsheet experiment. This is because students may not be representative of actual spreadsheet developers in commercial organizations. While this may be true to a certain extent, it is not possible to engage a large group of practitioners to take part in this kind of experiment. However, students as surrogates for practitioners do provide a

Table 5
Profile of the perceptions of the spreadsheet exercises

	First exercise Mean (SD)	Second exercise Mean (SD)	*t*-value
Time taken to complete	20.40 (10.38)	13.10 (8.14)	−7.69***[a]
Whether exercise was: (1=easy, 7=difficult)	2.73 (1.13)	2.88 (1.41)	1.32
Whether exercise was: (1=stressful, 7=relaxing)	5.04 (1.35)	4.84 (1.43)	−1.63
Whether exercise was: (1=interesting, 7=boring)	4.05 (1.28)	4.21 (1.16)	1.30
Whether background was: (1=sufficient, 7=insufficient)	2.60 (1.30)	3.06 (1.50)	3.70***
Whether time passed: (1=quickly, 7=slowly)	3.16 (1.17)	3.22 (1.23)	0.51

[a] ***$p<0.001$.

suitable sampling source to control for extraneous variables such as age, educational level, and expertise that may confound the results. Further, since our sample comprises business students, they will be potential users/developers of spreadsheet models in their working careers. As previous research (e.g. Panko, 1996; Panko & Sprague, 1996) used undergraduates and/or MBAs with varying spreadsheet experience and found no significant differences in error rates across different groups, our results should be generalizable to other spreadsheet users. Thus, the limitation on external validity is not severe.

8. Conclusions

This paper makes a significant contribution to existing literature in two ways. First, although qualitative errors in terms of duplication and jamming errors have been discussed in the literature, to the authors' knowledge, this paper is the first to actually measure it. Second, this paper examines the important issue of the correction and introduction of errors during 'what-if' analysis. This issue is important because the usefulness of spreadsheet models is often derived from their ability to perform 'what-if' analysis or scenario building solution.

The results show that it is important to understand that errors in spreadsheets can result in new errors being made during 'what-if' analysis. In this regard, understanding the frequency and causes of the types of spreadsheet errors is fundamental in ensuring the validity and reliability of spreadsheets.

The results of this study show that spreadsheet errors are indeed common. Among the quantitative errors, mechanical errors appear to be most easily detected and corrected. However, it is disturbing that although some errors are corrected during 'what-if' analysis, the number of new errors made is actually similar (for logic errors) or greater (for mechanical and omission errors) than the number of errors corrected.

Among the qualitative errors, jamming errors occur more frequently than duplication errors. It is reassuring that the number of jamming errors that are corrected during 'what-if' analysis is greater than the number of new jamming errors. In contrast, very few duplication errors are corrected during 'what-if' analysis. In fact, the number of new duplication errors is about five times that of the number of duplication errors that were corrected.

Jamming errors are positively correlated with logic errors while duplication errors are negatively correlated with mechanical errors. Neither jamming nor duplication errors appear to have any significant relationship with omission errors. Further research is needed to examine in greater detail the relationship between quantitative and qualitative errors.

Because very few errors were detected during the 'what-if' analysis, it is disturbing to note that there appears to be a tendency for the students to assume that their first model is correct. This suggests that there is a need to emphasize teaching of good spreadsheet design when providing spreadsheet instructions. The current emphasis on teaching spreadsheets seems to be more on the mechanics of building spreadsheets and their usefulness rather than the concept of proper spreadsheet design with adequate documentation. The concept of the cell, its purpose and applications should be emphasized. In building a spreadsheet model to address business problems with the hope of providing feasible solutions, it is important to note that spreadsheet design and development ought to have similar quality assurance as in the design and development of other business applications.

Hence, proper design procedures similar to programming and systems development can perhaps be adopted for spreadsheet development. For example, the systems development life cycle (SDLC) can actually be used as a systematic approach for spreadsheet development (Goss, Dillon & Kendrick, 1989). Design of spreadsheets should also be in logically related blocks (Davies & Ikin, 1987). This is important as most, if not all, corporate budgeting spreadsheets consist of many interrelated functional sub-blocks. There is little doubt that debugging and peer review or spreadsheet checking should be made an inherent part of spreadsheet development.

Because business undergraduate students are likely to make use of spreadsheets during their working lives, they should be made aware of the potential hazards of poor spreadsheet design. The importance of policies and guidelines in spreadsheet development should also be highlighted. Students should cultivate good habits and practices in spreadsheet design and development. In addition, spreadsheet users must realize that errors are often not easily detected and that new errors can be easily introduced during 'what-if' analysis. They should be taught to view completed spreadsheets, regardless of whether the spreadsheets are built by themselves or by others, with a healthy dosage of scepticism.

Acknowledgements

We thank the editor and anonymous reviewers for their insightful comments and suggestions about the paper. We also thank Dennis Galletta and Ray Panko for sharing with us their papers and experimental materials.

Appendix A

A.1. First spreadsheet exercise

You are to build a spreadsheet model to help you create a bid to build a wall. You will offer two options — concrete or brick.

Both walls will be built by teams of two laborers. Each team will work three 8-hour days to build either type of wall.

The wall will be 20 feet long, 6 feet tall, and 2 feet thick.

Wages will be $10 per hour per laborer. You will have to add 20% to wages to cover other overheads (e.g. CPF).

The concrete wall will cost $3 per cubic foot. The brick wall will cost $2 per cubic foot.

Your bid must add a profit margin of 30% to your expected cost.

Time taken to complete spreadsheet exercise: ______ minutes.

A.2. Second spreadsheet exercise

Both walls will be built by teams of three laborers. Each team will work four 7-hour days to build either type of wall.

The wall is 1.5 feet thick.

Add 25% for overheads to wages.

The brick wall will cost $2.80 per cubic foot.

Your bid must add a profit margin of 35% to your expected cost.

All other conditions remain the same as in the previous problem (see below if you need to refer to previous problem).

Time taken to complete spreadsheet exercise: ______ minutes

Appendix B. Solution for first exercise (adapted from Panko, 1996)

B.1. Labor hours

Days per person	3
Hours per day	8
Hours per person	24
People	2
Hours	48

B.2. Labor costs

Pay per hour	$10
Pay	$480
Fringe benefit rate	20%
Fringe benefits	$96
Labor costs	$576

B.3. Wall volume

Height	6
Length	20
Thickness	2
Volume	240

B.4. Cost of materials

	Brick	Concrete
Cost per cubic foot	$2	$3
Cost of materials	$480	$720

B.5. Total cost and bid

	Brick	Concrete
Labor+materials	$1056	$1296
Profit margin	30%	
Markup	316.8	388.8
Bid	$1373	$1685

References

AMA (1988). *Report on end-user and departmental computing.* New York: American Management Association.

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

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

Cragg, P. G., & King, M. (1993). Spreadsheet modelling abuse: an opportunity for OR? *Journal of the Operational Research Society*, *44*, 743–752.

Crawford, M. D. (1994). Re: Microsoft and Lotus spreadsheet errors. *The Risks Digest*, *16*.

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

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

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

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

Galletta, D. F., Hartzel, K. S., Johnson, S., Joseph, J., & Rustagi, S. (1996–1997). Spreadsheet presentation and error detection: An experimental study. *Journal of Management Information Systems, 13* (Winter), 45–63.

Goss, E., Dillon, T., & Kendrick, J. (1989). Bittersweet spreadsheets: application development needs control. *The Woman CPA*, *July*, 20–24.

Hall, M. J. J. (1996). A risk and control oriented study of the practices of spreadsheet application developers. In *Proceedings of the 29th Hawaii International Conference on System Sciences* (pp. 364–373). Maui, Hawaii, 4–7 January. IEEE Computer Society Press.

Janvrin, D., & Morrison, J. (1996). Factors influencing risks and outcomes in end-user development. In *Proceedings of the 29th Hawaii International Conference on System Sciences* (pp. 346–355). Maui, Hawaii, 4–7 January. IEEE Computer Society Press.

Lerch, F. J. (1988). *Computerized financial planning: discovering cognitive difficulties in model building*. Unpublished Ph.D. dissertation, University of Michigan, Ann Arbor.

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

Olson, J. R., & Nilsen, E. (1987–1988). Analysis of the cognition involved in spreadsheet interaction. *Human Computer Interaction*, *3*, 309–349.

Panko, R. R. (1988). *End user computing: management, applications, and technology* (1st ed). New York: John Wiley.

Panko, R. R. (1996). Hitting the wall: Errors in developing and debugging in a 'simple' spreadsheet problem. In *Proceedings of the 29th Hawaii International Conference on System Sciences* (pp. 356–363). Maui, Hawaii, 4–7 January. IEEE Computer Society Press.

Panko, R. R., & Halverson, R. P., Jr. (1995). Patterns of errors in spreadsheet development I: quantitative errors. Working Paper, Department of Decision Sciences, College of Business Administration, University of Hawaii.

Panko, R. R., & Halverson, R. P., Jr. (1996). Spreadsheets on trial: A survey of research on spreadsheet risks. In: *Proceedings of the 29th Hawaii International Conference on System Sciences* (pp. 326–335). Maui, Hawaii, 4–7 January. IEEE Computer Society Press.

Panko, R. R., & Halverson, R. P. Jr. (1997). Are two heads better than one? (at reducing errors in spreadsheet modeling?). *Office Systems Research Journal*, *15* (1), 21–32.

Panko, R. R., & Sprague, R. H. Jr. (1996). Hitting the wall: Errors in developing and code inspecting a 'simple' spreadsheet model. *Decision Support Systems*, *22*, 337–353.

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