THE USE OF INFLUENCE DIAGRAMS IN THE DESIGN OF SPREADSHEET MODELS: AN EXPERIMENTAL STUDY

Peter A. O'Donnell
Decision Support Systems Laboratory
Monash University
Melbourne, Australia
peter.odonnell@sims.monash.edu.au

ABSTRACT

The spreadsheet is one of the most commonly used personal computer applications. Many studies have shown that a very high percentage of spreadsheet models contain serious errors. The use of influence diagrams, which are a visual representation of the relationships among the variables in a problem, to plan the development of a spreadsheet model may reduce the likelihood of the spreadsheet model containing an error. This paper describes an experimental study that tests whether the use of influence diagrams improves spreadsheet modelling practice. The results of the study suggests that influence diagrams do help users to better understand problems, and as a result reduce the likelihood of omitting important variables from spreadsheet models.

INTRODUCTION

The spreadsheet is one of the most commonly used personal computer applications (Caine & Robson, 1993a, 1993b; Panko, 1988; Sperling, 1997). Spreadsheet models are now an integral part of the information analysis and decision making activity of many managers. Tens of millions of spreadsheet models are created and used each year. However, both survey and experimental investigations have found that a significant proportion of spreadsheet models contain serious errors (Brown & Gould, 1987; Edge & Wilson, 1990; Hall, 1996; Hayen & Peters, 1989; Panko & Halverson, 1994; Panko & Sprague, 1998; Simkin, 1987; Teo & Tan, 1999). Cragg & King (1993) state that empirical data suggests that 25% of all spreadsheets used in business contain some kind of flaw or bug. One metric used to measure spreadsheet errors is the cell error rate. This statistic is usually expressed as the number of errors per 100 cells. Cell error rates of between 1 and 5 errors per 100 cells are commonly reported (Panko, 1996). At first glance these cell error rates may seem low. However, it is typical for one cell in a spreadsheet to depend on the value in another which depends upon the value of another and so on. If one of the base level cells contains an error, that error is propagated through all the cells that depend on it. These cell cascades, as Panko (1996) calls them, mean that a low cell error rate will lead to a high probability of the spreadsheet model calculating incorrectly.

Influence diagramming is a modelling technique intended to aid the process of conceptual modelling. An influence diagram is a visual representation of the variables in a problem and of the relationships between these variables. If the developer of a spreadsheet model develops an influence diagram and then uses that influence diagram as the basis of their spreadsheet design, the resulting spreadsheet model may be of a higher quality than would otherwise be the case. This paper describes an experimental study intended to test this theory.

SPREADSHEETS

Spreadsheet software represents an easy to use and deceptively powerful class of modelling tool. Whilst, spreadsheets are best suited to financial models, they are also used in a wide variety of model types and application domains. Spreadsheets are commonly used to create simulation, optimisation, forecasting, statistical, network analysis, transport and inventory models (Flaherty, Morgan, Lombardo, de Silva, & Wilson, 1999; Ragsdale, 1998). Perhaps the most important function provided by a spreadsheet is the ability to perform "what-if" analysis. As a user changes the value in one cell, all the cells whose formulas make use of that cell are updated. All the cells that refer to those cells are then also updated and so on as the change propagates down the cell cascade. The effect of a single change to a cell might cause the instantaneous recalculation of an entire model. Ralph Kimball refers to this functionality as the 'twinkling' lights of a spreadsheet (Nardi & Miller, 1991). The visual and interactive nature of the recalculation of a spreadsheet encourages users to change inputs and investigate different problem scenarios.

Spreadsheet models are very easy to develop. Many users, with little or no formal training, can learn to quickly build complex models. The development environment, being visual and interactive, encourages users to make "on-the-fly" changes to models. Typically users will create formulas by entering them directly into the spreadsheet as they think about the problem they are working on. The cells of the spreadsheet, arranged in rows and columns, become the conceptual model (Landry, Malouin, & Oral, 1983) used by the developer to organise their thoughts about the problem.

Spreadsheets are a unique class of software in that they are the only software application in which end-users actually construct their own programs. Formulas can be quickly and easily created by moving the cursor around the screen and by using menus to access built-in functions. Unfortunately spreadsheets don't stop end-users from

making errors or help them find them. The spreadsheet doesn't know when a number has been input incorrectly or when a formula has variable missing. Whilst it is possible to create some input error trapping devices within a spreadsheet and some applications include rudimentary debugging facilities, these seem to be little used and ineffective.

Spreadsheet models containing errors have been used as the basis of significant decisions with disastrous and embarrassing consequences. For example, government officials in Australia are often forced to admit mistakes in spreadsheet models and the information released that is based upon them. When a mistake is discovered, a correction can be published however much political embarrassment is caused (for example, Kemp, 1996-97). In one case a Senate committee was told that a particular court case involving three judges had cost $818,00. Later it was discovered that a triplication of the costs associated with the judges had occurred, the actual costs were only $493,106 (Soden, 1996). Commercial organisations are reluctant to allow knowledge of spreadsheet errors and their effects to become widely known. Panko has published a bibliography of the few publicly reported instances of spreadsheet errors in commercial organisations (Panko, 2000). In one startling and widely reported case a data entry error caused a general ledger to be in error by $2.4 billion dollars. The company, an investment funds manager, announced that the fund involved had made a profit of $2.3 billion when it had actually lost $0.1 billon. (Godfrey, 1995; Savitz, 1994). The mistake was not noticed for several months.

A well structured and easy to understand system is just as easy to construct as a system that is poorly structured, difficult to understand and unreliable. However, most end-user developers don't understand the value of systems development techniques that create reliable maintainable systems. As a result, error-ridden spreadsheet models are common. The paradox of the use of spreadsheets is that the ease of use and power of the interactive modelling environment is also the reason why so many spreadsheets are developed with errors.

Ronen, Pally, & Lucas (1989) and also Hayen & Peters (1989) argue that the frequency and severity of spreadsheet errors can be reduced by the use of a structured approach to spreadsheet design. The proposals of Ronen et al., (1989) include a development life cycle for spreadsheet models and a spreadsheet flow diagramming technique similar to data flow diagrams that are used in the development of traditional information systems. In a field experiment Janvrin & Morrison (2000), have shown that a structured approach to spreadsheet design can help lower the number of mistakes made developing formulas that link different spreadsheets. Others, for example Edge & Wilson (1990) and Simkin (1987), have developed rules and procedures that should be followed to ensure the accuracy, consistency and completeness of spreadsheet models. Despite experimental studies that have shown that spreadsheet errors are difficult for people to detect and correct (Galletta, Hartzel, Johnson, Joseph, & Rustagi, 1996; Panko & Sprague, 1998) it is generally recognised that spreadsheets should be audited and tested using a formal process (Berglas & Hoare, 1999; Whittaker, 1999). An audit of a complex set of spreadsheet-based models developed by the Australian Department of Immigration to examine the effect of immigration on the government's budget revealed over 250 problems and issues which were corrected before the model was used (----, 1999). However, Hall's (1996) review of Australian spreadsheet development practice suggests that, in practice, few models are ever formally audited or tested.

Few tools or techniques are available to help model a problem prior to the development of a spreadsheet. Those tools that exist aren't used much due to the ad hoc nature of most spreadsheet development and the time-pressure that most spreadsheet developers are under. A design technique is required is that is easy for spreadsheet users to learn, quick to use, that helps to conceptualise a problem, and is relevant to spreadsheet design. The influence diagram seems to have these attributes.

Spreadsheet errors

One measure of the quality of a spreadsheet is to determine the number and type of errors present (Brown & Gould, 1987; Galletta et al., 1996; Panko & Halverson, 1994; Panko & Sprague, 1998; Teo & Tan, 1999). Cragg, Grant, & King (1991) developed a simple classification of error types found in spreadsheet models. Their classification was used to investigate the nature of errors found in operational spreadsheets. This error classification has been used in this study. The five major types in their classification are: erroneous formulae, incorrect ranges, omitted factors, data input errors and incorrect use of functions.

Erroneous formula: This means that the error exists within a formula. A typical example of this error is when the wrong cell co-ordinates are entered when composing the formula. For example, =A5+A6 may have been entered when the formula =A5+A7 was intended. Errors of this kind are very common (Brown & Gould, 1987; Chussil, 1988; Ronen et al., 1989).

Incorrect ranges: This error involves the incorrect specification of a range of cells in a formula. It is a very easy error to make and can be very difficult to detect. An example of this error would be a user entering =sum(B10:B16) when the formula =sum(B10:B15) was intended. As in this example, this error is associated with the use of functions like SUM and AVERAGE that use ranges of cells as an input (Ronen et al., 1989).

Omitted factors: This is an error in the logic of the spreadsheet (Chussil, 1988). It is committed whenever a variable is left out of a model or formula.

Data input errors: This category of errors is used to describe situations where invalid or incorrect data has been entered. For example a user may have entered 5,000 into a cell when 50,000 was intended.

Incorrect use of functions: The incorrect use of built-in functions can also introduce errors into spreadsheet models. For example, the incorrect use of the rounding function to format output to a certain number of decimal places can introduce errors when the numbers are used in other calculations (Brown & Gould, 1987).

INFLUENCE DIAGRAMS

Influence diagrams are directed graphs whose nodes represent the variables of interest in some problem and whose directed arcs represent the influence of the variables on each other (Holtzman, 1989). Figure 1 illustrates a simple influence diagram which consists of two nodes X and Y and one arc directed from node X to node Y. This diagram should be interpreted as indicating that the value of X has some influence on the value of Y (Owen, 1978), that is Y = f(X). Some prefer the term 'relevance' to 'influence' (Howard, 1990) arguing the arc directed from node X to node Y should be read as 'The value of the variable X is relevant when determining the value of the variable Y. The visual depiction of the influence diagram does not allow the specification of the nature of the relationship between nodes. However, it does allow the specification of the variables of interest when determining the value of a particular node.
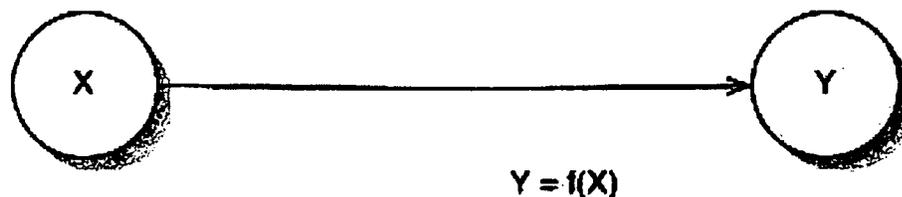

Figure 1. A Simple Influence Diagram (Owen, 1978)

The value of an influence diagram, as emphasised by Bodily (1988), is the visual representation of the relationships among the variables of interest in a problem. This visual representation aids communication about, and discussion and understanding of, the structure of a model.

Spreadsheets and influence diagrams

Any spreadsheet model can be expressed as an influence diagram. Alternatively, it is possible to represent a problem that is to be modelled using a spreadsheet as an influence diagram. That influence diagram can then be used to as an aid in the design of a spreadsheet model. The diagram specifies what variables are to be represented in the model. Each variable shown on the influence diagram model must be represented by a cell in the spreadsheet. The directed arcs of the influence diagram specify which variables are involved in which spreadsheet formula.

Figure 2 shows the influence diagram that represents the cell C2 on a spreadsheet if it contained the formula =A2/B2. In Figure 2, the cells A2 and B2 contain data input by the user and are represented on the associated influence diagram as border nodes. The formula for the cell C2, which contains references to the cells A2 and B2 indicates that arcs from the nodes representing those cells should be directed to the node representing the cell C2. Figure 3 shows a typical simple spreadsheet model (for break-even/profit analysis) and its representation as an influence diagram.
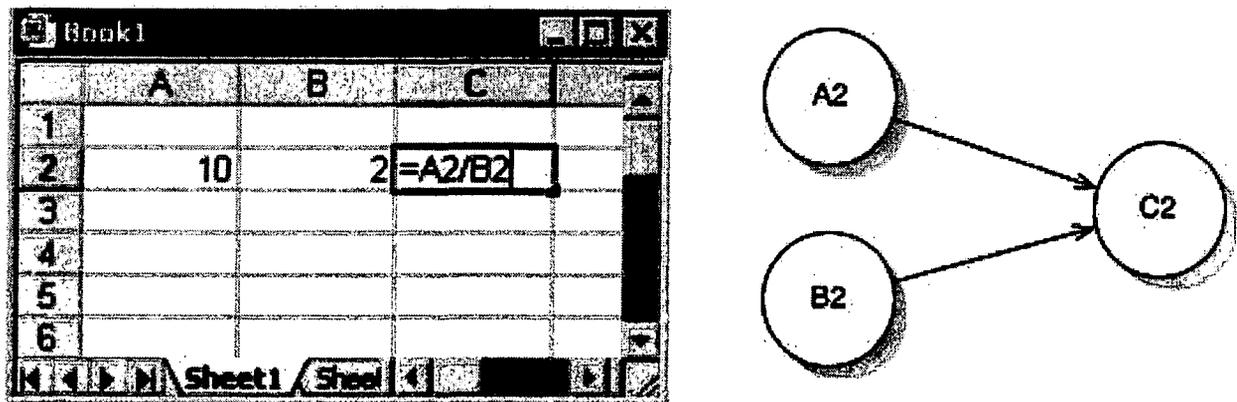
Figure 2. A simple spreadsheet formula and its influence diagram representation.

There are several potential advantages to developing an influence diagram prior to beginning to develop a spreadsheet model. The influence diagram is easily understood and as a result, errors or misconceptions can be identified more easily than they could be if they were incorporated into spreadsheet code. The influence diagram is easier to manipulate and change than a spreadsheet model. It can be argued that the spreadsheet model is less likely to contain a serious error if it has been based upon a carefully developed influence diagram due to the conceptualisation of the problem afforded by the influence diagram.

METHOD

This experiment involved the development of a spreadsheet model of in a laboratory setting. A schematic representation of the experimental method is shown in Figure 4. One group of subjects was asked to develop a spreadsheet model to help solve a problem. This group was the control group. Another group of subjects was asked to first develop an influence diagram based model of the same problem, and then develop a spreadsheet-based model. The models developed by the two groups were compared and analysed for errors.

Experimental procedure

The subjects in the experiment were third year students studying a decision support systems elective within an IT degree at an Australian University. These subjects should have had similar knowledge of the development of spreadsheet based models and of influence diagramming. A study of spreadsheet modelling and influence diagramming is included in the decision support systems elective in which they were enrolled. Some data about the background of the subjects was collected and used in the analysis. At the end of the experimental procedure the subjects were also asked to complete a questionnaire designed to gauge their reaction to the experimental procedure.

Briefly the procedure was as follows: After volunteering to participate in the experiment, the subjects were randomly assigned, by drawing lots, into two groups. Each group of subjects, accompanied by an "experimenter", were taken to a personal computer laboratory. Each laboratory was similarly configured, providing the subjects with access, via a network server, to the spreadsheet application with which they were familiar (Lotus 1-2-3). Each subject was given a kit containing their instructions, the problem they were to work on, a numbered work pad for drawing influence diagrams (not included in the kit for the control group), a short numbered questionnaire, and a numbered disk. After completing the exercise, subjects saved their spreadsheet model on the disk provided and returned their kit to the experimenter. The experimenter then gave each subject a post-procedure questionnaire. After completing their post-procedure questionnaire the subjects were thanked by the experimenter and left the laboratory.

A total of 49 students participated in the experiment. The random allocation of each student into either the control or experimental group unfortunately didn't produce two even sized groups. A total of 29 subjects were allocated to the control group and 20 were allocated to the experimental group. This number of subjects is not uncommon in spreadsheet research. Sample sizes in previous studies have ranged from 9 (Brown & Gould, 1987) to 113 (Galletta et al., 1996).
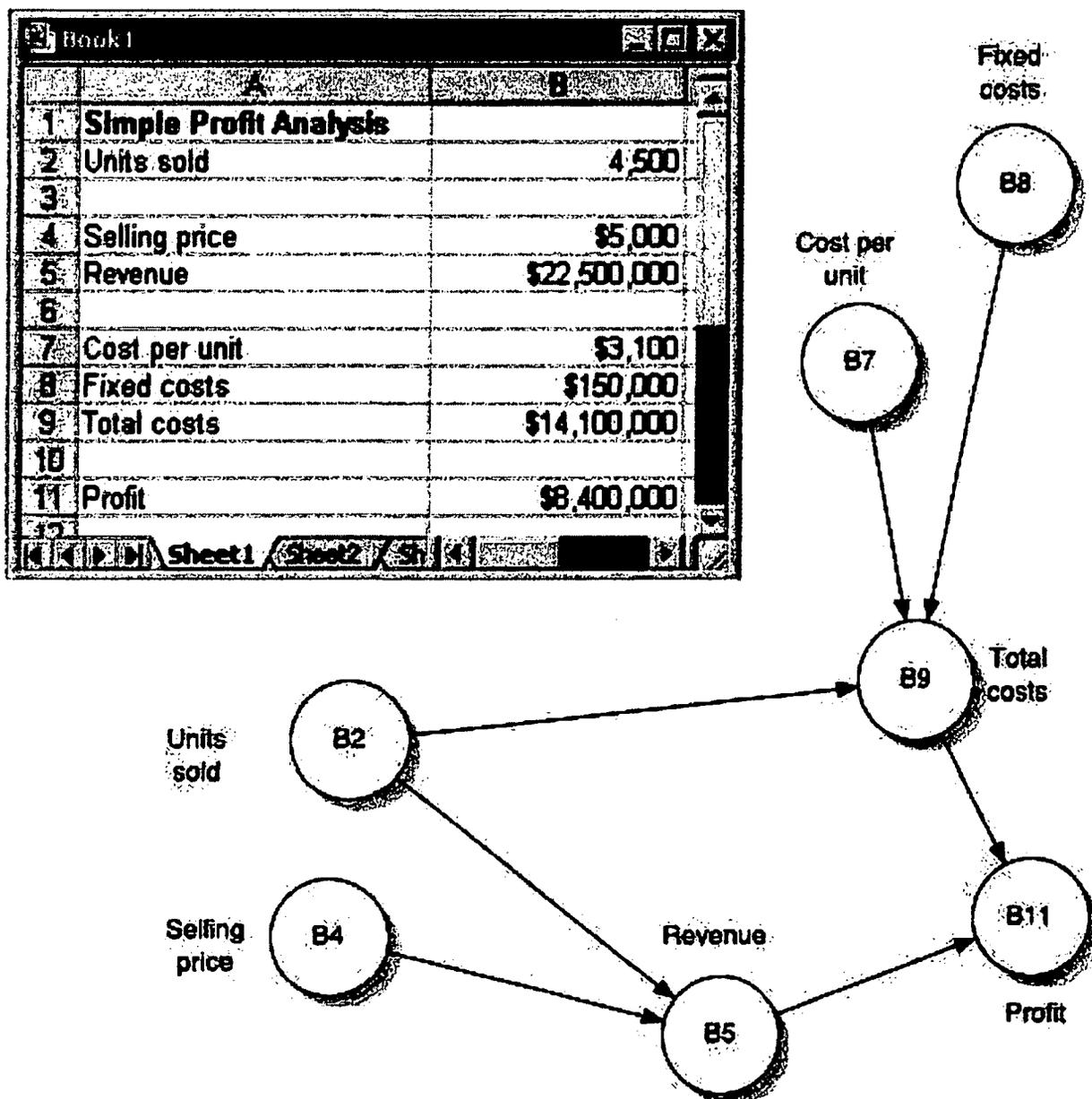
Figure 3. A typical (simple) spreadsheet model and its influence diagram representation.

Task

The target problem used in this work is a standard problem that has been used in the experimental testing of the role of different factors on the quality of spreadsheets (Panko & Halverson, 1994). A description of the problem given to the subjects is provided in the appendix. Panko (1996) reflecting on his earlier study thought it highly likely that many of the spreadsheet errors that were made by subjects working on this problem would have been caused by their lack of domain knowledge. The task assumes that the subjects have basic spreadsheet knowledge and know enough about accounting to construct a pro-forma income statement. In conducting a follow up experiment he ensured that the task used was simple in that no specialist domain knowledge was required. In this second experiment a high rate of errors was again found. The subjects in this experiment had previously completed a spreadsheet assignment that required the development of a similar pro-forma income statement. In addition to this all had studied at least one introductory accounting subject as part of there IT degree.
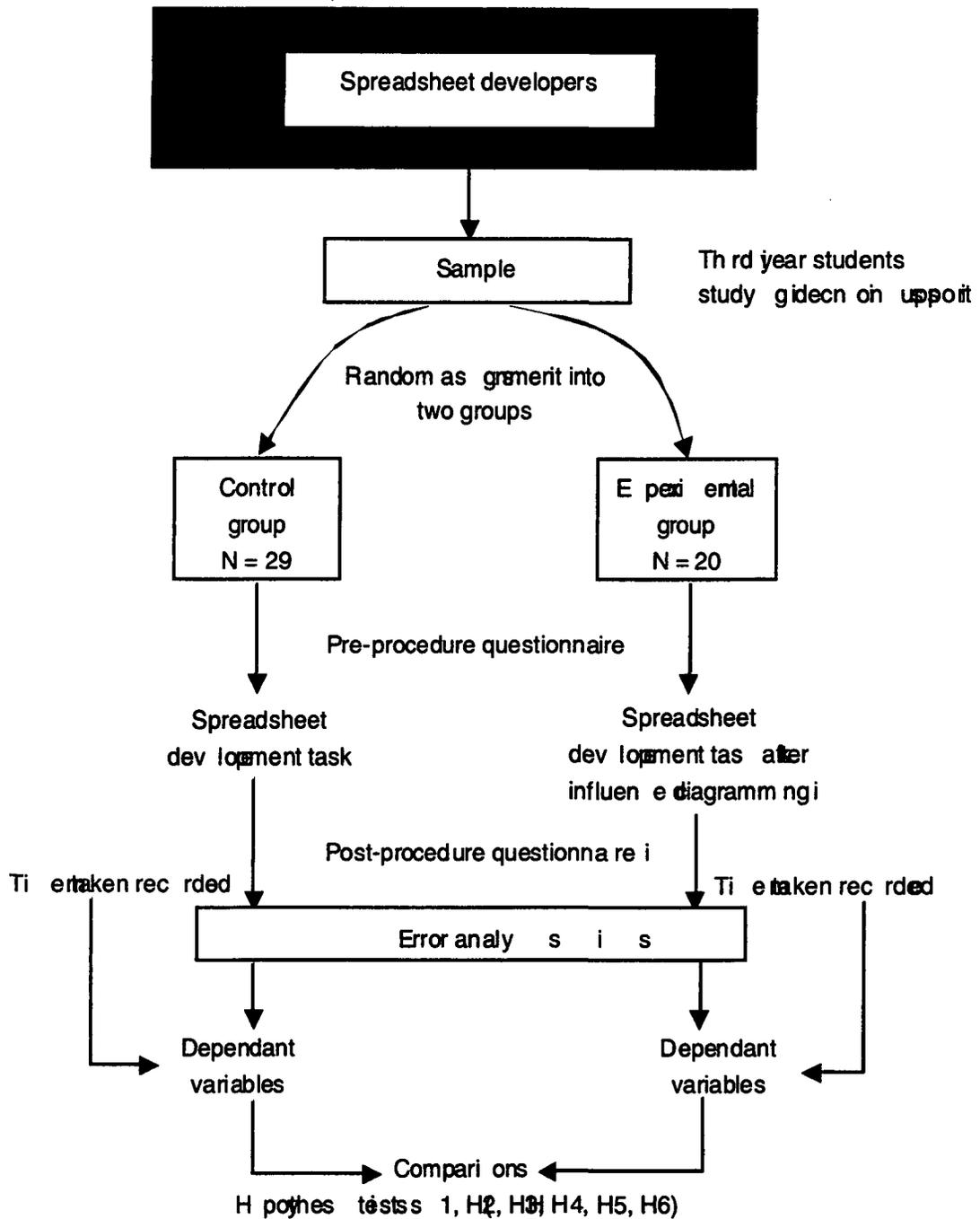
Figure 4. Schematic representation of the experimental method

Research hypotheses

Hypotheses 1 to 5 relate to one of the error classifications discussed earlier. Hypothesis 6 will test whether there is any difference in the time taken to complete the task between the two groups of subjects.

Hypothesis 1: The use of influence diagrams as the basis of spreadsheet design will lead to fewer errors of the type classified as "errors in formulae".

Hypothesis 2: The use of influence diagrams as the basis of spreadsheet design will lead to fewer errors of the type classified as "incorrect ranges".

Hypothesis 3: The use of influence diagrams as the basis of spreadsheet design will lead to fewer errors of the type classified as "omitted factors".

Hypothesis 4: The use of influence diagrams as the basis of spreadsheet design will lead to fewer errors of the

type classified as "data input errors".

Hypothesis 5: The use of influence diagrams as the basis of spreadsheet design will lead to fewer errors of the type classified as "incorrect use of functions".

Hypothesis 6: The use of influence diagrams as the basis of spreadsheet design will not add to the overall time taken to develop a spreadsheet.

Measurement

To measure the number of occurrences of each type of error, each model was compared to a "standard" answer for the problem. No attempt was made to make a judgement about the seriousness of the errors found; the errors were simply counted. Considerable flexibility was used when comparing the models to the standard answer, as the subjects were given no specific instructions about the format or structure of the model that they were to develop.

A research assistant performed the error analysis. This person was provided with the disks that the subjects had used to save their model. These disks were randomly numbered. No indication was given to the research assistant conducting the error analysis that could enable them to identify whether a particular model belonged to the experimental or control group.

The time taken to develop the models was recorded in minutes by the experimenter. When the experimental procedure began the experimenter noted the time. As each subject completed the task and submitted their model the time was noted - enabling the time taken by each subject to complete the task to be easily calculated.

RESULTS

Demographics

There was very little difference between the average ages of the experimental and control groups (23 and 24 years respectively). The oldest subject was 44 years old, the youngest was 19 years old. There were 20 male and 9 female subjects in the control group. The experimental group consisted of 11 male and 9 female subjects.

The subjects were asked if they had experience developing spreadsheet models prior to the decision support systems course. Overall, 59% of the subjects (29 out of 49) claimed prior experience with spreadsheet modelling. The subjects were asked about their major sources of spreadsheet skills. Table 1a shows the sources of spreadsheet training of the subjects. The subjects ranked their spreadsheet development as either 'excellent', 'very good', 'adequate', 'fair' or 'poor'. Table 1b shows how the two groups ranked their spreadsheet skills. This data confirms the prior expectation that the subjects were not spreadsheet experts and that formal training within their degree course was their major source of spreadsheet knowledge.

The experimental subjects were asked to indicate their major sources of influence diagramming skills. Most had only learned of influence diagramming during the study of the decision support systems elective (13 out of 19). The others had been taught influence diagramming in other units of study in their degree program (6 out of 19). This data confirms the expectation that the subjects in the experimental group had no significant exposure to influence diagramming other than that obtained from the decision support systems elective.

Table 1. Sources of spreadsheet skills

	Control Group[3]	Experimental Group[4]
'DSS' unit	17	7
Other subjects	4	4
From a text book or manual	13	9
At work	2	2
On a training course	7	3

Table 2. Subjects rating of their spreadsheet skills

	Control Group	Experimental Group
Excellent	2	2
Very Good	5	1
Adequate	14	12
Fair	6	5
Poor	4	0

[3] Not all subjects in this group answered.

[4] Subjects were allowed to select more than one item.

Upon completion of the experimental procedure, subjects were asked to rate the difficulty of the task they performed on a 5-point Likert scale. A rating of 1 indicated that they found the task very easy, a rating of 5 indicated that they found the task very difficult. Most rated the task as moderately difficult. The mean rating of the subjects in the control group was 3.16 and the mean rating of the subjects in the experimental group was 2.94 (the standard deviations were 1.06 and 0.85 respectively). There was no significant difference between the ratings given by the two groups (t = 0.77).

A surprisingly high number of subjects after completing the exercise, given that they thought that it was only of moderate difficulty, thought that their spreadsheet contained errors. (No prior indication had been given to the subjects that the spreadsheets they developed would be analysed for errors.) In the experimental group 55% (11 of 20) of the subjects believed their spreadsheet contained an error. In the control group 62% (18 of 29) of the subjects believed that their model contained an error.

Summary of the errors found

As anticipated a large number of errors were found in the spreadsheet models developed by the experimental subjects. A summary of the number and types of errors found is shown in Table 3. All of the models developed by subjects in the control group contained at least one error and the majority of the models developed by subjects in the experimental group contained at least one error. Spreadsheets developed by subjects in the control group contained an average of 4.21 errors. Those developed by subjects in the experimental group contained 3.5 errors. The most common type of errors were the errors classified as 'erroneous formula' and 'omitted factors'. Errors of the types 'incorrect range' and 'incorrect use' of functions were rare.

The error rate is high compared to other studies using the same experimental task. For example, individuals developing spreadsheets in Panko & Halverson's (1994) study made 2.36 errors on average (compared to 3.92 for all subjects in this study). In that study 79% of the spreadsheets developed by individuals contained at least one error (compared to 94% for all subjects in this study). There is no clear reason for the higher error rate found in this study. However, in Panko & Halverson's (1994) study a different technique was used to count errors and some errors were not included in the counts.

Table 3. Summary of the number and types of errors found.

	Control group			Experimental group			Total
	Count	Mean	Std Dev.	Count	Mean	Std Dev.	Count
Number of subjects	29			20			49
Erroneous formula	53	1.83	1.71	44	2.20	1.64	97
Incorrect range	4	0.14	0.52	0	0.00	-	4
Omitted factors	48	1.66	2.24	12	0.60	0.88	60
Data input errors	12	0.41	0.78	12	0.60	1.23	24
Incorrect use of functions	5	0.17	0.66	2	0.10	0.45	7
Total number of all errors	122			70			192
Average number of errors per spreadsheet	4.21			3.5			3.92
% of spreadsheets with errors	100%			85%			94%

Hypothesis testing

Upon observation of the data it appeared that the error-count data could not be assumed to be parametric. The Lilliefors test for normality was conducted on each of the data sets. The results of the Liffiefors test confirmed that a non-parametric test would be required to perform the hypothesis testing.

The test used to perform the most of the hypothesis testing was the Wilcoxon Rank Sum test (Wilcoxon, 1945). This test uses a z-score to determine if the difference between two data sets can be attributed to chance. For each hypothesis, the intention was to discover whether the error rate for the experimental group was lower than that of the control group, hence a one-tailed test can be used. The critical value of z for a one-tailed hypothesis test at a 95% level of confidence is 1.65.

The calculations associated with the Wilcoxon Rank Sum test for each hypothesis are given in Tables 4a, 4b, 4c, 4d and 4e. These calculations are summarised in Table 5. These results show that the null hypothesis is accepted for all error types ($H1_0$, $H2_0$, $H4_0$ and $H5_0$) except errors associated with omitted factors. For errors associated with omitted factors the null hypothesis can be rejected and the alternative hypothesis ($H3_A$) can be accepted. That is, there is sufficient evidence to indicate that errors of omitted factors of the experimental group are less in general than that of the control group.

Table 4a. Calculations for Wilcoxon rank sum test - Erroneous formula.

Error count	Frequency (Experimental Group)	Frequency (Control Group)	Ranks	Mean rank	Frequency by average rank (Experimental Group)	Frequency by average rank (Control Group)
0	5	5	40-49	44.5	222.5	222.5
1	1	9	30-39	34.5	34.5	310.5
2	6	10	14-29	21.5	129	215
3	2	1	11-13	12	24	12
4	5	2	4-10	7	35	14
5	1	1	2-3	2.5	2.5	12.5
8	0	1	1-1	1	0	1
	$N_{control} = 20$	$N_{experimental} = 29$			$T_{Experimental} = 447.5$	$T_{Control} = 777.5$

Table 4b. Calculations for Wilcoxon rank sum test - Incorrect formula.

Error count	Frequency (Experimental Group)	Frequency (Control Group)	Ranks	Mean rank	Frequency by average rank (Experimental Group)	Frequency by average rank (Control Group)
0	20	27	3-49	26	520	702
2	0	2	21-24	1.5	0	3
	$N_{control} = 20$	$N_{experimental} = 29$			$T_{Experimental} = 520$	$T_{Control} = 705$

Table 4c. Calculations for Wilcoxon rank sum test - Omitted factors.

Error count	Frequency (Experimental Group)	Frequency (Control Group)	Ranks	Mean rank	Frequency by average rank (Experimental Group)	Frequency by average rank (Control Group)
0	13	12	25-49	37	481	444
1	2	1	21-24	22.5	45	45
2	5	11	5-20	12.5	63	138
3	0	1	4-4	4	0	4
4	0	1	3-3	3	0	3
7	0	1	2-2	2	0	2
10	0	1	1-1	1	0	1
	$N_{control} = 20$	$N_{experimental} = 29$			$T_{Experimental} = 589$	$T_{Control} = 637$

Table 4d. Calculations for Wilcoxon rank sum test - Data input errors.

Error count	Frequency (Experimental Group)	Frequency (Control Group)	Ranks	Mean rank	Frequency by average rank (Experimental Group)	Frequency by average rank (Control Group)
0	15	21	14-49	31.5	473	662
1	2	5	7-13	10	20	50
2	0	2	5-6	5.5	0	11
3	2	1	2-4	3	6	3
4	1	0	1-1	1	1	0
	$N_{control}$ = 20	$N_{experimental}$ = 29			$T_{Experimental}$= 500	$T_{Control}$ = 726

Table 4e. Calculations for Wilcoxon rank sum test - Incorrect use of functions.

Error count	Frequency (Experimental Group)	Frequency (Control Group)	Ranks	Mean rank	Frequency by average rank (Experimental Group)	Frequency by average rank (Control Group)
0	19	27	4-49	26.5	504	716
2	1	1	2-3	2.5	3	3
3	0	1	1-1	1	0	1
	$N_{control}$ = 20	$N_{experimental}$ = 29			$T_{Experimental}$= 506	$T_{Control}$ = 719

Table 5. Summary of the hypothesis tests using Wilcoxon rank sum test.

	Error type	$T_{Experimental}$	$T_{Control}$	$E(T)$	sT	Z	Result
H1	Errors in formula	447.5	777.5	500	49.16	-1.07	Accept H0, Reject HA
H2	Incorrect ranges	520	705	500	49.16	0.41	Accept H0, Reject HA
H3	Omitted factors	589	637	500	49.16	1.81	Reject H0, Accept HA
H4	Data input	500	726	500	49.16	0.00	Accept H0, Reject HA
H5	Incorrect use of functions	506	719	500	49.16	0.12	Accept H0, Reject HA

$$E(T) = nExperimental\ (\ nExperimental + nControl + 1)\ /2$$
$$nExperimental = 20,\ nControl = 29,$$
$$Z = T_{Experimental} - E(T)\ /\ sT,$$
$$\text{Rejection region (95% confidence) is } Z > 1.65$$

Hypothesis 6 was tested using a t-test as the data used (the time taken in minutes) was parametric. The results show that there is no significant difference between the mean time taken by subjects in the control (mean 49.3, std dev 13.0) and experimental (mean 50.6, std dev 16.7) groups to complete the exercise (t = -0.283).

DISCUSSION

Findings

The primary question addressed in this investigation is whether the use of influence diagrams improves the quality of spreadsheet models. The hypotheses testing shows that they do. The subjects in the experiment who drew an influence diagram prior to developing their spreadsheet model did a better job of identifying the

variables relevant to the problem than the other subjects. As a result they made less errors of the type omitted factors. This finding can be explained by the fact that drawing an influence diagram represents an explicit stage in the design of the model where thought is given to the nature and type of the variables involved. The influence diagram becomes part of the developers conceptual model. Without giving explicit thought to that process, the subjects who didn't draw a diagram were likely to leave a variable out of their model. The rows and columns of the spreadsheet that they are manipulating as they think about the problem becomes the major framework they use for thinking about the problem. In most problem situations this conceptualisation is not likely to be effective. One argument against drawing an influence diagram is that it is an extra task that must be performed - that of constructing the diagram - and that this will slow down the development process. In this study the subjects in the experimental group didn't take any extra time overall to develop their model. It is reasonable to conclude that the time spent drawing the diagram enabled them to develop the system in less time - resulting in the same overall time. The argument that drawing an influence diagram prior to building a spreadsheet model will slow down overall development appears to be a weak one.

It must also be noted that the experimental group made just as many of the other types of errors as the control group. It was thought possible that the drawing of an influence diagram would lead to a more considered and careful approach to system construction which would help lessen errors related to typing mistakes and errors in formulas that might have resulted from less careful attention to the task. This seems not have been the case.

Study limitations and further research

It must be noted that the study is subject to several limitations that may limit the generalisability of the findings. First, undergraduate students with little 'real' experience of spreadsheet development were used. However, in a previous study by Panko & Sprague's (1998) there was no difference between the performance of undergraduate students and more experienced MBA students in terms of the number and type of errors made.

Another limitation relates to the used of a standard 'canned' problem. The problem - whilst realistic was not real. The subjects had no personal interest in the problem. They were not rewarded in any way for participation in the experiment nor was their performance to be rewarded or punished in any way. The use of participants working on a model that they have a personal interest in might lead to better performance and fewer errors. (It might also lead to a better influence diagram as well - so the benefits of drawing a diagram might also be magnified).

The laboratory environment in which the experiment was conducted is also a little removed from the real environment in which people build spreadsheets that are actually used for decision analysis. In their own office, a spreadsheet developer has other people to talk to and ask for help, manuals to refer to and perhaps access to old spreadsheets - parts of which they might be able to reuse. This might significantly change the nature of the development task, having an effect of the efficacy of influence diagrams. A field experiment using experienced spreadsheet developers, following a similar experimental process to the one used in this study, working in their normal work environment is currently being conducted to investigate this. However, it should be noted that in previous studies there has been little difference between the error rates found in spreadsheets developed in laboratory experiments and those found in field experiments and field audits (Panko, 1998).

CONCLUDING COMMENTS

Spreadsheets are commonly used as a decision making aid. Many decisions are made using spreadsheets that contain errors. A lot of research has been conducted to investigate the nature and frequency of these errors. Unfortunately little work to date has been performed to develop techniques and methods to help improve spreadsheet modelling practice. The techniques and methods that do exist are rarely applied in practice.

This study has investigated whether drawing an influence diagram before developing a spreadsheet model improves the quality of the resulting model. A laboratory experiment, using senior undergraduate IT students as subjects, was conducted. A high number of errors was found in the spreadsheet models developed. The subjects in this experiment who drew an influence diagram before developing their models made fewer errors of the type 'omitted factors'. That is their models were less likely to have a relevant variable omitted in error. Drawing the influence diagram didn't add to the overall time taken to build their model.

In short, the results of the study show that if people draw an influence diagram before beginning to construct their spreadsheet then they will build a better model – due to the improved conceptualisation of the problem afforded by the construction of the influence diagram. Influence diagrams are simple to draw and understand, and simple to learn to draw. A brief tutorial on influence diagrams could easily and should be introduced to any course or training material that teaches spreadsheet modelling.

ACKNOWLEDGEMENTS

REFERENCES

----. (1999, September 1999). **Impact of Migrants on the Commonwealth Budget.** Australian Immigration Research, Australian Department of Immigration and Multicultural Affairs. Available: http://www.immi.gov.au/research/acil1.htm [2000, 28 April].

Berglas, A., & Hoare, P. (1999). Spreadsheet errors, risks and techniques. **Management Accounting,** *77*(7), 46-47.

Bodily, S. E. (1988). **Modern Decision Making : A Guide to Modelling With Decision Support Systems** (2nd Edition ed.). New York, NY: McGraw-Hill.

Brown, P. S., & Gould, J. D. (1987). An empirical study of people creating spreadsheets. **ACM Transactions on Office Information Systems,** *5*(3), 258-272.

Caine, D. J., & Robson, A. J. (1993a). Models for decision making. **Management Services**(January), 28-30.

Caine, D. J., & Robson, A. J. (1993b). Spreadsheet modelling: Guidelines for model development. **Management Decision,** *31*(1), 38-44.

Chussil, M. J. (1988). Computer-based planning: A PC survival guide. **Journal of Business Strategy,** *9*(1), 38-42.

Cragg, P. B., Grant, B., & King, M. (1991). An exploratory study of spreadsheet modelling practice. Paper presented at the **2nd Australian Conference on Information Systems,** Wollongong.

Cragg, P. B., & King, M. (1993). Spreadsheet modelling abuse: An opportunity for OR? **Journal of the Operational Research Society,** *44*(8), 743-752.

Edge, W. R., & Wilson, E. J. G. (1990). Avoiding the hazards of microcomputer spreadsheets. **Internal Auditor,** *47*(2), 35-39.

Flaherty, J., Morgan, P., Lombardo, R., de Silva, B., & Wilson, D. (1999). **A Spreadsheet Approach to Business Quantitative Methods.** Melbourne, Aust.: Logistics Group, RMIT University.

Galletta, D. F., Hartzel, K. S., Johnson, S. E., Joseph, J. L., & Rustagi, S. (1996). Spreadsheet presentation and error detection: An experimental study. **Journal of Management Information Systems,** *13*(3), 45-63.

Godfrey, K. (1995). **Computing error at Fidelity's Magellan fund,** [Discussion archive]. The Risks Digest. Available: http://catless.ncl.ac.uk/Risks/16.72.html#subj1.1 [2000, 28 April].

Hall, M. J. J. (1996). A risk and control-oriented study of the practices spreadsheets application developers. Paper presented at the **29th Annual Hawaii International Conference System Sciences.**

Hayen, R. L., & Peters, R. M. (1989). How to ensure spreadsheet integrity. **Management Accounting**(April), 30-33.

Holtzman, S. (1989). **Intelligent Decision Systems.** Reading, Mass: Addison Wesley.

Howard, R. A. (1990). From influence to relevance to knowledge. In R. M. Oliver & J. Q. Smith (Eds.), **Influence Diagrams, Belief Nets and Decision Analysis.** Chichester, UK: John Wiley and Sons.

Janvrin, D., & Morrison, J. (2000). Using a structured design approach to reduce risks in end user spreadsheet development. **Information and Management,** *37*(1), 1-12.

Kemp, R. (1996-97). **Question without Notice (Speech): Midyear Review 1996-97** (Senate Hansard). Canberra, Australia: Parliament of Australia.

Landry, M., Malouin, J.-L., & Oral, M. (1983). Model validation in operations research. **European Journal of Operations Research,** *14*, 207-220.

Nardi, B. A., & Miller, J. R. (1991). Twinkling lights and nested loops" distributed problem solving and spreadsheet development. **International Journal of Man-Machine Studies,** *34*, 161-184.

Owen, D. L. (1978). The use of influence diagrams in structuring complex decision problems. Paper presented at the **2nd Lawrence Symposium on Systems and Decision Sciences,** Berkeley, CA.

Panko, R. R. (1988). **End user computing: Management, applications and technology.** New York, NY: Wiley.

Panko, R. R. (1996). Hitting the wall: Errors in developing and debugging a 'simple' spreadsheet model. Paper presented at the **29th Annual Hawaii Conference on System Sciences.**

Panko, R. R. (1998). What we know about spreadsheet errors. **Journal of End User Computing,** *10*(2), 15-21.

Panko, R. R. (2000). **Reports of Spreadsheet Errors in Practice,** [html page]. Available: http://panko.cba.hawaii.edu/ssr/Cases.htm [2000, 28 April].

Panko, R. R., & Halverson, R. P. J. (1994). Individual and group spreadsheet design: Patterns of errors. Paper presented at the **Twenty-Seventh Hawaii International Conference on Systems Sciences,** Kihei, Hawaii.

Panko, R. R., & Sprague, R. H. J. (1998). Hitting the wall: errors in developing and code inspecting a 'simple' spreadsheet model. **Decision Support Systems,** *22*(4), 337-353.

Ragsdale, C. (1998). **Spreadsheet Modelling and Decision Analysis: A Practical Introduction to Management Science**.: South-Western.

Ronen, B., Pally, M. A., & Lucas, H. C. J. (1989). Spreadsheet analysis and design. **Communications of the ACM**, *32*(1), 84-93.

Savitz, E. J. (1994, December 12). Magellan loses its Compass. **Barron's**, *84*.

Simkin, M. G. (1987). How to validate spreadsheets. **Journal of Accountancy**, *November*, 130-138.

Soden, W. (1996). **Transcript of a meeting of the Legal and Constitutional Legislation Committee (Senate)**. Canberra, Australia: Parliament of Australia.

Sperling, E. (1997). Spreadsheets. **Computer Reseller News**, *738*(74), 1.

Teo, T. S. H., & Tan, M. (1999). Spreadsheet development and 'what if' analysis: quantitative versus qualitative errors. **Accounting, Management and Information Technologies**, *9*(Jul-September), 141-160.

Whittaker, D. (1999). Spreadsheet errors-and techniques for finding them. **Management Accounting**, *77*(9), 50-51.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. **Biometrics**, *1*, 80-83.

APPENDIX: THE PROBLEM

The problem the participants were asked to solve follows. It is known as the galumpkes word problem and it was taken from Panko & Halverson (1994) who used it in a similar laboratory-based experiment.

Your task is to produce a two-year pro-forma income statement using a spreadsheet -- a projection of a company's financial performance. The company sells galumpkes, which are small food warmers used in restaurants. The owner will draw a salary of $80,000 per year. There is also a manager of operations, who will draw a salary of $60,000 per year. The tax rate is expected to be 25% in each of the two years. Each galumpke will require $40 in materials cost and $25 in labour cost the first year. These numbers are expected to change to $35 and $29 in the second year. There will be a capital purchase of $500,000 in the first year. For depreciation, assume 10% straight-line depreciation with no scrap value. Unit sales price is expected to be $200 the first year and $180 in the second year. There will be three sales people. Their average salary is expected to be $30,000 in the first year and $31,000 in the second. The rent will be $3000 per month. The company expects to sell 3000 galumpkes in the first year. In the second, it expects to sell 3200.