



Murdoch
UNIVERSITY

MURDOCH RESEARCH REPOSITORY

<http://researchrepository.murdoch.edu.au/>

McGill, T.J. and Dixon, M.W. (2001) *Spreadsheet knowledge: An exploratory study*. In: *Managing Information Technology in a Global Economy: 2001 IRMA International Conference, 20 - 23 May, Toronto, Ontario, Canada.*

<http://researchrepository.murdoch.edu.au/856/>

It is posted here for your personal use. No further distribution is permitted.

Spreadsheet Knowledge: An Exploratory Study

Tanya McGill and Michael Dixon
Division of Business, Information Technology and Law
Murdoch University, WA, Australia

Tel: 61 8 9360-2798

Fax: 61 8 9360-2941

[{mcgill, m_dixon}@murdoch.edu.au](mailto:{mcgill,m_dixon}@murdoch.edu.au)

ABSTRACT

Spreadsheet software is flexible and powerful and is the most commonly used software for creating user developed applications. This study explores the nature of spreadsheet knowledge and its relationship with factors that have been identified as contributing to errors in spreadsheets. The results show a positive relationship between both experience and training and levels of knowledge about spreadsheet features. However, neither experience nor training were found to be associated with spreadsheet development knowledge or quality assurance knowledge. The results of the study do, however, suggest that end users' perceptions of their skill are relatively consistent with their actual spreadsheet knowledge. Some implications of these results are discussed.

INTRODUCTION

User developed applications (UDAs) support decision making and organisational processes in the majority of organisations (McLean, Kappelman, & Thompson, 1993) and end user development of applications provides users with a valuable and popular alternative to the traditional process of systems development. Spreadsheets are the most commonly used software for creating UDAs (Amoroso & Cheney, 1991). Spreadsheet software is relatively easy to learn and use, flexible and powerful.

Unfortunately, many studies have shown that spreadsheets commonly contain significant errors that can compromise their value as aids to decision making (Brown & Gould, 1987; Cragg & King, 1993; Kreie, 1998; Panko, 1996). The literature suggests that factors contributing to spreadsheet errors include developer inexperience, poor design approaches, inappropriate application types, problem complexity, time pressure, and absence of review procedures (Janvrin & Morrison, 2000). Several of these problems may be associated with a lack of spreadsheet knowledge. This study attempts to explore the nature of spreadsheet knowledge and its relationship with factors that have been identified as contributing to errors in spreadsheets.

It might be expected that those end user developers with more experience would create spreadsheets with fewer errors. However there is little evidence to suggest that this is so. Panko and Sprague (1996) found no difference in the number of errors in spreadsheets created by undergraduates with little experience and MBA students with more experience. McGill (2000) also found no difference in the quality of spreadsheets between those developed by high experience end user developers and low experience end user developers, despite the fact that the high experience developers were more realistic in their perceptions of spreadsheet quality.

End users often develop spreadsheet applications in an informal, iterative manner (Cragg & King, 1993; Kreie, 1998) and this is believed to contribute to the high error rate in spreadsheets. Several methodologies for developing spreadsheets have been proposed (e.g. Ronen, Palley, & Lucas, 1989; Salchenberger, 1993) and a number of studies have investigated the role of design techniques in spreadsheet development. Janvrin and Morrison (2000) explored the impact of using a structured design approach and found a reduction in errors. Babbitt, Galletta and Lopes's (1998) study of spreadsheet development by novice users also suggested that users who plan and test their spreadsheets will develop better quality spreadsheets.

The amounts and types of training previously received might also be expected to be related to the quality of spreadsheets. Spreadsheet users generally receive little training (McGill, 2000; Taylor, Moynihan, & Wood-Harper, 1998) and the major means of training is self-study (Benham, Delaney, & Luzi, 1993; Chan & Storey, 1996; McGill, 2000). It has been suggested that when end users are self-taught the emphasis is predominantly on how to use the software rather than broader analysis and design considerations (Benham et al., 1993). There are many books that teach introductory spreadsheet skills typically giving a detailed, step-by-step coverage of examples that illustrate the main product features. However, the very proliferation of these features in recent software versions means that, increasingly, the fundamentals of 'what' the end user is attempting to do are being obscured by the multiplicity of ways 'how' to achieve it. Examples are presented as solutions to problems without the design stages being made explicit. Thus end users may have a narrow knowledge focused on spreadsheet features but lacking in techniques for developing spreadsheets that are user-friendly, reliable, and maintainable. Taylor, Moynihan and Wood-Harper (1998) found that few, if any, quality principles are applied in end user development and McGill (2000) found that spreadsheets developed by end users with high levels of training were of no better quality than those developed by end users with low levels of training.

Very little research has looked explicitly at spreadsheet knowledge. Panko (1996) claimed that various measures of prior knowledge failed to distinguish between those who made errors and those who did not and Kreie (1998) also found no relationship between spreadsheet knowledge and spreadsheet quality. However, she speculated that this was because all of the subjects in her study had high levels of spreadsheet knowledge. Her spreadsheet knowledge instrument also focussed specifically on spreadsheet features rather than spreadsheet design knowledge.

RESEARCH QUESTIONS

This study attempts to explore the nature of spreadsheet knowledge and its relationship with factors that have been identified as contributing to errors in spreadsheets. From the above review of the relevant literature it might be suspected that neither experience nor the types of training that are predominantly received will lead to knowledge and skills necessary for the development of good quality spreadsheets. Hence, the first research question investigated was:

Is there an association between years of spreadsheet experience and level of spreadsheet knowledge ?

Regular use of a spreadsheet will provide repeated opportunities to encounter advanced features of the software. The user will see toolbar icons for, as yet, unused features and when attempting to solve a new problem has recourse to the help menu. However, regular use of a

spreadsheet will not necessarily ensure that the end user learns more about system design processes or about quality assurance. It was therefore hypothesized that:

- H1: Greater experience is associated with greater knowledge of spreadsheet features.
- H2: Greater experience is not associated with greater knowledge about development processes.
- H3: Greater experience is not associated with greater knowledge about quality assurance.

The second research question investigated was:

Is there an association between amount of spreadsheet training and level of spreadsheet knowledge ?

Whilst the purpose of spreadsheet training is to increase spreadsheet knowledge and skill, it appears that much spreadsheet training does not emphasize design of spreadsheets or quality assurance principles (McGill, 2000). It was therefore hypothesized that:

- H4: Higher levels of spreadsheet training are associated with greater knowledge of spreadsheet features
- H5: Higher levels of spreadsheet training are not associated with greater knowledge about development processes.
- H6: Higher levels of training are not associated with greater knowledge about quality assurance.

The third research question considered the relationship between end user developers' perceptions of their spreadsheet skill and their spreadsheet knowledge. Organizations rely heavily on end user developers' perceptions of the quality and usefulness of user developed applications (Panko & Halverson, 1996) so any insight into the realism of their perceptions could be valuable. Therefore the third research questions asks:

Are end user's perceptions of spreadsheet skill consistent with their levels of spreadsheet knowledge?

No specific hypotheses were generated for this question.

METHOD

The participants in the study were 60 predominantly mature aged students enrolled in undergraduate business degrees. All had some previous exposure to spreadsheet use. Students were recruited during class and completed a questionnaire on the spot. It was stressed that completion of the questionnaire was voluntary and that it formed no part of their assessment.

The questionnaire consisted of two sections. The first section asked questions about the participants and their previous training and experience with spreadsheets, and the second section tested their spreadsheet knowledge. Spreadsheet experience was measured in years. Level of previous spreadsheet training and perceived level of spreadsheet skill were measured using a 5 point scale where 1 was labelled 'None' and 5 was labelled 'Extensive' (Al-Shawaf, 1993).

The second section of the questionnaire was a 32 item multiple choice test of spreadsheet knowledge (see Appendix 1 for example items). Each item was presented as a multiple choice question with 5 options. In each case the 5th option was 'I don't know' or 'I am not

familiar with this feature'. Fourteen of the items related to knowledge about the features and functionality of spreadsheet packages. Kreie's (1998) spreadsheet knowledge instrument was used as a starting point for the development of these items. Nine of the items tested knowledge of spreadsheet development. These items were developed specifically for the study and drew upon published methodologies for the development of spreadsheets (Ronen et al., 1989; Salchenberger, 1993), whilst attempting to ensure that subjects were not disadvantaged by a lack of specialised terminology. The items covered areas such as the need for planning and methods of testing. Nine items were also included to test knowledge of quality assurance of spreadsheets. These items were developed specifically for the study using Rivard et al.'s (1997) instrument to measure the quality of end user developed applications as a source of material. The 32 items were examined for content validity by four information technology academics who have been involved in teaching spreadsheet use and design. The instrument was shown to be reliable with a Cronbach's alpha of 0.77 (Nunnally, 1978). Knowledge scores for each category were obtained by summing the number of correct responses.

RESULTS

Table 1 summarizes the previous experience and training of the participants and their perceptions of their skill with spreadsheets. The subjects had an average of 4.35 years experience using spreadsheets with a minimum of just a few weeks and a maximum of 13 years. The average level of previous spreadsheet training was 2.72 (out of 5) and the average perceived level of skill was 2.85 (out of 5).

Table 1 also summarizes the spreadsheet knowledge scores obtained by the participants. They had an average score of 8.17 (58.33%) on the spreadsheet features component, an average score of 6.20 (68.9%) on the knowledge of design component and an average of 5.88 (65.37%) on the knowledge of spreadsheet quality component. The average overall knowledge score was 20.55 (63.28%). Although the average spreadsheet features score was lower than the average scores for design knowledge and quality assurance knowledge this is most likely an artefact of the individual questions asked and not a reflection of absolute levels of knowledge.

	Mean (N= 60)	Minimum	Maximum	Std. Deviation
Years of experience	4.35	0	13	3.30
Training level	2.72	1	4	1.08
Perceived skill	2.85	1	5	1.02
Knowledge of features	8.17 (58.33%)	3	14	2.57
Knowledge of design	6.20 (68.88%)	2	9	1.54
Knowledge of quality assurance	5.88 (65.37%)	1	9	1.92
Total knowledge score	20.25 (63.28%)	9	31	4.98

Table 1: Summary information about the participants and their knowledge of spreadsheets

The first research question considered the relationship between years of experience using spreadsheets and spreadsheet knowledge. To address this question, Pearson correlation

coefficients were calculated between years of spreadsheet experience and the knowledge scores for each participant (see Table 2). There was a significant positive correlation between years of spreadsheet experience and the overall knowledge score ($r=0.369$, $p=0.004$). However this relationship did not hold for all of the component scores. Years of spreadsheet experience was significantly correlated with knowledge of spreadsheet features ($r=0.524$, $p<0.001$), but not with design knowledge ($r=0.047$, $p=0.723$) or with knowledge of spreadsheet quality ($r=0.216$, $p=0.097$). Thus the study provided support for the first three hypotheses.

The role of previous spreadsheet training in acquiring spreadsheet knowledge was addressed by calculating Pearson correlation coefficients between the level of training and the knowledge scores for each participant (see Table 2). There was a significant positive correlation between level of previous training and overall knowledge score ($r=0.317$, $p=0.014$). However this relationship did not hold for all of the component scores. Level of training was significantly correlated with knowledge of spreadsheet features ($r=0.416$, $p=0.001$), but not with design knowledge ($r=0.158$, $p=0.228$) or with knowledge of spreadsheet quality ($r=0.140$, $p=0.228$). Thus the study provided support for hypotheses four to six.

The third research question related to end users' perceptions of their own spreadsheet skill. The relationship between perceived skill and spreadsheet knowledge scores was examined using Pearson correlations (see Table 2). There was a significant positive correlation between perceived skill and overall knowledge score ($r=0.500$, $p<0.001$). There was also a significant positive correlation between perceived skill and knowledge of spreadsheet features ($r=0.616$, $p<0.001$) and between perceived skill and design knowledge ($r=0.289$, $p=0.025$). Perceived skill was not found to be significantly correlated with knowledge of spreadsheet quality ($r=0.241$, $p=0.064$), although with a probability of 0.064 there is some suggestion of a positive relationship here as well. This marginal result should be followed up in future research.

N = 60	Knowledge of features	Knowledge of design	Knowledge of quality assurance	Total knowledge
Years of Experience	0.524***	0.047	0.216	0.369**
Training level	0.416**	0.158	0.140	0.317*
Perceived skill	0.616***	0.289*	0.241	0.500***

Table 2: Correlations between the spreadsheet knowledge scores and experience, training and perceived skill

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

DISCUSSION

This study investigated the relationship between types of spreadsheet knowledge and several factors that have been identified as associated with spreadsheet quality: previous spreadsheet experience, previous spreadsheet training and perceived skill with spreadsheets. The results

support the notion that there are different types of spreadsheet knowledge required by end user developers and show that neither experience nor the sorts of training commonly received lead to increases in spreadsheet development knowledge or quality assurance knowledge. The results of the study do, however, suggest that end users' perceptions of their skill are relatively consistent with their actual spreadsheet knowledge. Whilst the findings should be considered preliminary because of the use of student subjects, the results raise concerns about the types of training that end user developers receive.

The amount of time that the subjects had been using a spreadsheet was found to be significantly correlated with their level of knowledge of spreadsheet features. This finding is as predicted by the first hypothesis, but needs to be reconciled with the results of Panko and Sprague (1996) and McGill (2000). In both these studies experience was not found to be related to quality aspects of user developed spreadsheets. However, the apparent difference may be due to two factors. Firstly, end users may have high levels of spreadsheet knowledge but not necessarily use it when developing applications. In the McGill (2000) study subjects with high levels of experience were more realistic in their perceptions of spreadsheet quality despite not developing spreadsheets of better quality, suggesting access to greater reserves of spreadsheet knowledge. Secondly, knowledge of spreadsheet features may not be sufficient for developing quality applications.

Spreadsheet experience was not found to be related to levels of either design knowledge or quality assurance knowledge. End users with years of experience did not know more about spreadsheet development or quality assurance than relatively novice end users, despite having greater knowledge of spreadsheet features. Given the heavy reliance of organizations on end users' perceptions of quality and usefulness, this finding raises concerns about end users' abilities to develop applications, and highlights the need for more formal means of ensuring that end user developers have the requisite knowledge.

The amount of previous spreadsheet training was found to be significantly related to the subjects' levels of knowledge of spreadsheet features. This finding is reassuring, as training is the major vehicle available for facilitating end user development and user developers must have a broad knowledge of spreadsheet features to be able to develop the necessary functionality in their applications. However, no relationship was found between level of training and either spreadsheet design knowledge or quality assurance knowledge. Whilst this finding is consistent with the findings of Benham, Delaney and Luzzi, (1993) it raises major concerns about the types of training that end user developers receive. Future research should investigate the relationship between training content and the types of spreadsheet knowledge end users have.

End user perceptions of their own spreadsheet skill were significantly correlated with overall spreadsheet knowledge, knowledge of spreadsheet features, knowledge of spreadsheet design and marginally correlated with knowledge of quality assurance. This finding is very positive for organizations as it suggests that end user developers have a clear insight into their own abilities. With the majority of organizations imposing no quality control procedures on user developers (Bergeron & Berube, 1990; Cale, 1994; Panko & Halverson, 1996) this insight is essential. Organizations must be able to rely upon the judgements of their end users as to the fitness for use of applications, and if end users' perceptions of their own skill are not realistic then these judgements would be suspect.

CONCLUSIONS

This study has provided both encouraging and disturbing findings. The positive relationship between end user perceptions of spreadsheet skill and spreadsheet knowledge levels is very positive as it suggests that the confidence organizations place in the judgements of end user developers is not misplaced. However, the lack of relationship between previous training and spreadsheet design knowledge or spreadsheet quality assurance knowledge is of concern. With the lack of external quality control procedures imposed on user developers (Bergeron & Berube, 1990; Cale, 1994; Panko & Halverson, 1996), a number of authors have suggested that training may be the most effective tool for minimizing risks associated with end user computing (Cragg & King, 1993; Edberg & Bowman, 1996; Nelson, 1991). However as the results of this study suggest, increasing levels of training per se is no guarantee of improvements in end user knowledge. Future research should investigate the impact of training that emphasizes application development methods and procedures, especially in the area of quality assurance.

REFERENCES

- Al-Shawaf, A.-R. H. (1993). *An Investigation of the Design Process for End-user Developed Systems: An Exploratory Field Study*. Unpublished Ph.D., Virginia Commonwealth University.
- Amoroso, D. L., & Cheney, P. H. (1991). Testing a causal model of end-user application effectiveness. *Journal of Management Information Systems*, 8(1), 63-89.
- Babbitt, T. G., Galletta, D. F., & Lopes, A. B. (1998). Influencing the success of spreadsheet development by novice users. *Proceedings of the Nineteenth International Conference on Information Systems*, 319-324.
- Benham, H., Delaney, M., & Luzzi, A. (1993). Structured techniques for successful end user spreadsheets. *Journal of End User Computing* (Spring), 18-25.
- Bergeron, F., & Berube, C. (1990). End users talk computer policy. *Journal of Systems Management* (December), 14-16, 32.
- Brown, P. S., & Gould, J. D. (1987). An experimental study of people creating spreadsheets. *Transactions on Office Information Systems*, 5(3), 258-272.
- Cale, E. G. (1994). Quality issues for end-user developed software. *Journal of Systems Management* (January), 36-39.
- Chan, Y. E., & Storey, V. C. (1996). The use of spreadsheets in organizations: Determinants and consequences. *Information & Management*, 31, 119-134.
- Cragg, P. G., & King, M. (1993). Spreadsheet modelling abuse: An opportunity for OR? *Journal of the Operational Research Society*, 44(8), 743-752.
- Edberg, D. T., & Bowman, B. J. (1996). User-developed applications: An empirical study of application quality and developer productivity. *Journal of Management Information Systems*, 13(1), 167-185.
- Janvrin, D., & Morrison, J. (2000). Using a structured design approach to reduce risks in end user spreadsheet development. *Information & Management*, 37(1), 1-12.
- Kreie, J. (1998). *On the Improvement of End-User Developed Systems Using Systems Analysis and Design*. Unpublished PhD, University of Arkansas.

- McGill, T. J. (2000). User Developed Applications: Can End Users Assess Quality *Challenges of Information Technology Management in the 21st Century. 2000 IRMA International Conference.* 106-111.
- McLean, E. R., Kappelman, L. A., & Thompson, J. P. (1993). Converging end-user and corporate computing. *Communications of the ACM*, 36(12), 79-92.
- Nelson, R. R. (1991). Educational needs as perceived by IS and end-user personnel: A survey of knowledge and skill requirements. *MIS Quarterly*, 15(4), 503-525.
- Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill.
- Panko, R. R. (1996). Hitting the wall: Errors in developing and debugging a "simple" spreadsheet model. <http://www.cba/hawaii.edu/panko/research/risks/h6wall.htm>.
- Panko, R. R., & Halverson, R. P. (1996). Spreadsheets on trial: A survey of research on spreadsheet risks. *Proceedings of the Twenty-Ninth Hawaii International Conference on System Sciences*, 2, 326-335.
- Panko, R. R., & Sprague, R. H. (1996). *Hitting the wall: Errors in developing and code inspecting a "simple" spreadsheet model (96-02)*: Working Paper 96-02, Department of Decision Sciences, University of Hawaii.
- Rivard, S., Poirier, G., Raymond, L., & Bergeron, F. (1997). Development of a measure to assess the quality of user-developed applications. *The DATA BASE for Advances in Information Systems*, 28(3), 44-58.
- Ronen, B., Palley, M. A., & Lucas, H. C. (1989). Spreadsheet analysis and design. *Communications of the ACM*, 32(1), 84-93.
- Salchenberger, L. (1993). Structured development techniques for user-developed systems. *Information & Management*, 24, 41-50.
- Taylor, M. J., Moynihan, E. P., & Wood-Harper, A. T. (1998). End-user computing and information systems methodologies. *Information Systems Journal*, 8, 85-96.

APPENDIX 1 - SAMPLE QUESTIONS

Spreadsheet Features Questions:

If you want the numbers in your spreadsheet to appear as currency (that is with \$ signs, etc), you would use the:

- Edit feature.
- Data feature.
- Format feature.
- Label feature.
- I am not familiar with this spreadsheet feature.

If you copied the formula =A\$1*B1 from cell C1 to cell C2, the formula in cell C2 would be:

- =A\$1*B2.
- =A\$2*B2.
- =A\$1*B1.
- =A1*B2.
- I am not familiar with this spreadsheet feature.

What function does an evaluation (e.g. Is C1 = 10?) and executes either a 'true' or a 'false' action based on the outcome of the evaluation? (Assume the function is preceded by the appropriate symbol for Lotus 1-2-3 or for Microsoft Excel).

- a. BRANCH.
- b. SELECT.
- c. COMPARE.
- d. IF.
- e. I am not familiar with this spreadsheet feature.

Spreadsheet Design Questions:

When you need to create a new spreadsheet, the **FIRST** thing you should do is:

- a. Plan the layout of the spreadsheet on paper.
- b. Work out exactly what the spreadsheet has to do.
- c. Start up your spreadsheet program.
- d. See if you have a previous spreadsheet that you could adapt.
- e. I don't know.

Dividing your spreadsheet into sections is important because it:

- a. Makes it look more professional.
- b. Enhances the compatibility.
- c. Makes it easier to use and change.
- d. Increases the data storage capacity.
- e. I don't know.

Which of the following is **NOT** a reason for planning your calculations on paper:

- a. It allows you to make sure you understand the calculation before worrying how to create a formula for it in your spreadsheet package.
- b. It makes it easier to get someone else to check your logic.
- c. It reduces the likelihood of making errors.
- d. It saves computer processing time.
- e. I don't know.

Quality Assurance Questions:

Which of the following is **NOT** a criterion for an effective spreadsheet?

- a. It is small.
- b. It is accurate.
- c. It is easy to change.
- d. It is standardised and consistent.
- e. I don't know.

A spreadsheet is more likely to be useful over a long period of time if:

- a. Errors are easy to identify.
- b. It is easy to understand the calculations it uses.
- c. It has detailed documentation.
- d. All of the above are true.

- e. I don't know.

Which of the following is **NOT** a characteristic of a well-designed spreadsheet?

- a. Each section of the spreadsheet has a unique function.
- b. It can be printed out on one page.
- c. Corrections are easy to make.
- d. All headings and labels provide clear information about the data they relate to.
- e. I don't know.