

A Survey of MBA Spreadsheet Users

by

Kenneth R. Baker, Lynn Foster-Johnson, Barry Lawson, and Stephen G. Powell

*Spreadsheet Engineering Research Project
Tuck School of Business
Dartmouth College
Hanover, NH 03755*

Abstract

We developed an internet-based questionnaire on spreadsheet use, and we administered the questionnaire to a large number of users to document how spreadsheets are currently being developed and used in the business world. In this paper, we discuss the results drawn from a large sample of MBA graduates. These results describe current spreadsheet use in organizations, largely confirming the gap that exists between actual practice and best practice. Moreover, our results indicate that this gap occurs in two areas: it appears not only in the skills of individual users but also in the policies of large organizations.

Keywords: Spreadsheets, survey results, training, standards, spreadsheet risk

A Survey of MBA Spreadsheet Users

1. Introduction

Spreadsheets are common in the world of business. As they have become more widely accepted, spreadsheets have been employed for increasingly critical business applications. They are regularly used for clerical tasks, for modeling and analysis, and for communication. The popularity of spreadsheet use also has a downside. Stories of business failures, lawsuits, and governmental investigations appear in the press or on the Internet, with errors in spreadsheet use and calculations cited as the reason for the mishap.¹ The passage of Section 404 of the U.S. Sarbanes-Oxley Act has forced companies to actively address spreadsheet usage and errors. As a result, many firms find themselves developing spreadsheet policies and documenting spreadsheet practices. Nevertheless, researchers, auditors, and consultants frequently express the concern that spreadsheet use defies the norms of discipline that can be found in other business activities (Cragg and King, 1993, Cale, 1994, Croll, 2005). They point out that spreadsheet use implies certain risks and costs, and in that light, there is a need for companies to pay attention to the way spreadsheets are managed in the organization (Ronen, et al. 1989, Panko and Halverson, 1997, Berglas and Hoare, 1999, Scheubrein, 2003, Croll, 2004).

In the Tuck Spreadsheet Engineering Research Project (SERP), we have been examining current organizational practice as it relates to the use of spreadsheets, with the aim of promoting best practices for creators and users of spreadsheets. An early step in this research has been to document how spreadsheets are currently being used, and for this purpose, we created a detailed questionnaire that we administered on the Internet. In this paper, we discuss results drawn from our survey. These results help sketch a picture of current spreadsheet use, largely confirming the gap that exists between actual practice and best practice. Moreover, our findings show that this gap is not limited to the skills of individual users; it also extends to the policies of large organizations. Thus, our results begin to identify the major needs facing the community of spreadsheet users.

In the next section, we review survey results that have appeared in the research literature. In Section 3, we describe the SERP questionnaire and the population to whom the survey was administered. In Sections 4 and 5, we give an overview of the survey results, highlighting the most important findings. Section 4 is organized around spreadsheet use, and Section 5 is oriented to management policies. Finally, in Section 6, we discuss the implications of the results and describe other uses of the questionnaire that might shed additional light on spreadsheet use.

¹ A collection of such stories has been compiled by the European Spreadsheet Risk Interest Group (EUSPRIG) at www.eusprig.org/stories.htm.

2. Framework and Literature Review

Our project work began with a series of field visits and open-ended interviews with spreadsheet developers and users at a number of companies. Our observations led us to a seven-stage model that describes the life cycle of a typical spreadsheet. The stages in this life cycle are: designing, testing, documenting, using, modifying, sharing, and archiving. Although we can list the stages, the typical path is not serial. Instead, a spreadsheet may revisit a previous stage—perhaps several times—during its useful life. Figure 1 shows the seven stages and the main evolutionary paths.

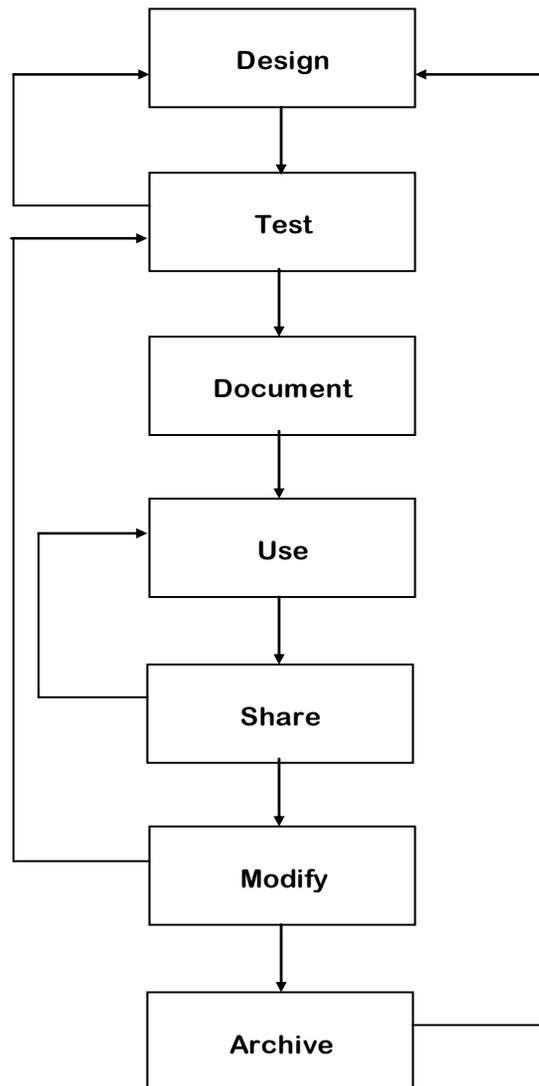


Figure 1. The seven major stages of a spreadsheet life cycle.

Our interviews also explored the role of company policies and their influence on spreadsheet development and use. On this topic, we drew from the research literature (e.g., Floyd et al., 1995) the insight that management policies should cover standards, training, and controls. Our exploratory questions were also organized around these themes.

Spreadsheets have been around for over 25 years, but there have been relatively few published surveys that provide a broad-based look at spreadsheet practices. The most important surveys we were able to find were due to Sajaniemi and Pekkanen (1988), Gable, et al. (1991), Schulteis and Sumner (1994), Floyd, et al. (1995), Hall (1996), Chan and Storey (1996), Speier and Brown (1996), and Pemberton and Robson (2000). We elaborate below on the findings in these surveys, using our life-cycle framework. Some of these surveys looked beyond spreadsheets to other kinds of end-user computing; others looked at spreadsheets but not spreadsheet users. Most of the surveys involved fewer than 100 respondents, and nearly all of the surveys are old enough that the primary spreadsheet software under consideration was Lotus 1-2-3. Our goals in the SERP survey were to reach a much larger sample than previous surveys and to portray an era in which Excel is the dominant software for spreadsheets.

Design

The Design phase of the spreadsheet development is the most critical for incorporating best practices because design decisions influence the entire life of the spreadsheet. Good design practices can accelerate development and avoid lengthy cycles of rework. Researchers have demonstrated, for example, that spreadsheet errors are difficult to detect when spreadsheets are poorly designed (Teo and Tan, 1997, Teo and Lee-Partridge, 2001).

Good practice in spreadsheet design suggests that, prior to working on the computer, the developer should sketch the final spreadsheet or at least create a plan for it. This step describes the scope of the project and provides an opportunity to organize the model's logic. We found little mention in the literature that this practice is critical to spreadsheet design, much less evidence of frequent practice. Hall (1996) reported that only 52% of her sample put plans on paper prior to implementation (although the figure was higher for expert developers), but 77% of the respondents indicated that they "should have done" this practice. Cragg and King (1993) noted that 60% of their spreadsheets had been built without prior design or planning. They surmised that the frequent modifications to the models were due to the lack of forethought and design.

As in other forms of computer programming, best practice calls for modularization in design. For example, in the structured design approach of Janvrin and Morrison (2000), each module is built on a separate worksheet. Nevertheless, Hall (1996) found that most of the respondents in her sample designed their spreadsheets without separate areas for inputs and outputs. Among the respondents, only 45% provided a separate module for calculations, and 51% provided a module for parameters and constants. Cragg and King

(1993) found that only 55% of the spreadsheets they studied were clearly separated into sections, and none of the spreadsheets completely separated data from formulas.

Best practice also requires the ability to tailor the design of a spreadsheet to the way it will ultimately be used. We saw evidence of this practice in the literature, although only certain aspects of this practice appeared to be implemented consistently. For example, it is important to incorporate input and output controls within the spreadsheet design (Kruck and Sheetz, 2001; Yoon, 1995). Good design requires the ability to determine the necessary specifications and features, although Torkzadeh and Lee (2003) found that their survey participants had relatively low perceptions of their abilities to accomplish these tasks.

Against this background, we are interested in what our survey can tell us about the role of planning in design and the prevalence of basic structural initiatives (e.g., modularization and separation) that might be taken during the design stage.

Test

The testing phase of spreadsheet design is commonly recognized as necessary, although most companies don't have a formal policy on this practice (Cale, 1994). Kruck and Sheetz (2001) reported that 79% of the studies in their review recommended testing and debugging.

A number of studies also find that different types of testing are important to adequately assess spreadsheet accuracy (Janvrin & Morrison, 2000; Panko & Sprague, 1998). Bradley (2003) provides a concise overview of various testing techniques.

Most spreadsheets are tested to some degree prior to use. Hall (1996) found that the most frequent testing method was checking formulas with test data, with 71% of her sample indicating that they did so. As evidence of the general acceptance of this practice, she found little difference between expert developers and experienced users. However, using test data at the limits of the normal range was employed by only 33% of the sample, and a minority of the sample (42%) indicated that they used test data containing errors.

We might expect that the degree of required testing increases as the span of users increases. Whereas most assessments explicitly recommend testing for spreadsheets when they are to be used by the developer, the strong sense is that testing is critical when spreadsheets are to be used by multiple departments (Cale, 1994).

With respect to testing, our survey sought to learn whether testing is commonplace and what formal methods are in use.

Document

Documentation is a critical aspect of spreadsheet development which also plays an important role in increasing user understanding and sustaining user satisfaction (Doll and

Ahmed, 1985; Doll and Torkzadeh, 1987). Documentation also plays a role in other aspects of spreadsheet use. For example, Torkzadeh and Doll (1993) found that the quality of spreadsheet documentation was positively associated with the user's understanding of the content, perceived accuracy, ease of use, and timeliness of information access. Pryor (2006) gave a comprehensive overview of documentation forms and uses.

Good design practice includes documentation to anticipate the need for maintenance.. On this topic, Hall (1996) found that only 32% of her sample provided online instructions for use, although 63% indicated that they should have implemented this practice. One of the simplest practices, listing the file names of previous versions, was done by only 13% of the respondents.

Documentation of spreadsheets is seldom addressed formally in organizations or applied systematically. Cale (1994) reported that almost 90% of the 52 firms he surveyed either agreed or strongly agreed that lack of documentation is a potentially serious problem. The majority of his respondents' employers had no company policy or had an unwritten policy on documenting spreadsheets. Schultheis and Sumner (1994) found that documentation ranked third in their list of controls (after verification and training). They report that there was "very little documentation" in the spreadsheets, and a quarter of the spreadsheets had no documentation at all.

Good design practice also suggests that some documentation should exist in paper copy, although electronic documentation is increasing. Hall (1996) found that 75% of her sample had no hardcopy documentation of their spreadsheets. Only 23% of her sample documented design details, formulas, and assumptions and known limits. Fewer (21%) provided instructions for use. With regard to documentation in testing, only 13% kept information on expected and actual test results. Cragg and King (1993) reported similar findings. In their sample, only 10% of the spreadsheets had assumptions documented. In fact, 50% of the spreadsheets in the Cragg and King study had no documentation of any kind. Floyd, et al. (1995) reported that only a few of their end users indicated that all spreadsheets had to be documented. Slightly more than a quarter thought that spreadsheets needed documentation only when there are multiple users.

With respect to documenting, our survey sought to learn how often designs are documented and what formal methods are used.

Use

It is generally acknowledged that spreadsheets are used in just about every area of business (Croll, 2005). Hall (1996) found that most of the spreadsheets were run on a regular basis (67%) and a smaller proportion was run one or two times (17%). The remaining 16% were run occasionally after very long gaps in time. Pemberton and Robson (2000) reported that 48% of their sample used spreadsheets at least 3 times a week, 17% used them one or twice a week, and 12% used them once or twice a month.

Various spreadsheet tools are often suited to the different functional requirements in business. Hall (1996) reported that most of the spreadsheets in her survey used numerous tools in their design. Macros were used in about 45% of the spreadsheets, and graphics were used in 38%. Almost half (47%) of the spreadsheets used the IF function. Lookup functions and tables were used in 27%, and 66% of the spreadsheets used both absolute and relative referencing. Cragg and King (1993) found that half of their spreadsheets used macros. Pemberton and Robson (2000) report that 35% of their sample used database functions frequently, while another 35% used them occasionally. Almost 40% of their respondents used the sorting tool frequently. Interestingly, fewer respondents used summary and statistical facilities. Almost 60% of the sample never used summary measures and slightly over 70% never used regression or correlation analyses. A full 90% of their sample never used advanced statistical analyses.

Chan and Storey (1996) found that the respondents to their survey used most of the available functions, but only moderately. The higher scoring spreadsheet functions were “what-if”, mathematical/statistical, and macro languages. Database, financial, and goal-seeking features were used somewhat less. Chan and Storey also found that the use of spreadsheet features was positively correlated with the number of hours spent carrying out business analysis. They also found that the more proficient the user, the more likely the use of spreadsheet features. The features most strongly associated with proficiency were macros, “what-if” analysis, financial functions, and graphics.

Knowing that there is some self-selection in the sample of responses to our survey, we were interested in verifying patterns of use and learning about the variety of tools that are being employed.

Modify

Modifying spreadsheets is a frequent occurrence. The average lifetime is often quite short before business demands require changes in the spreadsheet. Cragg and King (1993) found that while one-third of the spreadsheets in their study were in their first version, the majority had been updated continually. Over 85% of the models had been modified after the initial implementation. On average, the models had been updated seven times.

The question of who should make modifications has been met with various responses. Floyd et al. (1995) found that about 25% of their users indicated that modifications to a spreadsheet should be made only by the authors of the spreadsheet. About 20% indicated that only the users, and not necessarily the authors, could make changes to the spreadsheet. Hall (1996) found that only 27% of the existing spreadsheets were modified.

Our survey probed the ways in which modification takes place and explored who is responsible for changes.

Share

Sharing of spreadsheets is becoming much more widespread with the use of networks and cross-functional teams. However, controls around spreadsheet sharing have lagged. Schultheis and Sumner (1994) identified such frequent risks as use beyond the developer within the firm, use beyond the firm, and the number of users. Hall (1996) found that relatively few of the spreadsheets were developed by the ultimate user (17%) and that only 15% of end users kept a distribution record. Most of the spreadsheets were used by someone other than the developer. Almost one-third of the spreadsheets (31%) were used by one department, while 23% were used by many departments, and 29% were shared outside the organization. Most of the spreadsheet output stayed in circulation for some time, with over half (55%) being used for longer than one month. Indeed, Schultheis and Sumner had noted that, with the relatively short tenure of many spreadsheet developers, there is a substantial risk that spreadsheet life extends past the tenure of the developer.

Protections can be implemented to reduce the risks associated with spreadsheet sharing. Cragg and King (1993) found that cell protection was used in 30% of the systems they investigated. Clear identification of the version was also infrequent. McGill (2002) found that end-user developers and experts differed in their assessments of applications related to security and integrity of organizational data. Specifically, experts noted significantly fewer applications where unauthorized users could not easily access all of the data and users were required to have a unique password.

With respect to sharing, our survey explored the frequency and extent of the practice, along with the protections that were put in place.

Archive

Archiving is important to maintain institutional knowledge and provide a database of existing and available spreadsheets ready for use. Although Figure 1 shows archiving as the last stage of the life cycle, the diagram also shows that archiving potentially affects the very first steps in design. However, the archiving of spreadsheets is seldom addressed in the research literature. One of the most common practices in archiving is the simple task of making electronic backup files of spreadsheets. Hall (1996) reported that this was a common practice by people in her sample. Although some firms recognize that spreadsheet archives can be useful, our survey sought to find out how common archiving really is.

Finally, a related literature describes studies (not surveys) of spreadsheet use, relying on experiments and interviews. Recent examples include Kreie, et al. (2000), McGill and Klobas (2005), Grossman, Mehrotra, and Özlük (2005), and Croll (2005). Although this literature is more recent than the surveys listed above, the articles deal with rather small samples and address narrowly-focused research questions. They contribute to our understanding of spreadsheet practice but in very specialized ways. They complement the information drawn from our survey.

3. The SERP Survey

We used the seven life-cycle stages and the three policy dimensions to organize the questions in our survey. Thus, the structure and coverage of the questions were based on research findings as well as our own field work. In addition, there were demographic questions that described the respondents themselves. In all, the questionnaire contained 67 items, some of which were open-ended, and took about 15-20 minutes to complete.

We first tested the questions on a small sample of graduate students to improve the clarity of the questions. We also solicited input from several researchers active in the European Spreadsheet Risk Interest Group. They provided suggestions for additional coverage and changes in wording, yielding the final form of the questionnaire.

Several populations have been invited to fill out the questionnaire. In this paper, we focus on the MBA population. We were able to contact alumni of two prominent business schools, one American program and one European program, and invite them to fill out the survey. Participation was voluntary, and we anticipated that spreadsheets would be relatively important to most of those who responded. In fact, we asked about the “level of importance for spreadsheets in your job,” and only about 1% replied that spreadsheets were unimportant. At the other end of the spectrum, 79% replied that spreadsheets were either “very important” or “critical” in their work.

We ran several statistical tests to justify combining the two alumni groups into one sample, based mainly on demographic information.² For example, a breakdown of the samples by age showed that the largest proportion of the first school’s sample (54%) represented ages of 30 and younger; this figure was 47% for the second school’s sample. At the other end of the spectrum, 8% of the first school’s sample and 3% of the second represented ages over 60. In the first sample, 22% were female, whereas the figure was 26% in the second. The majority of both samples held either managerial or executive positions. In the first school’s sample, 33% characterized themselves as being in managerial positions, with 44% in executive positions. In the other school’s sample, the figures were 36% and 50%. Non-managerial roles accounted for 14% in the first sample and 8% in the other. The two samples were also similar in organizational industry profile. Among the respondents, 53% of the first sample indicated that they worked in the service sector (e.g., banking, retail, consulting), compared to 48% of the second. The second most frequent industry represented by both samples was manufacturing, with 18% of the first sample and 19% of the second. Another 16% of the first and 13% of the second sample were employed in organizations best characterized by “other.” There were slight differences between the samples in the primary functional job areas. For the first sample, the most frequent affiliations were finance, other, and marketing, in that order. For the second, the ranking was other, finance, and marketing. Respondents of both samples worked in similarly-sized organizations. The first sample contained 49% who worked in large organizations (over 1000 employees), compared to 43% in the second sample. At

² For questions with multiple answers, we used a chi-square test with a significance level of $p = 0.10$. For questions with two possible answers, we used a test of differences between proportions, with the same p -value.

the other end of the scale, 17% of both samples worked in the smallest organization size (10 or fewer employees).

We cannot claim that this is a representative sample of people who use spreadsheets, nor is it representative of any particular function or industry. The sample is homogeneous to the extent that it contains alumni of MBA programs: they were trained to work on challenging business problems and to perform useful analysis in the business world. Nevertheless, the sample is quite diverse, as measured by age, functional responsibility, and industry; and the sample of 846 respondents is quite a bit larger than any sample previously reported in the literature on this subject. We would expect our respondents to be relatively skilled compared to the average of all spreadsheet users, with many having advanced skills. It is important to remember that these are likely to be people for whom spreadsheets play a significant role. Over 95% of the sample characterized themselves as having either “some expertise” or “extensive expertise” in using spreadsheets. In short, we believe that our sample represents an important segment of the spreadsheet community.

4. Survey Results: The Life Cycle

As mentioned earlier, our survey contained 67 questions, some of which were open ended. Rather than give a complete accounting of the results, we limit our discussion to those questions that seemed to shed the most light on patterns of spreadsheet use and on the gap between current and best practice. For more details, interested readers can visit our project’s website (<http://mba.tuck.dartmouth.edu/spreadsheet/index.html>).

4.1 Designing Spreadsheets

Slightly over 90% of the sample indicated that at least some of their time at work was devoted to *creating* spreadsheets. About 81% of the sample worked alone when creating a spreadsheet; the rest worked with a small team or a project group, or else they actively sought the advice of peers.

About 70% of spreadsheets created were built from scratch, and very few of the respondents replied that they “never” create spreadsheets from scratch. The other 30% were adapted or borrowed from existing spreadsheets in the company, although the actual use of a spreadsheet archive appeared to occur no more than about 10% of the time.

As mentioned earlier, spreadsheets are used for a variety of purposes, and this fact was borne out by the purposes identified in our sample, as shown in Table 1.

Analyzing data (e.g. financial, operational)	90%
Evaluating alternatives	55%
Determining trends and making projections	54%
Tracking data (e.g. budgets, sales, inventories)	54%
Maintaining lists (e.g. names and addresses)	29%

Other	8%
-------	----

Spreadsheets come in many sizes, as depicted in Table 2. A very rough estimate of the average size would be about 2500 cells, but it is the variety, more than the average, that seems significant here.

Under 100 cells	10%
101 to 1000 cells	51%
1001 to 10,000 cells	28%
10,001 to 100,000 cells	9%
Over 100,000 cells	2%

Many spreadsheet developers can recount a story in which a spreadsheet they originally built for their own use became part of an established business process and was used by many other users. Spreadsheets that evolve in this way are often ill-suited for their current use because they were not initially designed with that use in mind. Nevertheless, it was common, among our respondents, to find that their spreadsheets could be widely used. Table 3, where only one answer was permitted, provides the details.

For my personal use	12%
Shared with one or two others	48%
Used by a number of people	30%
Often become permanent assets	10%

What's the first step in creating a spreadsheet? In our labs and classrooms, we usually observe students begin by entering numbers and then formulas directly into the spreadsheet. Best practice, on the other hand, calls for planning at the outset. Among our respondents, regrettably, the most frequent first step was to enter something directly into the computer, as shown in Table 4. Many started by borrowing an existing design, but less than 20% sketched the spreadsheet as a first step, and only a small percentage built an algebraic model at the outset.

Borrow a design from another spreadsheet	26%
Sketch the spreadsheet on paper	18%
Write the fundamental relationships using algebra	3%
Enter the data and formulas directly into a computer	49%
Other	4%

Sketching the spreadsheet up front is one way to avoid the need for extensive rework and modification at a later stage of development. Because cycles of rework are often the source of errors in spreadsheets, not to mention a delay in development time, methods that reduce rework can be valuable. Sketching the spreadsheet is therefore a recommended best practice, but, as Table 4 reveals, it is not necessarily a common practice.

The survey asked two other explicit questions related to best practices in design, drawing on two well-established design principles. First, it is commonly accepted that spreadsheets, especially complicated ones, should be designed by breaking them into modules, building and debugging the modules somewhat separately, and then integrating the modules. Table 5 indicates how often the sample followed this precept.

Never	5%
Sometimes	37%
Usually	40%
Always	18%

Another common recommendation is to separate data (inputs) from formulas (calculations). This means that formula cells should not contain input data. Instead, input data should appear separately (in complex models, that may even mean the use of a separate worksheet), and formulas should reference those cells. Table 6 indicates how often this principle was followed.

Never	5%
Sometimes	34%
Usually	42%
Always	19%

In both of these tables, we see that only about 60% of our respondents “usually” or “always” followed these best-practice steps. This result suggests that there is considerable room for improvement in design practice, even at the level of the simplest and most generic methods.

4.2 Testing and Documenting Spreadsheets

Most computer programmers find that developing the software is more engaging than testing it or documenting it. Except for the most disciplined professional programmers, the testing and documenting steps always seem to get less attention than they deserve. With respect to testing spreadsheets, one possibility is that developers are not familiar with the various techniques that could be used. Our survey asked what methods they employed, as shown in Table 7.

Use common sense	67%
Test extreme cases	42%
Test performance for plausibility	41%
Use a calculator to check selected cells	41%
Examine formulas individually	41%
Use Formula Auditing Toolbar	24%
Display all formulas	14%
Use Error Checking option	6%
Use Go To-Special	3%

Other tools	5%
-------------	----

From the perspective of effective testing, the results again show a gap between current practice and best practice. “Common sense” seems to be the only testing mechanism used by the majority of respondents. Research indicates that developers tend to be overconfident about the quality of their spreadsheets (Brown and Gould, 1987, Davies and Ikin, 1987, Floyd, Walls, and Marr, 1995, and Panko and Halverson, 1997), so a developer’s common sense is likely to be fallible. Relatively common tools for testing, such as the Formula Auditing Toolbar (the “blue arrow” tool), the Formula Audit Mode (a display of all formulas), or the enhanced error-checking tools available in recent versions of Excel, all seem underutilized. About 42% of the sample indicated that they “usually” or “always” tested a spreadsheet, and our coarse estimate is that testing consumes only about 6% of the time devoted to spreadsheet work, whereas some authors claim (see, for example, Olphert and Findlay, 2004) that testing should ideally consume 25-40% of the work.

When it comes to documentation, the picture may even be more bleak. Only 26% of the sample indicated that they “usually” or “always” documented a spreadsheet, and our estimate is that documentation consumes less than 5% of the time devoted to spreadsheet work. The main techniques used by those who do some documenting are displayed in Table 8.

Text in spreadsheet	63%
Cell comments	56%
Documentation sheet in workbook	23%
Separate document	10%
None of the above	2%

4.3 Using Spreadsheets

As we mentioned earlier, most development activity involves working alone, but when it comes to spreadsheet use, very few people work in isolation. Although we might be inclined to think in terms of a “developer” and a “user” playing roles analogous to consultant and client, the picture is far more complex. As Table 9 demonstrates, it is common to have more than one user.

None	13%
1 other person	20%
2-5 other people	53%
6-10 other people	8%
More than 10 other people	6%

The survey also asked about frequency of use. As shown in Table 10, there is quite a variety, but weekly use and monthly use are the most prevalent.

Daily	18%
Once or twice a per week	38%
Monthly	29%
Quarterly	10%
Annually	2%
Less than once a year	4%

We made an effort to learn about the specific Excel features in use. The questionnaire offered a list of 14 Excel functions and tools and asked respondents to indicate, for each one, the extent of use. The scale was qualitative (never, sometimes, often, frequently, daily). Following common practice, we converted that data to a 1-5 scale for the purposes of numerical comparisons. On that basis, we were able to rank the tools in order of use, as shown in Table 11.

Data Sort tool	3.42
IF function	3.36
Finance functions	3.20
Chart Wizard	3.13
Find/replace	3.10
Function Wizard	2.99
Lookup functions	2.75
Conditional Formatting	2.61
Formula Audit tools	2.57
Pivot Tables	2.50
Data Table tool	2.36
Macros	2.32
Goal Seek	2.17
Solver	1.91

Obviously, these tools are not altogether comparable. Using an IF statement in a spreadsheet requires different knowledge than using a macro, and the opportunities to do so differ as well. Nevertheless, the ranked list gives us some insight into the spreadsheet capabilities that are used routinely.

4.4 Sharing and Modifying Spreadsheets

As shown earlier in Table 9, sharing spreadsheet information is a common phenomenon. But this does not always mean handing the spreadsheet over to another user. Table 12 shows the different forms that sharing may take. (Respondents were allowed multiple answers.)

I rarely share any part of a spreadsheet	7%
I provide a summary of results	42%

I provide parts of the spreadsheet	27%
I share the entire model	69%

Most of the time, Table 12 indicates, sharing amounts to providing someone else with the full spreadsheet. In that light, it is somewhat surprising that there is relatively little inclination to use protection methods. Table 13 elaborates on the most common alternatives. (Respondents were allowed multiple answers here.)

Table 13. Methods of protecting a spreadsheet	
None	73%
Password protection	16%
Cell protection	16%
Data validation	4%
Other	4%

Table 13 portrays another gap between current practice and best practice. Given that unprotected spreadsheets are vulnerable, and that sharing is commonplace, we would expect to see more reliance on simple protections.

Whether or not sharing leads to errors, we do know that spreadsheets are often modified. The process of modification requires that the development tasks (design, test, document) be revisited. Interestingly, this does not necessarily mean that the work is done by the original developer, as Table 14 attests.

Table 14. Person who makes the modifications	
The original developer	69%
A new developer	19%
Users	37%

More than a third of the time, modifications are made by users, which raises the question of whether users' skills in designing, testing, and documenting are likely to be comparable to those of developers. The risk of errors creeping in with modification, and the suitability of the original design for future sharing, might both be affected by the skill level of a new developer.

4.5 Archiving Spreadsheets

Our questions about the use of a spreadsheet archive were not motivated by concerns about legal and financial record keeping, although there are certain types of firms for which such considerations are mandatory. Instead, we were curious about the use of an archive as a kind of library, where developers could examine the types of spreadsheets that previous developers had built and used. As mentioned earlier, the first step in designing a spreadsheet, for some developers, is to look at an existing spreadsheet and possibly adapt it. Table 15 shows that this kind of use occurs infrequently, probably less than half the time.

Table 15. Frequency of using archived spreadsheets

Seldom, if ever	51%
Occasionally	41%
Frequently	7%

Although our respondents may sometimes rely on archived spreadsheets, they didn't seem as inclined to contribute to an archiving procedure, as indicated by the responses in Table 16.

I do not catalog	63%
Creator	11%
Version	17%
Title	26%
Date	26%
Department	3%

In brief, the creation and use of spreadsheet archives may be underutilized. This is one of the areas where an organizational initiative might pay off well in terms of faster development times, less rework, and fewer errors. Two questions need to be addressed: (1) what is the best way to structure an archive so that relevant materials can be located quickly, and (2) does the use of archived spreadsheets improve the quality of designs?

5. Survey Results: Organizational Policies

When people speak of current practice involving spreadsheets, they usually refer to the practice of individuals. Nevertheless, organizations also establish practices explicitly or implicitly, and these practices, in turn, influence spreadsheet use. We focused on three areas: training, standards, and protection against risk.

5.1 Training

First, we asked about the types of training that organizations provided. Table 17 shows the different responses.

None	47%
In-house training	34%
Training by external party	19%
One basic session is available	4%
Several sessions, including advanced topics	13%
Spreadsheet specialist who assists designers/users	3%
Other	5%

Two observations seem noteworthy. First, about half the organizations provide no training, which seems surprising given how widely used spreadsheets have become. Second, the use of a “spreadsheet specialist” is rather unusual. We find it unusual in light of the fact—often encountered in our interviews—that many users learn and enhance their skills with the help of an informal “spreadsheet guru” in their department.

Even when training is available, it appears to be rather basic. Respondents who reported that their training included data analysis, macros, or specialized add-ins were in the minority. Tables 18 and 19 summarize our results on the amount of training offered and used.

None	58%
1 or 2 days	28%
3 to 5 days	10%
More than 5 days	8%

None	79%
1 or 2 days	16%
3 to 5 days	3%
More than 5 days	2%

Thus, the training that is offered remains underutilized, although part of the reason may be that the coverage is too elementary for the needs of the target population. When asked about the biggest impediment to participating in company-sponsored training programs, the most frequent answer was “not enough time.” However, we can imagine this answer might be given as well to questions about other kinds of self-improvement activities. More specifically, we asked about incentives provided by employers for participating. Table 20 shows the replies.

None	42%
Organization pays cost of training	23%
Organization provides paid time off	7%
Training is a prerequisite for promotion	2%
Not applicable	23%

Here the categories for “none” and “not applicable” may overlap for companies that offered no training, but the obvious result is that incentives are limited.

5.2 Standards and Policies

In some of our field interviews, we have encountered organizations that have promulgated standards for spreadsheets. These may be largely cosmetic, such as the use of a corporate logo and a common font size, or they may be much more detailed, such as the use of a specific template for analyzing client needs. However, in our experience, such initiatives are rare. This observation was supported by the responses in our survey, as shown in Table 21.

No standards	72%
No written standards, only informal guidelines	21%

Basic written standards	4%
Detailed written guidelines and protocols	3%

We recognize that the scope of “informal” standards can include a wide spectrum of possibilities, but if we focus on written material on this subject, it would seem to exist in fewer than 10% of the organizations represented in our survey. Otherwise, people may suspect that standards exist, but there is little hard evidence. This type of attitude is consistent with the results shown in Table 22.

Table 22. Standards and policies are followed	
Seldom	15.0%
Usually	28.6%
Always	6.0%
Don't know	50.4%

Clearly, if half the population doesn't know whether standards and policies are being followed, there is ample room for improvement. Note here that standards can address features of spreadsheets, but they can also describe the procedures used in managing spreadsheets, such as the stages of the spreadsheet life cycle.

5.3 Managing Risk

The basis for considering spreadsheet risk is straightforward: spreadsheets have come to play an increasingly important role in business software, but there is also growing evidence that errors in spreadsheets are common (Brown and Gould, 1987, Cragg and King, 1993, Panko and Halverson, 1997). As noted earlier, it is not difficult to find reports describing situations where considerable economic losses stemmed from spreadsheet errors.

Basically, risk exists if mistakes are likely to occur or if large amounts of money are at stake. Companies are used to dealing with technology risk, supplier risk, inventory risk, political risk, and the like, but spreadsheet risk is an area that has only recently begun to receive attention. Asked about the importance of spreadsheets to the organization as a whole, 97% of the sample classified it as “moderately important,” “very important,” or “critical.” And, as shown in Table 23, a majority felt that the risks to their organization posed by spreadsheets are either a medium level or a high level.

Table 23. Risk posed by spreadsheets in the organization	
High risk	13%
Medium risk	39%
Low risk	40%
No risk	8%

A question aimed at identifying the person in the organization who would be responsible for addressing issues of risk elicited a variety of answers, but “don't know” was by far the most frequent.

Only about 19% of the respondents indicated that their organization was fully aware of the risk, but when we looked more closely at these responses, a distinct pattern emerged. The firms that were identified as more fully aware of risk were most inclined to be implementing best practices, from modular design to the use of data validation.

One mechanism for dealing with risk at the level of spreadsheet development is to employ specialized auditing software. In recent years, a host of auditing programs, compatible with Excel, have been invented and made available on the Internet (Grinde, 2004; O'Beirne, 2006). However, this kind of software is evidently undiscovered: less than 1% of the respondents knew of the use of such software in their organizations.

6. Summary and Conclusions

From reading the literature on end-user computing and spreadsheets, we form the impression that spreadsheets are growing in use and importance faster than organizations can cope with the consequences. This impression is certainly supported by much of the data in our survey. Two main conclusions seem evident:

- A gap exists between current practice and best practice; and this gap exists even at the level of basic and generic use.
- Management policies have not caught up with the issues raised by the increased reliance on spreadsheets.

To deal with these issues, the most effective place to start is with management policies. Putting in place a set of procedures to mitigate spreadsheet risk is analogous to implementing a quality management program: standards and training, for example, could represent key parts of the initiative. A deeper awareness of spreadsheet risk would also filter down to the elements of spreadsheet design and use, leading to more disciplined testing and documentation, greater motivation to use protections, and enhanced use of auditing software.

As stated earlier, the strengths of our survey are its size and coherence, but we cannot determine from this survey alone how typical these responses from MBAs are of the broader population. We have administered the survey to several other populations, and in subsequent papers, we hope to offer comparisons of these populations.

Our survey instrument could be used in other situations, and we welcome inquiries about administering the survey in different settings, with research purposes in mind. Readers interested in following up this way can contact the authors through the SERP website, <http://mba.tuck.dartmouth.edu/spreadsheet/index.html>.

References

- Berglas, A. and P. Hoare (1999) "Spreadsheet Errors: Risks and Techniques." *Management Accounting* 77, 46-47.

- Bradley, H. (2003) "Error-Free Excel." *Macworld* 20, 78.
- Brown, P. and J. Gould (1987) "An Experimental Study of People Creating spreadsheets." *ACM Transactions on Office Information Systems* 5, 258-272.
- Cale, E. (1994) "Quality Issues for End-User Developed Software." *Journal of Systems Management* 45, 36-39.
- Chan, Y. and V. Storey, (1996) "The Use of Spreadsheets in Organizations: Determinants and Consequences." *Information and Management* 31, 119-134.
- Cragg, P. and M. King (1993) "Spreadsheet Modelling Abuse: An Opportunity for OR?" *Journal of the Operational Research Society* 44, 743-752.
- Croll, G. (2004) "A Typical Model Audit Approach." *EuSpRIG Conference Proceedings*, London.
- Croll, G. (2005) "The Importance and Criticality of Spreadsheets in the City of London." *EuSpRIG Conference Proceedings*, London.
- Davies, N. and C. Ikin (1987) "Auditing spreadsheets." *Australian Accountant*, 54-56.
- Doll, W. and M. Ahmed (1985) "Documenting Information Systems for Management: A Key to Maintaining User Satisfaction," *Information & Management* 5, 221-226.
- Doll, W. and Torkzadeh, G. (1987) "The Quality of End-User Documentation." *Information and Management* 12, 73-78.
- Floyd, B., J.Walls, and K. Marr (1995) "Managing Spreadsheet Model Development," *Journal of Systems Management* 46, 38-43.
- Gable, G., C. Yap, and M. Eng (1991) "Spreadsheet Investment, Criticality, and Control," *Proceedings of the Twenty-Fourth Hawaii International Conference on System Sciences*, 3, 153-162.
- Grinde, R. (2004) "Spreadsheet Development, Auditing, and Control Tools." *INFORMS Conference Proceedings*, Denver.
- Grossman, T., V. Mehrotra, and O. Özlük (2005) "Spreadsheet Information Systems are Essential to Business." *Working paper*.
- Hall, M. (1996) "A Risk and Control-Oriented Study of the Practices of Spreadsheet Application Developers." *Proceedings of the 29th Annual Hawaii International Conference on Systems Sciences*, 364-373.
- Janvrin D. and J. Morrison (2000) "Using a Structured Design Approach to Reduce Risks in End User Spreadsheet Development." *Information and Management* 37, 1 - 12.
- Klobas, J. and T. McGill, (2004) "Spreadsheet Knowledge: Measuring What User Developer Know." *Journal of Information Systems Education* 15, 427-436.
- Kreie, J., T.Cronan, J.Pendley, and J. Renwick (2000) "Applications Development By End-Users: Can Quality Be Improved." *Decision Support Systems* 29, 143-152.
- Kruck, S. and S. Sheetz (2001) "Spreadsheet Accuracy Theory." *Journal of Information Systems Education* 12, 93-106.

- Lawrence, R. and J. Lee (2004) "Financial Modelling of Project Financing Transactions." *Institute of Actuaries of Australia Financial Services Forum*. 19.
- McGill, T. (2002) "User-Developed Applications: Can End Users Assess quality?" *Journal of End User Computing* 14, 1-15.
- O'Beirne, P. (2006) "Spreadsheet Auditing and Inspection Tools." www.sysmod.com/sslinks.htm#auditing.
- Olphert, C. and J. Findlay (2004) "Validation of Decision-Aiding Spreadsheets: The Influence of Contingency Factors." *Journal of the Operational Research Society* 55, 12-22.
- Panko R. (1998) "What We Know About Spreadsheet Errors." *Journal of End User Computing* 10, 15-21. See also <http://panko.cba.hawaii.edu/ssr/Mypapers/whatknow.htm>
- Panko, R. and R. Halverson (1997) "Are Two Heads Better than One? (At Reducing Errors in Spreadsheet Modeling?)" *Office Systems Research Journal* 15, 21-32.
- Panko R, and R. Sprague (1998) "Hitting the Wall: Errors in Developing and Code Inspecting a 'Simple' Spreadsheet Model." *Decision Support Systems* 22, 337 - 353.
- Pemberton, J., and A. Robson (2000) "Spreadsheets in Business." *Industrial Management and Data Systems*, 100, 379-388.
- Pryor, L. (2006) "What's the point of documentation?" *EuSpRIG Conference Proceedings*, Cambridge, UK.
- Ronen, B., M. Palley, and H. Lucas (1989) "Spreadsheet Analysis and Design. *Communications of the ACM* 32, 84-92.
- Sajaniemi, J., and J. Pekkanen (1988) "An Empirical Analysis of Spreadsheet Calculation." *Software – Practice and Experience* 18, 583-596.
- Scheubrein, R. (2003) "Elements of En-User Software Engineering." *INFORMS Transactions on Education* 4, 37-47.
- Schultheis, R. and 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.
- Speier, C. and C. Brown. (1996) *Proceedings of the 29th Annual Hawaii International Conference on System Sciences*, Kihei, 374-383.
- Teo, T. and J. Lee-Partridge (2001) "Effects of Error Factors and Prior Incremental Practice on Spreadsheet Error Detection: An Experimental Study." *Omega* 29, 445-456.
- Teo, T. and Tan, M. (1997) "Quantitative and Qualitative Errors in Spreadsheet Development." *Proceedings of the 30th Annual Hawaii International Conference on System Sciences*, Wailea, 149-156.
- Torkzadeh, G. and W. Doll (1993) "The Place and Value of End-User Documentation." *Information and Management* 24, 147-158.
- Torkzadeh, G. and J. Lee (2003) "Measures of Perceived End-User Computing Skills." *Information and Management* 40, 607-615.
- Yoon, Y. (1995) "Applications-Oriented Spreadsheet Instruction in Financial Model Building: An Extension." *Financial Practice and Education* 5, 143-147.