Spreadsheet Risk, Awareness, and Control

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

Using data from a survey of MBA graduates, we developed a simple diagnostic tool to help assess the level of spreadsheet risk in an organization. We also found that in firms where spreadsheets pose a high risk, users are more aware of this risk. We review the best practices they employ at various stages of spreadsheet use, and we describe the steps that managers can take to control spreadsheet risk.

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

Spreadsheet Risk, Awareness, and Control

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 spreadsheets also has a downside. Stories of business failures, lawsuits, and governmental investigations sometimes appear in the press or on the Internet, with errors in spreadsheet use cited as the reason for the mishap [1]. The advent of Section 404 of the Sarbanes-Oxley Act has focused management attention on the control of spreadsheets. 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. They point out that spreadsheet use carries with it certain risks and costs and therefore companies should pay more attention to the way spreadsheets are managed.

Although it is becoming more evident that spreadsheets can pose a risk to companies, there is little practical information to aid managers in assessing the degree of risk, determining how aware their employees are of that risk, and developing ways to mitigate the risk. While spreadsheet use is almost ubiquitous in the business world, formal management practices governing spreadsheet use are rare, and relatively little is known about the practices that are used in companies where spreadsheet risk is recognized. This paper addresses the problem of spreadsheet risk in practical ways by

- introducing a diagnostic tool that can aid managers in determining the risk spreadsheets pose to their companies,
- informing managers about the spreadsheet practices used by companies where risk awareness is high,
- recommending steps that managers can take to reduce the risk represented by spreadsheet use.

In the next section we present a brief questionnaire that can help identify organizations that may be facing significant spreadsheet risks. Then we discuss our own survey data, specifically comparing companies characterized as being aware of spreadsheet risk with those characterized as being relatively unaware. This comparison reveals several best practices that can be found at the more aware companies. Finally, we discuss actions that are available to managers who wish to mitigate spreadsheet risk.

What's your risk score?
In our work [2] with the Spreadsheet Engineering Project (SERP), we've found that the role of spreadsheets in an organization is directly linked to the degree of risk that spreadsheets pose for that organization. Specifically, we found that the degree of risk that spreadsheets pose is directly linked to the importance of spreadsheets to the work of the organization, the sophistication of spreadsheet models, the number of spreadsheet users, and frequency of use. Based on our

research [3], we developed and validated a five-question diagnostic questionnaire that will help the managers ascertain the degree of risk that spreadsheets present for their firms (See Table 1).

In a survey of 1252 respondents, we found that firms most frequently earned scores in the Medium Risk category (47%). About 24% of the scores were in the Low Risk range, while 29% were in the High Risk category.

The main part of our sample was composed of over 700 MBA alumni from two prominent business schools, one in the U.S. and one in Europe [4]. These respondents represented a wide variety of firms and functions, and their responses to the questionnaire provided us with a unique look into spreadsheet use in the contemporary business world. From the survey, we found that spreadsheet risk tends to reflect certain key features of an organization. Based on our statistical analysis, we developed the five-question diagnostic tool in Figure 1. We then validated our results on a second sample, which was composed of over 500 respondents obtained from the mailing lists of two prominent software houses. As was the case for the MBA alumni sample, these respondents represented a variety of firms and functions. Finally, we combined the two samples to create the survey data we discuss in the following sections.

Table 1. Spreadsheet Risk Questionnaire

Spreadsheet Risk Diagnostic Tool

Answer each question about spreadsheet use in your organization to the best of your knowledge. Add the scores (number in parentheses by each answer) to get a total spreadsheet risk score.

1. How important are spreadsheets in your organization?

_____ Not at all important (1)
_____ Somewhat important (2)
_____ Important (3)
_____ Very important (4)

2. What is the size of the spreadsheet models generally created?

_____ under 100 cells (1)
_____ 101 to 1,000 cells (2)
_____ 1,001 to 10,000 cells (3)
_____ 10,001 to 100,000 cells (4)
_____ over 100,000 cells (5)

3. How many other users are there for a typical spreadsheet?

_____ None (1)
_____ 1 other person (2)
_____ 2-5 other people (3)
_____ 6-10 other people (4)
_____ More than 10 other people (5)

4. How often is a spreadsheet used after it is developed?

_____ Annually (1)
_____ Quarterly (2)
_____ Monthly (3)
_____ Once or twice a per week (4)
_____ Daily (5)

5. What are spreadsheets used for in your organization? (Check all that apply)

_____ Analyzing data (e.g. financial, operational) (1)
_____ Determining trends and making projections (1)
_____ Statistical analysis (1)
_____ Optimization (e.g. Solver, *What's Best*) (1)
_____ Simulation (e.g. *Crystal Ball, @Risk*) (1)

Total Score:_____

The sum of the scores associated with the answers to the five questions yields a Total Risk Score for the company. Three categories of risk differentiate between those firms where the potential for spreadsheet risk is most likely low, and where risk is most likely high.

CATEGORIES OF RISK
12 or below = **Low Risk**; 13-16 = **Medium Risk**; 17 or above = **High Risk**
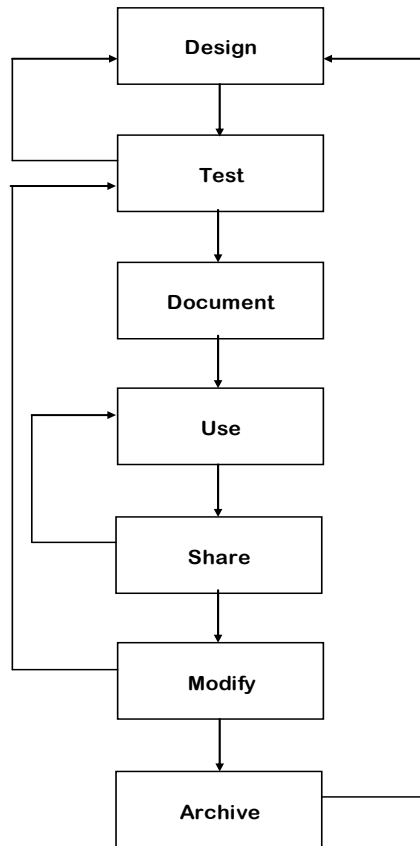
Awareness of Risk

Apart from the level of perceived risk in an organization, some firms have at least acknowledged the risk associated with spreadsheet use. Awareness of spreadsheet risk is critically important for managing that risk. Companies that are more aware of spreadsheet risk are more likely to encourage best practices among spreadsheet users. Thus, in analyzing our survey results, we used reported awareness of risk as a guide to the companies that promoted best practices.

At each stage of spreadsheet use, the opportunity for risk exists. The stages of a spreadsheet life cycle are: designing, testing, documenting, using, modifying, sharing, and archiving. (See Figure 1).This risk accumulates as the spreadsheet moves through the various stages of its life cycle. Opportunities for poor practice, carelessness, and serious mistakes exist throughout those stages, and the risk may even be magnified as the spreadsheet matures. Awareness of risk at each stage ultimately leads to the adoption of best practices.

For each stage, we next discuss the risks and identify important best practices used at firms that are most aware of spreadsheet risk. We refer to these firms as high-awareness (HA) firms. As a contrast, we also provide information from participants who work at low-awareness (LA) firms.

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

```
            ┌──────────────┐
        ┌──→│    Design    │←──┐
        │   └──────────────┘   │
        │          ↓           │
        │   ┌──────────────┐   │
    ┌───┼──→│     Test     │   │
    │   │   └──────────────┘   │
    │   │          ↓           │
    │   │   ┌──────────────┐   │
    │   │   │   Document   │   │
    │   │   └──────────────┘   │
    │   │          ↓           │
    │   │   ┌──────────────┐   │
    │   └──→│     Use      │   │
    │       └──────────────┘   │
    │              ↓           │
    │       ┌──────────────┐   │
    │       │    Share     │   │
    │       └──────────────┘   │
    │              ↓           │
    │       ┌──────────────┐   │
    └──────→│    Modify    │   │
            └──────────────┘   │
                   ↓           │
            ┌──────────────┐   │
            │   Archive    │───┘
            └──────────────┘
```

Best Practices over the Spreadsheet Life Cycle

Design. The design phase of spreadsheet development is the most critical for incorporating best practices because design consequences last through the entire life of the spreadsheet. Good design practices can accelerate development and reduce mistakes. Researchers have demonstrated that spreadsheet errors are difficult to detect when spreadsheets are poorly designed [5]. We found three areas of design practice where there was a marked difference in practices between high-awareness (HA) and low-awareness firms (LA).

Plan the spreadsheet first. Best practice in spreadsheet design suggests that, prior to working directly on the computer, a developer should sketch the spreadsheet or create a plan for it. This step delineates the scope of the project and provides an opportunity to organize the model's logic, avoiding extensive rework at later stages of development. In LA companies, 60% of the respondents reported that they typically began by entering data and formulas directly into the computer. In HA companies, by contrast, fewer than half indicated that they proceeded directly to the keyboard.

- *Create modules*. As in other forms of computer programming, modularization in design reduces the likelihood of errors. For example, in financial analysis, a module could be devoted to the components of revenue, another module to the components of cost, and a third module to the calculation of after-tax profit. Slightly over 71% of the HA sample indicated that they always or usually created modules, compared to 55% of the LA sample.

- *Separate Inputs from Calculations*. Another desirable practice is to create separate areas for data and formulas. Formula cells should generally not contain input data. Instead, input data should appear separately, and formulas should reference those cells. Thus, a design could contain a module for data, a separate module for detailed calculations, and a third module for summary information and charts. The application of this practice was similar to the previous one: about 68% of the HA sample usually or always followed the practice compared to 57% of the LA sample. At the other end of the spectrum, respondents who never separated inputs from calculations were 2½ times more prevalent in the LA sample than in the HA sample.

Test. The testing phase of spreadsheet design is commonly recognized as necessary, although companies may not have a formal policy on the practice. Most spreadsheets are, however, tested to some degree prior to use. The most frequent testing method mentioned in the literature seems to be checking individual formulas with test data. Another testing practice is to replicate normal cases—that is, to generate observed outputs from historical input data. However, it is much less common to use test data at the limits of the normal range or to use erroneous inputs, both of which might be revealing. Except in the work of the most disciplined professional programmers, testing always seems to get less attention than it deserves.

Most spreadsheets are tested prior to use, although the degree to which this is important seems to vary according to perceived risk. We found that 37% of the HA companies reported that they always tested their spreadsheets, compared to 21% of the low-risk-spreadsheet firms. We also found the following testing methods were used to some degree in both kinds of organizations.

- *Check with a calculator*. Selected cells, usually those involving complicated formulas, can be verified by making a parallel calculation on a hand calculator. This type of check was used by 42% of the HA respondents and 37% of the LA respondents.

- *Test extreme cases*. A useful testing method uses test data at the limits of the normal range or even erroneous inputs, both of which might be revealing. We found that 54% of the HA sample used this method, compared to 41% of the LA sample.

- *Display all formulas*. This display, officially known as Formula Auditing Mode in Excel, was used by 20% of the HA respondents and 17% of the LA respondents.

- *Use the formula auditing toolbar*. This method, sometimes called the "blue arrow tool" in Excel, was used by 37% of the HA respondents and 20% of their LA counterparts.

- *Use common sense.* Using "common sense" seemed to be the only testing mechanism adopted by a majority of respondents from both types of companies (73% of the HA sample and 67% of the LA sample.) Aside from the fact that this is a vague method, research

indicates that developers tend to be overconfident about the quality of their spreadsheets [6], so a developer's common sense is likely to be fallible.

Although there are other testing methods, they appeared to be used infrequently. Thus, our inference is that testing is not done extensively, but the HA companies are further along in their use of testing techniques. Indeed, only 11% of the respondents from HA companies claimed that they never tested their spreadsheets, whereas the percentage was nearly double among LA respondents (21%).

Document. Documentation is a critical aspect of spreadsheet development that helps users understand how the spreadsheet works. Better understanding on the part of users leads to easier use, greater accuracy, more timely response, and greater satisfaction with the spreadsheet. Documentation is also important for subsequent maintenance when multiple users exist and later revisions are likely. We found that 11% of the HA respondents claimed they always documented their spreadsheets, compared to less than 5% of their LA counterparts.

Some basic pieces of information to document include: assumptions and limitations, data sources, changes, and contact information for the developer. This information can be provided on one or more separate worksheets. The most popular documentation methods used by HA companies also include:

- *Text in cells.* The most convenient and simple form of documentation is to write text in the cells of the spreadsheet. This mechanism was used by about 65% of both samples.

- *Cell comments*. Excel's cell-comment feature makes it possible to explain key formulas and logical steps. This option was used in 62% of the HA companies and 54% of the LA companies.

- *Documentation sheet*. Devoting a worksheet entirely to documentation is less common, although it is particularly appropriate for workbooks containing multiple worksheets. This approach was used in 36% of the HA companies and only 21% of the LA companies.

Use. It is generally acknowledged that spreadsheets are used frequently and in just about every area of business. Various spreadsheet tools have been introduced that are tailored to specific functional requirements in business, so spreadsheets are used across several functions as well as across companies.

When it comes to spreadsheet use, relatively 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 actual picture is far more complex. Sometimes, the developer is also the user, and the two roles merge. At other times, there are many users: members of a team, analysts in a second department, or managers in several departments and functions.

As our diagnostic tool indicates, the number of users can be an indicator of organizational risk. In fact, in our combined sample, 34% of the respondents from High Risk firms indicated that their spreadsheets were typically used by more than 5 people. In Low Risk firms, by contrast,

only 2% of spreadsheets are used by more than 5 people. As our diagnostic tool also indicates, larger spreadsheet models and more frequent use are also indicators of potential risk. However, these factors are seldom choices made by the organization: they simply reflect the way business is conducted.

Risk is also associated with how spreadsheets are deployed, and it is instructive to compare how high-awareness and low-awareness firms use spreadsheets. More than their LA counterparts, respondents from the HA companies indicated that they used spreadsheets for tracking data, analyzing data, making projections, and evaluating alternatives. The LA respondents were comparatively more active at maintaining lists. The HA respondents also tended to employ spreadsheets for sophisticated purposes, such as

- *performing statistical analyses* (69% vs. 56%)
- *implementing optimization models* (55% vs. 47%)
- *running simulation studies* (34% vs. 26%)

When we looked in some detail at specific spreadsheet features and functions, we found that the HA sample used more of them and used them more frequently.

Share. Two opposing trends seem to play out in the sharing of spreadsheets. On the one hand, users are increasingly developing their own spreadsheets. Self-use has been encouraged by a friendly and sophisticated user interface as well as the desire of many organizations to become leaner. On the other hand, the use of networks and cross-functional problem-solving has promoted team use and other forms of sharing. Both trends account for increased use of spreadsheets, even if the driving forces have been different. Among the HA respondents, 75% stated that they worked alone when developing a spreadsheet, whereas the figure was 87% among the LA respondents.

We found that the HA companies tended to share spreadsheets more often than their LA counterparts. Sharing of the entire model was prominent (71% vs. 68%, respectively), although sometimes they shared summaries of the results (42% vs. 35%).

Protections can be implemented to reduce the risks associated with spreadsheet sharing. Cell protection and version control have received greater attention as concerns have increased about security and integrity of organizational data. Given that unprotected spreadsheets are vulnerable and that sharing is commonplace, we expected to find more reliance on simple protections than our survey indicated. However, we found that a majority of the users from both types of firms still employ no form of protection. The HA respondents were more inclined to implement simple protections. This is how the practice of the two groups compares on usage of three types of protection:

- *Cell protection*. Worksheets and key calculations can be protected from accidental invasion by users. The HA sample was more likely to employ this method than the LA sample (29% vs. 23%).

- *Password protection*. With little effort, worksheets and workbooks can be protected by the use of passwords. Again, the HA sample was more likely to implement this approach (30% to 21%).

- *Data validation*. Excel's Data Validation tool prevents the user from entering input data that is outside a specified range. This technique is simple to use, but most users are unfamiliar with it. The HA respondents were again more likely to take advantage of the tool, 15% to 8%.

Modify. Modifying spreadsheets is a common activity. The time between modifications is often quite short because business demands change frequently. We found that for both kinds of companies, most respondents indicated that the lifetime of their spreadsheets ranged from a few weeks or months to a year or two. When spreadsheets have been used and are then subject to modification, version control becomes a problem. Here is how practice compares on two types of controls:

- *Version numbers*. A majority of the HA companies (unlike their LA counterparts) at least used version numbers (54% vs. 37%).

- *Version control*. Other forms of version control for modifying spreadsheets exist, and some firms have developed elaborate control procedures. Some form of version control was used more frequently by HA companies than by LA companies, 76% to 63%

Responsibility for making the required modifications depends on the circumstances. We found that, for the HA companies, more spreadsheet modifications were made by new developers (28% vs. 14%) or by users (37% vs. 30%), than by the original developers.

Archive. The need for a spreadsheet archive is often motivated by legal and financial requirements. Aside from such considerations, an archive can serve as a kind of library, where developers may examine the types of spreadsheets that previous developers have created. Archiving is important to maintain institutional knowledge and to provide a database of existing and available spreadsheets ready for use. Such an initiative should pay off well in terms of reducing development time, minimizing rework, and preventing errors. (Although archiving is listed as the last stage of the life cycle, our diagram also shows that archiving potentially affects the very first steps in design.) The effective use of an archive requires two commitments: spreadsheets that have proven effective must be stored in the archive, and spreadsheet developers must explore the archive when building a new model. Here is how practice compares on the use of archives:

- *Borrow the design.* We found that 21% of the HA respondents borrowed a design from another spreadsheet, compared to 19% of the LA respondents.

- *Make proven spreadsheets widely available*. We found that 17% of the HA respondents noted that their archived spreadsheets "frequently" served as a knowledge or reference base for subsequent creators or users, compared to 13% of the LA respondents.

Different types of information are necessary to maintain an effective spreadsheet archive. Both kinds of companies recorded information on the title and date of the spreadsheet in their spreadsheet archives, but HA companies kept more extensive records about their archived files.

Management Initiatives

If we step back and examine the survey results, we might conclude that spreadsheet risk arises from inadequate safeguards taken during the spreadsheet life cycle. It seems clear that spreadsheet developers spend too little time testing and documenting their work, and users pay too little attention to the various forms of protection that are currently available. Thus, when people speak of current "practice" involving spreadsheets, they usually refer to the practice of individuals, and it is not difficult to identify a gap between current practice and best practice. Nevertheless, *organizations* also establish practices explicitly or implicitly, and these practices, in turn, influence spreadsheet use by individuals. From our survey we drew lessons for managers in three areas of organizational practice: training, standards, and controls.

Training. First, we asked about the types of training that organizations provided. Figure 2 shows the different responses.

Figure 2. Types of training offered by firms in the survey
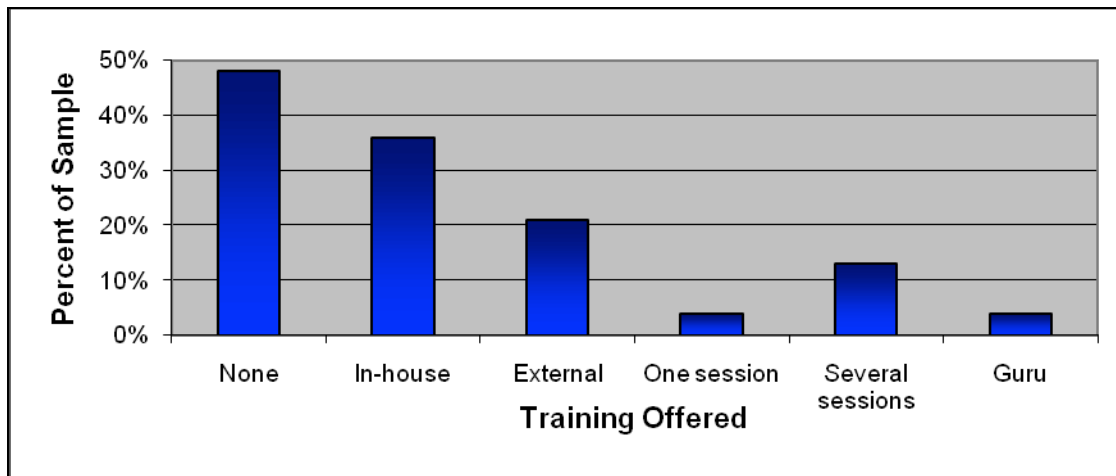
About half the organizations provide no training, which seems surprising given how widely spreadsheets have come to be used. In addition, the use of a spreadsheet specialist is rather unusual. We call it unusual in light of the fact that many users have told us they've learned and enhanced their skills 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. Figure 3 summarizes the results on the amount of training offered and used.

Figure 3. Training days offered and used
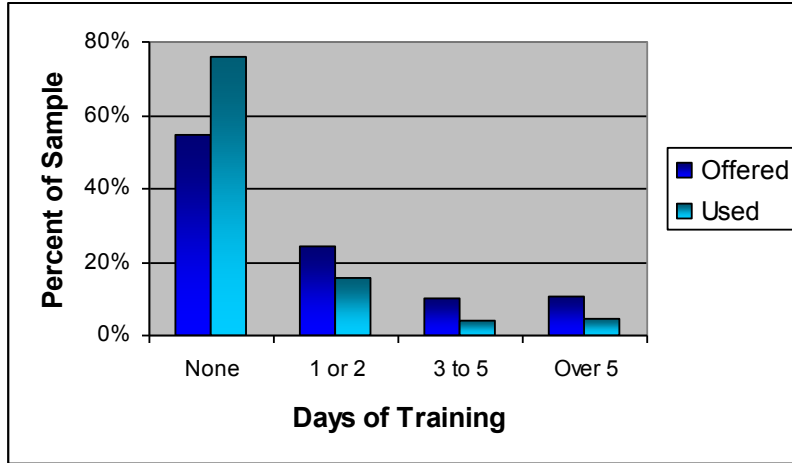
Thus, the training that is offered remains underutilized, although part of the reason may be that the coverage is too elementary for the needs of the target population. When asked about the biggest impediment to participating in company-sponsored training programs, the most frequent answer was "not enough time." However, we can imagine that same answer might be given as well to questions about other kinds of self-improvement activities. More specifically, when we asked about incentives provided by employers for participating, we found that less than a quarter of the firms paid for the cost of training, and other incentives were rare.

Standards and Policies. In our field interviews, we encountered organizations that have promulgated a set of required standards for spreadsheets. These requirements 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 uncertainties in a client's decision-making environment. However, in our experience, such initiatives are rare. This observation was supported by the responses in our survey, as shown in Figure 4.

Figure 4. Types of standards offered by firms in the survey.
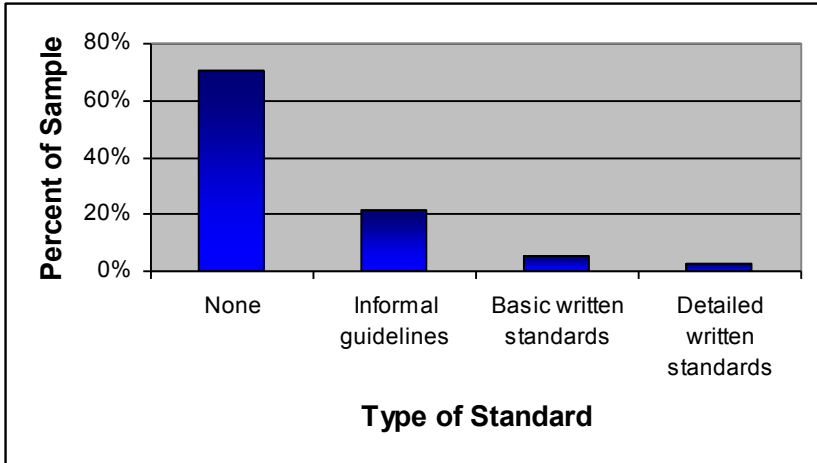
We recognize that the scope of "informal guidelines" may cover a wide spectrum of possibilities, but if we focus on written material on this subject, it would seem to exist in fewer than 10% of the organizations represented in our survey. Otherwise, people may suspect that standards exist, but they see little hard evidence. This attitude is consistent with the results shown in Figure 5.

Figure 5. Extent to which standards and policies are followed.
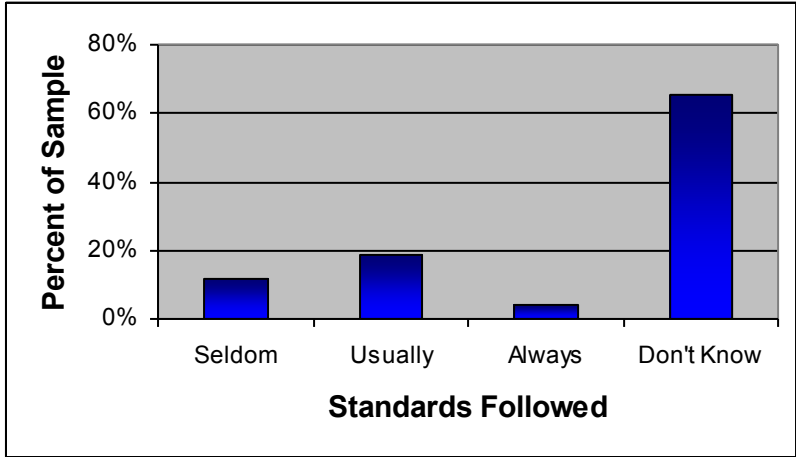
Clearly, if half the population doesn't know whether standards and policies are being followed, there is ample room for improvement. Note here that standards can address features of spreadsheets, but they can also describe the procedures used in managing spreadsheets at the various stages of the spreadsheet life cycle.

Controls. Risk exists if mistakes are likely 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 our sample classified it as "moderately important," "very important," or "critical." In particular, about half of the sample felt that the risks to their organization posed by spreadsheets were at either a medium level or a high level.

A deeper look at the replies revealed that more spreadsheet risk is detected by the more experienced spreadsheet users. And although we might have suspected otherwise, higher levels of spreadsheet risk are perceived in large companies than in small companies. Higher levels of spreadsheet risk are also detected by those users who are more inclined to employ best practices. This result suggests that making employees aware of the risks may motivate them to pay closer attention to the quality of their spreadsheet work.

Only about 18% of the respondents indicated that their organization was "fully aware" of spreadsheet risks, and fewer than this number knew of organizational strategies in place to deal with risk. 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.

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 [7]. However, this kind of software is not widely used: less than 3% of our respondents could attest that such software was being used in their organizations.

Summary and Conclusions

The literature on end-user computing and spreadsheets suggests that spreadsheets are growing in use and importance faster than organizations can cope with the consequences. That impression is certainly supported by much of the data in our survey. Two main conclusions were 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.

Some firms have become aware of spreadsheet risk more rapidly than others, and we found in our survey that those firms promote a number of "best practices." However, even in the high-awareness firms, such practices are far from universal. The most effective way to deal with risk is through management policies. Putting in place a set of procedures to mitigate spreadsheet risk is analogous to implementing a quality management program: training, standards and controls could represent key parts of the initiative.

- Training needs to be more accessible, more sophisticated, and visibly supported by the organization.

- Standards and policies need to be articulated in more detail, preferably in writing, and awareness of spreadsheet risk must be promoted more widely.

- Controls need to be put in place to ensure that standards are followed and that appropriate incentives exist.

Managerial initiatives of this sort would enhance the awareness of spreadsheet risk among individuals working with spreadsheets. This awareness, in turn, 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. In short, well-aimed management initiatives will lead to best spreadsheet practices.

Notes

1. A collection of such stories has been compiled by the European Spreadsheet Risk Interest Group (EUSPRIG) at www.eusprig.org/stories.htm.
2. Additional information about project activities can be found at the project's website, http://mba.tuck.dartmouth.edu/spreadsheet/index.html.
3. The risk diagnostic tool was developed using a sample of over 700 MBA alumni from two prominent business schools, one in the U.S. and one in Europe. These respondents represented a wide variety of firms and functions, and their responses to the questionnaire provided us with a unique look into spreadsheet use in the contemporary business world. We then validated our results on a second sample, which was composed of over 500 respondents obtained from the mailing lists of two prominent software houses.
4. For an overview of other surveys that have been aimed at spreadsheet use, see B. Lawson, et al., "A comparison of spreadsheet users with different levels of experience," SERP Working Paper, http://mba.tuck.dartmouth.edu/spreadsheet/index.html.
5. See, for example, 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.
6. The overconfidence phenomenon has been explored in several papers. See, for example, Davies, N. and C. Ikin (1987). Auditing spreadsheets. *Australian Accountant*, 54-56, Floyd, B., J.Walls, and K. Marr (1995) "Managing Spreadsheet Model Development," *Journal of Systems Management* 46, 38-43, and 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.
7. For more information on auditing software for spreadsheets see Patrick O'Beirne's website, http://www.sysmod.com/sslinks.htm#Nixon or the SPRIG website, http://sprig.section.informs.org/audit.html.