Abstract

Spreadsheets are thought to be highly prone to errors and misuse. In some documented instances, spreadsheet errors have cost organizations millions of dollars. Given the importance of spreadsheets, little research has been done on how they are used in organizations. We review the existing state of understanding of spreadsheet errors, concentrating on two recent studies. One analyzes errors in 50 operational spreadsheets; the other studies the quantitative impacts of errors in 25 spreadsheets from five organizations. These studies suggest that counts of error cells are not sufficient to understand the problem of errors. Average error cell counts reported in the literature range from 1 percent to 5 percent depending on definitions and methods used. However, some errors are benign while others are fatal. Furthermore, spreadsheets in some organizations appear to be error-free. Several types of new research are needed to understand the spreadsheet error problem more fully.

1. Introduction

Spreadsheets are the dominant desktop application in contemporary organizations. Millions of ordinary workers acting as end-user programmers design, build, and use spreadsheets every day. Whether these computer systems serve the organizations that use them in a positive way is therefore a question of enormous practical significance.

Many authors and spreadsheet practitioners have warned about the ubiquity and dire consequences of errors in spreadsheets. A widely cited estimate [3] is that 5% of all formulas contain errors. But the most relevant question is not how common errors are but how significant they are to their users. After all, there is strong evidence that humans make frequent errors in many areas but few of these errors are consequential [9].

This paper describes the current state of the art in spreadsheet error research. After briefly summarizing the existing literature, we concentrate on two recent large-scale studies of errors and their impacts. While these studies alter our understanding of spreadsheet errors in some ways, they also highlight many areas in which our knowledge is lacking. The paper closes with a discussion of some promising lines of research.

2. Existing Literature

We recently summarized the research literature on spreadsheet errors [5]. That paper organized the literature into five areas:

- categories of errors
- impact of errors
- frequency of errors
- creation and prevention of errors
- detection of errors

We summarized the existing literature on categories of errors as follows:

- Classifications are generally offered without specifying the context or purpose for which the classification is intended.
- The existing classifications do not include sufficient examples of specific errors that satisfy each category.
- Classifications are generally not rigorously tested to demonstrate that multiple users can consistently classify actual errors into the proper categories.
- The boundary between quantitative errors and qualitative errors remains vague.

A number of studies have attempted to measure the frequency of errors in spreadsheets. Panko [3] summarized the results of seven field audits in which operational spreadsheets were examined. His results show that 94% of spreadsheets have errors and that the average cell error rate (the ratio of cells with errors to all cells with formulas) is 5.2%. However, in our analysis of these studies we concluded that these estimates are unreliable given that the studies are not sufficiently documented or comparable. A better estimate of the cell error rate based on these studies would be 1.3% of cells, but only 13 spreadsheets can be included in this average (Note that corrected cell and spreadsheet error rates now appear on Panko’s website [4]).

The least studied aspect of spreadsheet usage is the quantitative impact that errors have on the results. What little evidence we have on the impact of errors
errors in audited spreadsheets is largely anecdotal. For example, Hicks (cited in [3]) reported that the errors found in the audit of a single large spreadsheet caused the results to be off by 1.2%. Clermont [1] found errors in an average of 3.03% of cells in three large spreadsheets but reported that “we did not find any tremendous erroneous result values that might have had severe negative effects on the company.” Lukasic (personal communication cited in [3]) found a 16% error in the results of one of two spreadsheets audited. Panko’s two interviewees [3] suggested that 5% of spreadsheets had “serious” errors.

Another source of information on the impact of errors is the stories that have appeared in the press documenting losses due to mistakes involving spreadsheets. The European Spreadsheet Risks Interest Group (EUSPRIG) maintains a web page [2] that documents dozens of these cases. Many of these incidents involve errors in formulas, but others involve errors in the use of spreadsheets, such as in sorting, or even errors in interpreting results. These types of errors, while potentially serious, are beyond the scope of the current research.

We can summarize the literature on the impacts of errors in spreadsheets as follows:

- No systematic studies have been reported that measure the quantitative impacts of errors in spreadsheets.
- Studies of the frequency of errors suggest around 1% of formulas are erroneous.
- Errors can be classified in many ways and no one classification has become accepted.

3. Errors in 50 Operational Spreadsheets

As we have seen, the previous studies of spreadsheet errors have been small in scale and in many cases are not well documented. Several years ago we set out to study a large sample of operational spreadsheets using a well-tested and fully-documented method [6, 7]. We stress that our subject was errors in completed, operational spreadsheets that had been in use for some time, not errors discovered in laboratory experiments. We used competent Excel users, not professional programmers, to conduct our study. And we did not have access to the spreadsheet developers. This study allowed us to identify a wide range of errors, but it had its limitations as well. One limitation was that we could only detect errors in the spreadsheets themselves, not errors in formulating the underlying problem or in the actual use of the spreadsheet. And, of course, we cannot guarantee that we found all the errors actually in these spreadsheets.

Our sample of 50 spreadsheets came from a variety of organizations and from several websites. While not a strict random sample, ours is typical of the general population of spreadsheets. No one organization or application area contributed more than five spreadsheets to the sample.

To assist us in analyzing these spreadsheets we used two software tools: XLAnalyst [11] and Spreadsheet Professional [10]. These software applications are add-ins to Excel that assist in auditing and building spreadsheets. XL Analyst evaluates 28 aspects of a spreadsheet, for example, formulas that evaluate an error. Spreadsheet Professional includes tools for building, testing, analyzing, and using spreadsheets. One of its most useful features for our purposes is the mapping tool, which translates each worksheet in a workbook into a map showing which formulas have been copied.

Our analysts were undergraduates, graduates, and recent alumni from our business or engineering schools. All of them had several years of experience with Excel, but none were professional programmers, even in Excel. Our training program took about 10 hours, which includes time to read the study procedure, analyze several sample workbooks, and receive feedback.

The protocol we used involves eleven steps:

1. Run the two software tools XLAnalyst and Spreadsheet Professional.
2. Transfer selected results from the software tools to a data record sheet.
3. Record the purpose of the workbook and each worksheet.
4. Examine the workbook for use of Excel functions.
5. Review the results of XL Analyst and use them to locate errors.
6. Review the Spreadsheet Professional maps and use them to locate errors.
7. Review the Spreadsheet Professional calculation tests and use them to locate errors.
8. Review all formulas not already reviewed for errors.
9. Conduct sensitivity analyses to uncover errors.
10. Rate the workbook on various aspects of spreadsheet design (e.g., use of modules).
11. Record the total time taken by the audit and record comments on any special situations encountered.

Since no generally-accepted taxonomy of spreadsheet errors exists we experimented with several before settling on the following six types of errors:

- Logic error - a formula is used incorrectly, leading to an incorrect result
- Reference error - a formula contains one or more incorrect references to other cells
- Hard-coding error - one or more numbers appear in formulas
• Copy/Paste error - a formula is wrong due to inaccurate use of copy/paste
• Data input error - an incorrect data input is used
• Omission error - a formula is wrong because one or more of its input cells is blank.

The senior researchers reviewed each workbook after the analysis was completed and made final judgments as to which errors would be accepted. Together accounted for less than 5% of instances and cells.

Not all of the 483 error instances we identified involved wrong results. In fact, roughly 40 percent of these were Hard-coding errors, in which a number was included in a formula rather than a cell reference; most of these did not produce wrong results. In order to determine how many errors actually gave rise to wrong results we categorized each instance as a Wrong Result or a Poor Practice. With this more restrictive definition, we found that 7 of the 50 spreadsheet workbooks we audited were error-free, for a spreadsheet error rate of 94 percent.

Table 1
Errors categorized by type

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Instances</th>
<th>Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard-coding</td>
<td>182 (37.7% of total)</td>
<td>2,111 (43.5% of total)</td>
</tr>
<tr>
<td>Reference</td>
<td>159 (32.9%)</td>
<td>1,074 (22.1%)</td>
</tr>
<tr>
<td>Logic</td>
<td>106 (21.9%)</td>
<td>1,389 (28.6%)</td>
</tr>
<tr>
<td>Copy/Paste</td>
<td>17 (3.4%)</td>
<td>206 (4.2%)</td>
</tr>
<tr>
<td>Omission</td>
<td>13 (2.7%)</td>
<td>65 (1.3%)</td>
</tr>
<tr>
<td>Data Input</td>
<td>6 (1.2%)</td>
<td>10 (0.2%)</td>
</tr>
<tr>
<td>Total</td>
<td>483</td>
<td>4,855</td>
</tr>
</tbody>
</table>

Although our inter-rater reliability was 92 percent, we found that errors come in hundreds of subtle variations that require thought and debate to categorize correctly. This experience deepened our skepticism of previous studies of errors in which the taxonomy of errors and the procedures used to find them were not spelled out.

Of the 50 spreadsheets we studied, three were found to have no errors of the types included in our protocol. This gives a spreadsheet error rate (percent of spreadsheets with at least one error) of 94 percent, which is the same as Panko’s estimate [3].

In the 47 spreadsheets with at least one error we found a total of 483 error instances involving 4,855 cells. Thus each error instance involves on average 10 cells. The average cell error rate implied by these results is 1.79 percent of all formula cells, considerably lower than the 5.2 percent error rate given by Panko [3].

In the remaining 43 spreadsheets we found a total of 281 instances of Wrong Results involving a total of 2,353 cells. Using this definition, the average cell error rate over all 270,722 formulas audited was 0.87 percent. This is, of course, substantially lower than the 1.79 percent cell error rate we estimated using our more inclusive definition of error, and much less than Panko’s estimate of 5.2 percent.

Our results suggest that the average cell error rate across all spreadsheets is in the range of 1 percent when we use a restrictive definition and 2 percent when we use an inclusive definition. But the average error rate does not tell us how the error rate varies across spreadsheets. The received wisdom is that errors are unavoidable in spreadsheets, as they are in most facets of life. This suggests that we should expect most spreadsheets to have a cell error rate close to the average.

Figure 1 shows the distribution of cell error rates using our original, inclusive definition. Using that definition, 46 percent of our sample spreadsheets had error rates below 2 percent; 70 percent had error rates below 5 percent. However, several spreadsheets had error rates above 10 percent, in fact one had a cell error rate of 28 percent. Figure 2 shows the error rates below 2 percent and 90 percent were below 5 percent. The highest single error rate was 17.
This study leads to the following tentative conclusions:

- Moderately-skilled Excel users can successfully analyze operational spreadsheets for errors with minimal training when provided with an explicit protocol to follow.
- Cell error rates average 1 or 2 percent, depending on the definition of error.
4. Impacts of Errors in 25 Spreadsheets

While the previous study revealed new and interesting patterns in the prevalence of errors in spreadsheets, it did not address one critical question: do errors have a substantial quantitative impact on the results? As our review of previous studies showed [5], this is a question that has received almost no attention in the past. Accordingly, we undertook a new study to estimate the quantitative impacts of errors [8].

Our sample in this study consisted of 25 spreadsheets, five each from organizations that allowed us access to their spreadsheet developers. This included two consulting companies, a large financial services firm, a manufacturing company, and an educational institution. These organizations identified five volunteers, each of whom provided one spreadsheet. We provided the following specifications to the volunteers for help in choosing a spreadsheet:

- contains 3-10 worksheets
- contains 250-10,000 formulas
- occupies 200-1,000 kb of memory
- has been in use for no more than a year
- contains no complex Visual Basic code
- is well understood by the developer
- has no broken external links

Not all the spreadsheets we audited conform to these specifications. In fact, the average number of sheets in our sample was 15.2 (the range was from 2 to 135 sheets) and the average size in kilobytes was 1,463 (the range was 45 to 7,450kb). Many of the spreadsheets in our sample were larger on one or more of the dimensions than specified above.

We used the same error taxonomy as in the previous study and the same general auditing protocol, although we eliminated many of the data gathering steps from the previous study as they were not necessary for our purposes.

Determining the quantitative impact of an error in a spreadsheet is not as straightforward as it might appear. First, some errors occupy a single cell while others occupy many cells. Do we consider each cell as a separate error and measure the impact of correcting it alone, or do we correct all the cells with a similar error and measure the overall impact? Second, some error cells influence many other cells while others impact no other cells. When a cell impacts many other cells it is not always obvious which of the impacted cells to use to measure the effect. Third, it is not always clear how to correct an error. For example, if erroneous inputs were used do we replace them with average inputs or extreme inputs? Finally, it is necessary to decide whether to measure errors in absolute or relative terms, and whether to combine all the errors in a given workbook into one overall error or to treat them separately.

We chose to measure the effect of each error separately and then to report the maximum error in a workbook. In most cases we corrected all the cells with a given type of error, treating this as one error with a single (overall) impact. When such an error impacted only the erroneous cells themselves, we computed the maximum change from the base case and took that as our error estimate. When such an error impacted a single dependent cell, we measured the impact of correcting all the error cells on that one cell. When an error cell had many dependent cells, we identified the most important dependent cell and measured the impact of correcting the error on that cell. Obviously, the magnitude of the error impacts we identified are dependent to some extent on the measurement system used, although we believe our choices are justifiable and our results fairly reflect the quantitative accuracy of the workbooks tested.

Table 2 summarizes our results. In column 1 we have used a two-digit code to label each spreadsheet. For example, spreadsheet 3.4 is the fourth spreadsheet from organization 3. The table gives the following information for each spreadsheet:

- number of issues raised in our audits
- number of errors confirmed in interviews
- number of errors with non-zero quantitative impact
- maximum percentage impact
- maximum absolute impact

Many different kinds of issues turned out not to be errors after we discussed them with the spreadsheet developers. For example, if we found one odd formula in a column of identical formulas we flagged it as an issue. Typically the developers would then tell us that this case was calculated differently from all the rest and that it was correct, although there was no such indication in the spreadsheet itself. Most of these cases, in fact, involved inconsistent formulas in a column or row.

Errors with zero quantitative impact also arose in a variety of ways. Sometimes a formula referred to the wrong cell for an input, but the numerical value in that cell was the same as in the correct input cell. We categorized this case as an error, but one with zero impact. We recognize that this is a dangerous practice, which could lead to a nonzero error during
the use of the spreadsheet. However, for the purposes of this study we categorized it as an error with zero impact.

Within this sample of 25 spreadsheets we identified a total of 381 issues. After we discussed these issues with the developers we found that nine spreadsheets had no errors; among the remaining 16 spreadsheets we found a total of 117 errors. Of these 177 errors, 47 had zero quantitative impact, leaving 70 errors with non-zero impact.

Table 2

Errors and impacts in 25 spreadsheets

<table>
<thead>
<tr>
<th>Organization-Workbook Size (KB)</th>
<th># Issues</th>
<th># Errors</th>
<th>Errors with No Impact</th>
<th>Maximum Percentage Impact</th>
<th>Maximum Absolute Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 842</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>0.0%</td>
<td>$0</td>
</tr>
<tr>
<td>1.2 3,291</td>
<td>50</td>
<td>6</td>
<td>1</td>
<td>28.8%</td>
<td>$32,105,400</td>
</tr>
<tr>
<td>1.3 3,556</td>
<td>18</td>
<td>7</td>
<td>4</td>
<td>137.5%</td>
<td>$110,543,305</td>
</tr>
<tr>
<td>1.4 1,376</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0.0%</td>
<td>$0</td>
</tr>
<tr>
<td>1.5 1,678</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>2.1 3,180</td>
<td>19</td>
<td>6</td>
<td>1</td>
<td>3.6%</td>
<td>$13,909,000</td>
</tr>
<tr>
<td>2.2 520</td>
<td>27</td>
<td>11</td>
<td>4</td>
<td>16.0%</td>
<td>$74,000,000</td>
</tr>
<tr>
<td>2.3 394</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>2.4 156</td>
<td>30</td>
<td>4</td>
<td>1</td>
<td>416.5%</td>
<td>$10,650,000</td>
</tr>
<tr>
<td>2.5 1,734</td>
<td>40</td>
<td>2</td>
<td>0</td>
<td>NA</td>
<td>$0</td>
</tr>
<tr>
<td>3.1 389</td>
<td>19</td>
<td>2</td>
<td>0</td>
<td>5.3%</td>
<td>$238,720</td>
</tr>
<tr>
<td>3.2 177</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.0%</td>
<td>$0</td>
</tr>
<tr>
<td>3.3 47</td>
<td>11</td>
<td>2</td>
<td>0</td>
<td>15.6%</td>
<td>$4,930,000</td>
</tr>
<tr>
<td>3.4 990</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0.0%</td>
<td>$0</td>
</tr>
<tr>
<td>3.5 286</td>
<td>23</td>
<td>1</td>
<td>1</td>
<td>0.0%</td>
<td>$0</td>
</tr>
<tr>
<td>4.1 606</td>
<td>27</td>
<td>22</td>
<td>10</td>
<td>116.7%</td>
<td>$13,355,445</td>
</tr>
<tr>
<td>4.2 576</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>141.8%</td>
<td>$272,000</td>
</tr>
<tr>
<td>4.3 1,571</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>4.4 2,116</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>4.5 1,932</td>
<td>79</td>
<td>44</td>
<td>17</td>
<td>39.1%</td>
<td>$216,806</td>
</tr>
<tr>
<td>5.1 7,450</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>5.2 415</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>5.3 1,900</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>5.4 309</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>5.5 427</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Totals</td>
<td>381</td>
<td>117</td>
<td>47</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As we pointed out above, there are two ways to measure the impact of errors: absolute and relative. Absolute impacts are important because they tell us how large errors are in the units of the spreadsheet. However, they cannot be compared easily across workbooks, since a million dollar error may be trivial in one spreadsheet and catastrophic in another. Relative (or percentage) errors more accurately reflect the significance of an error, but they have their shortcomings as well. One problem with relative errors is that percentage changes cannot be determined when the initial value is zero; another is that percentage changes in percentages are not generally as meaningful as percentage changes in
dollar amounts. Accordingly, we present our results here in both absolute and relative terms.

Recall that 47 of the errors we found had zero impact on the spreadsheet. This often came about when a formula had an erroneous reference, but both the erroneous and the correct input cells had the same value. Thus when the error was fixed the results did not change.

Twelve of the errors involved percentages; among these the average absolute change was 22 percent. Twenty-four of the remaining 58 errors involved absolute errors less than $10,000. However, some errors were huge: the largest single absolute error we found was over $100 million!

Our evidence suggests that spreadsheet practice is very different among the five organizations we studied. Figure 3 shows the distribution of issues, confirmed errors, and errors with non-zero impact in the five organizations we studied. In the five spreadsheets from Organization 5 we could identify only five issues to discuss with the developers and no errors were identified among those five issues. Organization 5 is a small consulting company with highly educated employees and a culture that demands excellence. The spreadsheets we analyzed from this firm were works of art: thoughtfully designed, well documented, easy to understand, and error free.

Figure 3
Issues, errors and errors with impact, by organization

Organization 4 had two spreadsheets with no errors and two with 22 and 44 errors, respectively. The quality of the spreadsheet practice in this organization clearly depends on just where one looks. In the case we found both the best of practice and the worse of practice in offices just a few miles apart.

In Organization 3, which is another consulting company, all the spreadsheets we analyzed had errors but in three cases no error had a measurable impact on the results. Even in the remaining two spreadsheets the errors were few in number and fairly small in terms of impact.

Organizations 1 and 2 are both very large. One is a financial firm and the other is a manufacturing firm. Some of the spreadsheets we analyzed from these companies were astonishingly large and complex. Perhaps for this reason, only two of the ten we analyzed were error-free (although four had no errors with impact). The quality of spreadsheet practice in both of these companies was inconsistent, with inadequate attention paid to spreadsheet design, simplicity, ease of use, documentation, and consistency.

We can summarize our findings as follows:
• Some organizations have spreadsheets that are essentially error-free.
• Within a single organization, spreadsheet practice can range from excellent to poor.
• Some organizations use spreadsheets that are rife with errors and some of these errors are of substantial magnitude.
• Many errors have zero impact, or impact unimportant calculations.
• There is little correlation between the importance of the application or the risk involved and the quality of the spreadsheet.
• Few spreadsheets contain errors that, in the eyes of their developers, would destroy their usefulness.

5. Directions for Research

Of all desktop applications for business, the spreadsheet must rank second behind word processing as the most commonly used. While errors made in word processing may be embarrassing or annoying, they probably rarely lead to financial losses. By contrast, errors made in spreadsheets quite obviously can and do lead to such losses. It is therefore surprising how little research has been conducted on the subject of spreadsheet errors.

While our research has introduced some new data and hypotheses into the literature, we believe its value lies in raising better questions rather than in settling them. Our studies suggest that more groundwork is needed to define useful categories of errors and to develop more effective methods for identifying errors in operational spreadsheets. We believe that current estimates of the average cell error rate are sufficient, since it is evident that error rates vary widely among spreadsheets and organizations. Much more work is needed to refine our understanding of how error rates vary across the universe of spreadsheets and on the impacts of errors.

We suggest that four types of studies are needed at this point:

• Retrospective studies: study the impact that spreadsheet use had on a single decision in an organization
• Causal studies: analyze the determinants of spreadsheet quality at the individual and organizational levels. For example, how do training, support, and culture influence quality?
• Best practices studies: how did exemplary individuals and organizations develop their spreadsheet skills?
• Action studies: study the impact of a spreadsheet quality control program in an organization.

6. References