Struggling to Excel: A Field Study of Challenges Faced by Spreadsheet Users

Chris Chambers
Oregon State University
chambech@eecs.oregonstate.edu

Chris Scaffidi
Oregon State University
cscaffid@eecs.oregonstate.edu

Abstract

Spreadsheets have become one of the most widely-adopted software technologies. They have proven useful for performing numeric computations as well as for organizing, manipulating, exploring, and visualizing data. Yet only one aspect of spreadsheets, formulas, has received extensive attention in field studies to date. In this paper, we describe a three-part field study that widens this focus to uncover a broader range of challenges that people encounter when creating and using spreadsheets. This study has revealed several opportunities to improve spreadsheet editors, including developing different modes for spreadsheet creation, improving support for spreadsheet reuse, and helping users to find and use features.

1. Introduction

Spreadsheet editors have evolved far beyond the simple calculation engine called VisiCalc that first appeared over 30 years ago. A major shift away from pure calculation toward data visualization and exploration began in 1983, with the introduction of charting features in Lotus 1-2-3. This shift continues today in the now-predominant editor, Excel, which offers a dizzying array of graphing and data analysis features.

With the addition of these and other features, spreadsheets have become useful for an increasingly broad range of tasks. For example, even by the mid-1990's, spreadsheets were predominantly used for creating reports rather than performing computations [6]. We suspect, though do not know for certain, that the introduction of more sophisticated charting features in the past 14 years will have further increased the usefulness of Excel for reporting. More recently, education researchers have even argued that spreadsheet editors (and, in particular, their charting wizards) are now mature enough to be used as a platform for teaching K-12 students various skills related to basic mathematics, data exploration, and data visualization [7][13].

If the accumulation of data visualization and exploration features is indeed making spreadsheets more useful to a wider population of people, this might help to explain the rapid rise in spreadsheet adoption over the past 30 years. Projections indicate that spreadsheet editors will soon be used by over 55 million people in American workplaces alone [19].

But on the other hand, the usefulness of these new features might be overstated. They might be no more than "software bloat," an accretion of features that few people use and even fewer use properly. Bloat clutters user interfaces and can make it hard for people to find the features that they want. Moreover, designing and implementing features that nobody wants is a waste of money and time that could be invested elsewhere.

The fact is that we do not know whether and how these new features are used in practice. To date, only formula usage has been examined in close detail by empirical field studies of spreadsheet usage. Consequently, we lack empirical guidance for further invention, refinement, or perhaps elimination of spreadsheet features for visual exploration and analysis of data.

In response, we have conducted a three-part field study analyzing real-world spreadsheets, online user comments and user interviews, in order to answer the following questions:

Q1. What features of spreadsheet editors are people actually using in practice? We are particularly interested in use of charting features, given their potential importance to reporting and education.

Q2. What problems do people encounter when using features? If people are trying to use features, and failing, then this might suggest that these features are valuable and need to be improved rather than simply eliminated. We are specifically interested in problems encountered when using spreadsheets in education.

Q3. How could spreadsheet editors be improved through addition, enhancement or elimination of features? Analysis of spreadsheet errors might suggest ways to provide better spreadsheet editors, and users might offer also useful suggestions, since they know their needs better than anybody else.

Overall, our study has shown that Excel offers the functionality and flexibility to perform most of the tasks that users express an interest in accomplishing, but people often struggle to find needed features, particularly when creating charts. Moreover, effectively communicating information through charts sometimes requires substantial manual tweaking. In the context of education, one consequence of these barriers is that many teachers express reluctance to use spreadsheets to instruct students. Based on these results, we have identified several opportunities for significant improvement in spreadsheet editors, including developing different modes for spreadsheet creation, improving support for spreadsheet reuse, and helping users to find and use features.

The remainder of this paper is organized as follows. Section 2 summarizes related work, including the numerous prior studies of errors in spreadsheet formulas. Section 3 describes our methodology for analyzing spreadsheets and user comments, as well as for interviewing users. Sections 4, 5, and 6 present the analyses of spreadsheets, online user comments, and user interviews. Drawing on these three parts of the study, Section 7 addresses the research questions presented above and discussions implications for future research.

2. Related Work

Prior field studies of spreadsheet usage have largely focused on formulas, particularly formula correctness. This work has shown that spreadsheets are frequently “built in an informal, iterative manner” that omits quality control [3], with up to 94% of spreadsheets typically containing at least one formula error [15]. Formula errors come in many different varieties [16][17] that ultimately can lead to major monetary losses [1].

These results show that formula correctness is an important objective that deserves significant research, but formulas constitute only a slice of the features offered by spreadsheet editors. One 1996 survey of 256 spreadsheet users found that other features such as charts and macros were used just as often as several varieties of functions (financial, database, and goal-seeking), with only statistical functions seeing more use than charts or macros [2]. Another 1996 survey of 106 users found that 70% of spreadsheets were used for laying out reports, with only 30% for calculation-intensive activities like modeling or analysis [6]. A 2006 survey of 831 information workers found that while formulas were used by 90% of respondents or their subordinates, other features were also frequently used, such as charts by 80% and macros by 42% [18]. Consistent with these earlier studies, a 2007 analysis of 4498 spreadsheets on the web (the EUSES Spreadsheet Corpus) found that only 4% of cells contained formulas, and only 44% of spreadsheets had any formulas at all [4]. Finally, a 2009 survey of 1597 users found that charts were used more frequently (on a Likert scale) than every function except the “if” function, and that macros were more rarely used than charts or functions [9].

The implication of these studies is that other features besides formulas appear to deliver a major part of the value offered by spreadsheet editors. A recurring theme is the importance of charts. In fact, this was apparent as far back as the 1996 studies, and numerous additional charting features have since been introduced. Just a few examples are features for augmenting charts with regression lines, regression statistics, moving averages, custom shapes, error bars, series lines, drop lines, drop shadows, extra axes, and embedded data tables.

Charting has also risen in importance because of its value for data visualization and exploration activities, which have increasingly become essential components of K-12 and college education. As far back as 1998, the National Research Council noted that spreadsheets were well-suited for helping students to visualize functions, in order to understand what it means to solve equations [10]. More recently, in 2004, the National Council of Teachers of Mathematics issued standards stipulating that students should be able to “understand histograms, parallel box plots, and scatterplots and use them to display data” by the end of 12th grade, and to create “line plots, bar graphs, and line graphs” by the end of 5th grade [11].

In response to these calls to action, researchers have been developing spreadsheets for teaching subjects such as algebra [13], digital logic [7], and oscillatory motion [14]. Various companies such as Key Curriculum Press[1] also offer specialized software such as the Geometer’s Sketchpad and the Fathom spreadsheet editor, which contain specialized feature sets for teaching specific subjects. None of these initiatives yet take advantage of all the diverse new features provided by modern spreadsheet editors. Based on the fact that teachers generally go through several levels of skill development when learning to integrate new technology into educational practice for their students [12], it might be anticipated that teachers would encounter multiple barriers when attempting to use the vast array of the latest spreadsheet features for educational goals.

To date, however, no field studies have yet assessed in detail whether and how people are using the many data visualization features introduced in the past decades, especially in the context of education. Our goal is to fill this gap, in order to help guide the development and refinement of spreadsheet editors.

[1] http://www.keypress.com/

3. Methodology

Studies of users outside the laboratory typically fall into two methodological categories, each with complementary strengths and weaknesses. Some studies focus on user experiences, often through surveys (e.g.: [2][9][18]) or direct observation (e.g., [8]). One weakness of surveys is that people might not accurately remember or portray past events. Other studies focus on digital artifacts created by end users (e.g., [4][15]), thereby obtaining insights from artifacts without the potentially-distorting biases introduced by interacting with or observing people. However, this method provides no information about what people think or want.

We have combined these complementary approaches by combining views based on artifacts as well as human experiences. The first part of our study consisted of analyzing over 400 spreadsheets. We then analyzed over 200 online forum threads of users' comments and questions about Excel. In addition, we interviewed 5 teachers to learn about their experiences with spreadsheet-based educational materials. Using these results in combination provided answers to the research questions outlined in Section 1.

3.1. Spreadsheet Analysis

To obtain a general view of spreadsheet usage, we randomly selected 400 of the 4498 spreadsheets in the EUSES Spreadsheet Corpus [4], which was chosen because many other researchers have relied on it for studying use of formulas. We did not want to impose a pre-ordained coding scheme on the data. Therefore, in keeping with grounded theoretical approaches [5], we gradually built up a set of codes for categorizing spreadsheets based on content, layout, and feature usage. (Spreadsheets in each category had a typical content, layout and usage of features.)

To verify the coding scheme's robustness and consistency, a second person reviewed it and the results to look for ambiguity in codes or potential errors in statistics. To further test the coding scheme, a third person independently coded a random subset of 25 spreadsheets. Agreement was 92% and Cohen's Kappa was 89%, indicating that the scheme was reasonably robust and reliably applied. Based on this evaluation, we then reviewed and finalized codes.

After analyzing the 400 randomly-selected spreadsheets for overall purpose, structure and feature usage, we then delved into the charting-specific features, which were central to the motivating questions that motivated this study. Specifically, we analyzed all 419 charts that we could programmatically identify in the EUSES corpus; they were located in a total of 97 spreadsheets. (Other corpus spreadsheets also can contain charts that cannot be identified automatically due to Excel API version incompatibility.) Coding of chart feature usage was tested as above, with an inter-reviewer agreement of 96% and a Kappa of 93%. We also analyzed charts to identify common user errors.

3.2. Forum Analysis

While the analysis above identified technical errors in spreadsheet charts, it could not reveal what problems people actually cared about—in particular, analyzing spreadsheets will never reveal problems that prevent people from creating spreadsheets in the first place.

To supplement the spreadsheet analysis, we obtained user comments and questions from a popular online forum for discussing Excel.[2] We randomly selected 200 threads from the "General" forum and 40 from the "Charting" subforum. Coding of user problems was tested as above, yielding an agreement of 92% and a Cohen's Kappa of 89% between reviewers.

3.3. EUSES User Interviews

The analyses described above gave us preliminary ideas about the answers to our research questions. To evaluate and clarify these ideas, as well as to focus more closely on the usefulness of Excel for education, we interviewed teachers with relevant experience. Specifically, we recruited participants by emailing the 37 people who had previously downloaded materials provided by the EUSES Consortium as a basis for teaching mathematics in K-12 classrooms [13]. The centerpiece of these materials is a set of sample spreadsheets with charts that walk students through how to solve various word problems. In our recruitment email, we offered a $20 Amazon gift card incentive.

Five former or current teachers agreed to a semi-structured 30-minute interview, which asked about experiences with these materials and with Excel in general, as well as what changes they want in Excel. Given the small sample, we did not code the responses but instead qualitatively used them to refine the results previously obtained through the spreadsheet and forum analysis.

4. Spreadsheet Analysis Results

In examining 400 randomly selected spreadsheets, we focused on their functional role, structure and feature usage. Based on these criteria, we identified five categories of spreadsheets (Table 2). After applying this coding scheme, we computed the distribution of spreadsheets among categories (Figure 1).

We found that *Data entry* spreadsheets were the most common, accounting for 56% of spreadsheets. Of our five categories, *Data entry* was closest to the stereotypical view of a spreadsheet's role as a calculation tool, yet even among these 221 spreadsheets abounding in numerical data, only 98 had any formulas.

[2] http://www.excelforum.com

Table 2. Codes for categorizing spreadsheets

Category	Description
Data entry	Spreadsheet has large amounts of numerical data that may or may not be tabular in structure. May have formulas, rarely charts, no form widgets.
Database	Spreadsheet consists mainly of textual tabular data. There is typically a heading for each column and a unique identifier for each row. No formulas, charts, or form widgets.
Data form	Spreadsheets contain check boxes or other form widgets, and often large amounts of explanatory text.
Data mix	Spreadsheet has a relatively balanced mix of numeric and text data. May have formulas. A subset of data may be graphed. No form widgets
Data viz	Spreadsheet consists predominantly of charts, along with numerical data that only exists in order to feed charts. Few formulas or text, no form widgets.

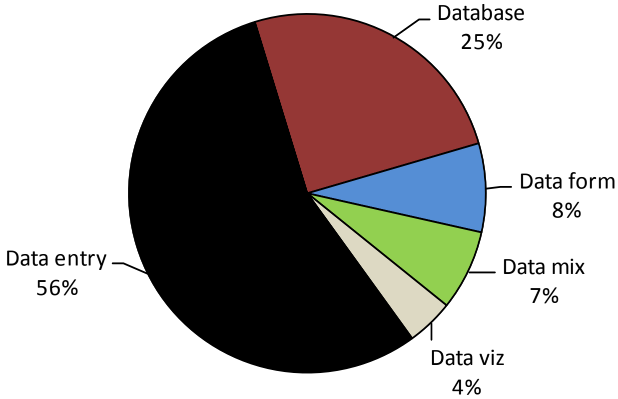

Figure 1. Distribution of spreadsheets by category

Nearly half of the spreadsheets fell into the other categories, of which *Database* spreadsheets rich with textual data were the most common at 25% of the total.

Data form spreadsheets used checkboxes or other form widgets in a manner similar to dialog windows in desktop applications or web forms in web applications. This category accounted for 8% of spreadsheets, suggesting that Excel is now finding some limited use as a platform for creating simple applications.

Data mix was the only category where spreadsheets combined significant proportions of numbers, text, formulas, and charts all in the same document. Such spreadsheets were rare, at only 7% of the total. Instead, (as reflected in the other categories), most spreadsheets each did mainly one thing such as storing numbers, storing text, presenting a form, or visualizing data.

Data viz was the least common kind of spreadsheet, demonstrated by only 17 of the 400. We also observed charts in 11 *Data entry* spreadsheets, with only 7% of spreadsheets overall containing any charts.

To provide further detail about how people use sub-features of Excel's chart editor, we programmatically scanned the entire EUSES spreadsheet corpus to obtain 97 spreadsheets that together contained 419 charts. We then analyzed what kinds of charts they were, what specific features they used, and what errors people frequently made with these features.

We found that most charts (76%) were either line or bar charts (Figure 2). The next most frequent were pie charts at 14%. The remaining 10% were scatter, stacked bar, or area charts (which are essentially stacked line charts).

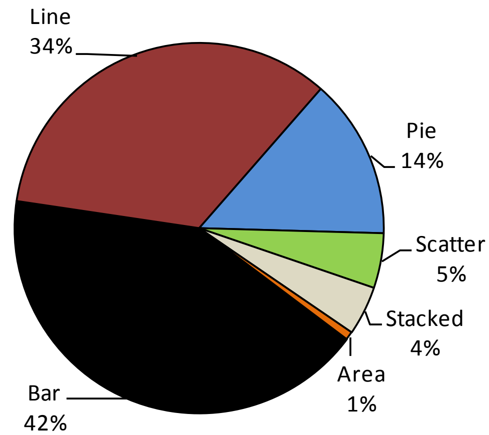

Figure 2. Distribution of charts by type

Just as certain types of charts were much more common than others, we found upon delving into the internals of charts that certain charting sub-features were much more frequently used than others (Table 1). In particular, 23% had labels on data points or bars, 5% contained trend lines, and 4% contained embedded tables showing data. Many sub-features such as series lines, drop lines, and error bars were very rare, appearing in a total of only 1% of charts.

We found that 16% of charts contained objects such as text boxes, lines and arrows that must be manually customized after insertion (unlike the automatically-generated objects such as labels and error bars, which require no manual customization). Manual objects allowed people to communicate information that could

Table 1. Frequency of use for sub-features that add objects to charts

Sub-feature object	Number	% of charts
Automatically-generated objects		
Data labels	96	23 %
Trend lines	19	5 %
Data tables	17	4 %
Other	5	1 %
Manually-inserted objects		
Text boxes	47	11 %
Lines	7	2 %
Arrows	6	1 %

not be expressed through automatically-generated objects. For example, some included textboxes stating facts that the reader should take away from the chart such as "Your project is in the 10th percentile, meaning that 90% of projects have a better APR difference". Manually-inserted objects were approximately as common in the oft-used line and bar charts (at 10% and 12%) as they were in less-used charts.

Finally, we analyzed errors demonstrated by charts. We categorized errors according to whether they were "critical" in the sense of making it impossible to correctly read the chart (e.g., broken cell references) or "non-critical" in simply making it difficult to read or interpret the chart (e.g., a barely readable font).

We found that 12% (51 charts) had critical errors. One common problem was broken data references in 3% of all charts, another was no data at all in 2% of all charts, and a third was that the charts were unreadable in 3% of all charts, often due to light foreground colors on a light background. These per-chart error rates are comparable to per-cell error rates found in prior studies of formula errors [15], meaning that each chart is just as prone to contain critical errors as a formula cell is.

We found that 43% (179 charts) had non-critical errors. By far, the most common problem was simply leaving out labels in situations that forced the viewer to refer to text near to the chart to interpret the chart (23%, 96 charts). The next most frequent problem was cluttering up the chart with unneeded information (7%, 28 charts), such as a legend whose content was redundant with labels attached to data points.

Overall, these results suggest that Excel's chart editor is somewhat difficult for people to use, particularly in terms of achieving readable charts. Moreover, effectively using it can at times require significant manual tweaking. In the next two sections, we evaluate and refine these inferences by studying users' experiences.

5. Forum Analysis Results

We analyzed 200 randomly-selected threads from a "General Excel" online forum[2] in order to learn what problems users complain about. We found seven categories of user problems (Table 3). The *Problem setup* category essentially covers difficulties before users even know what features are needed. For example, one user explained, "I would like to develop a spreadsheet to forecast the evolution of the corresponding accounts." The *Feature finding* category deals locating features in the user interface; for example, one user requested, "How can I keep the first line on the screen when I scroll down?" The last five codes related to actually using or configuring certain kinds of features. After applying this coding scheme, we found 21% of threads related to *Problem setup*, 19% related to *Feature finding*, and the remaining 60% related to actually using features (Figure 3). These statistics from the "General" forum serve as a baseline for comparing similar statistics from the "Charting" subforum, below.

Table 3. Codes for categorizing "General Excel" forum threads

Category	Description
Problem setup	Often the user will know what they want to do, but wants to know how to set the problem up
Feature finding	Problems with finding specific functionality through the user interface of Excel
Config	Problems dealing with use of Excel's settings, such as default colors or startup toolbars
Formulas	Problems dealing with use of formulas, such as asking about specific formulas or formula debugging.
Data viz	Problems dealing with use of advanced data visualization features, such as pivot tables or charting
Macro	Problems dealing with use of macros, including writing debugging and invoking
Integration	Problems dealing with use of features to integrate Excel with other programs, such as data import

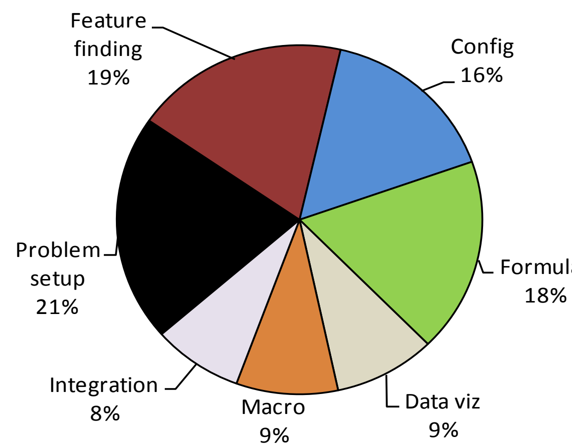

Figure 3. Distribution of general forum threads by user problem

To gain more insight into the chart-specific problems that people encounter, we examined 40 randomly-selected threads from the "Charting" subforum. (Of the seven categories shown in Table 3, only *Charting* and *Formulas* had their own subforums on the site.)

As we developed a coding scheme for the "Charting" subforum (Table 4), we recognized that many features sought by people could be described as searches for *Novel charts*: charts that Excel currently does not support directly, but which rather must be pieced together through a series of manual steps.

Table 4. Codes for categorizing "Charting" subforum threads

Category	Description
Problem setup	Users often present their data and ask for the best way to chart this or how to present it
Feature finding	Problems with finding specific functionality through the user interface of Excel
Novel chart	Creating charts that Excel does not directly support
Using	Questions about settings when using charts

Novel chart can be viewed as a composite variant of *Feature finding*. For example, one user asked, “How do I create a more than [sic] cumulative frequency polygon?” (A cumulative frequency polygon is essentially a cumulative histogram that uses a line chart rather than a bar chart to depict frequencies.) In response to this request, another user explained how to combine histogram features, formulas, and line chart features to construct the desired chart.

In contrast to complex *Novel chart* requests, *Feature finding* threads focused on locating particular individual features to perform specific operations. Some of these related to questions about how to modify specific types of charts, such as inserting a second x axis in a bar chart, formatting the slices of a pie chart based on specific conditions, or adding a regression line.

We found that *Problem setup* accounted for 17% of “Charting” threads, *Feature finding* or its *Novel chart* composite variant covered 70%, and *Using* accounted for the remaining 10% (Figure 4). Comparing these statistics to those of the “General” forum, the “Charting” threads are much more likely to focus on locating features than they are to focus on using features. That is, compared to Excel as a whole, the charting features seem particularly hard to find.

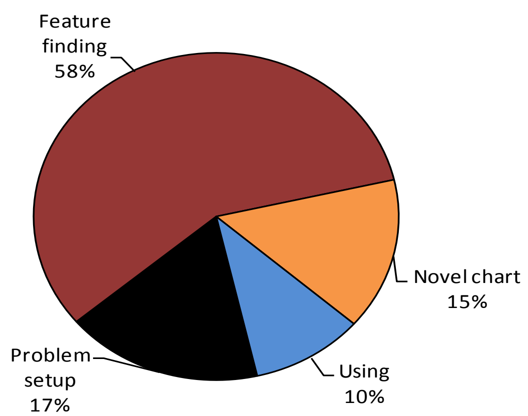

Figure 4. Distribution of Charting threads by user problem

6. Interview Results

The analysis in the previous sections tells us what problems people are having with Excel, but not why these problems matter. To complement this analysis and to show that these problems do have real consequences, we interviewed five people who downloaded the EUSES spreadsheet education materials. These materials consist of 27 sample spreadsheets for teaching various subjects in math, such as algebra, geometry, and statistics. Many of these spreadsheets include charts. We asked about respondents’ background, experiences with these materials, general views on Excel, suggestions for improving Excel and experiences using charts.

All five interviewees had experience teaching. Two currently taught K-12 math, one currently taught college math, and two were retired K-12 teachers. All had experience using spreadsheets and expressed a desire to use spreadsheets to support education. The two retired teachers and the college teacher actively try to develop spreadsheet materials for this purpose and serve as paid consultants for other teachers.

One challenge with using spreadsheets in the classroom was a difficulty in seeing how they fit in with specific educational goals. Specifically, both retired teachers indicated that they felt the spreadsheets needed to be more closely tailored to the particular topical needs of particular courses; one suggested that spreadsheets should be accompanied by customizable lesson plans.

Shifting attention away from the classroom to the editor in general, four interviews suggested that users struggle to find features in Excel. Two explicitly commented on trouble “figuring out where things are.” Two other interviewees appeared to reveal their own inability to find features when they suggested that Excel should provide certain functionality that Excel already does have (specifically, a wizard for helping people to write formulas, and support for box plots).

In regards to charting, Excel was found to be adequate by most of the teachers and able to create most of the charts they needed. However, two teachers found Excel to be lacking in several categories and are forced to use online graphing tools to remedy this. One stated “I use Excel for charting simple functions and pie charts, but often have to use online graphing tools for the more complicated charts.” Another teacher was required to demonstrate box plots, but could not get Excel to create the kind of charts she desired and also used an online tool.

Finally, interviewees thought that teachers are reluctant to use Excel in class unless if they feel a total mastery of Excel. One reason is that students might want help that the teacher cannot give; one teacher coped by just demonstrating Excel, rather than letting students use it. But even just using features to lecture can be intimidating, so one teacher wanted “a version of Excel that had less or no formulas so that I wouldn’t have to worry about messing up.” Another thought a simplified system like a graphing calculator might be ideal and several teachers and said that a simple data visualization tool would be beneficial as they would not have to teach Excel to their students or master it themselves to teach data visualization or create charts. While we had been aware that teachers have varying levels of skill at using new technologies in education [12], we were surprised that interviewees so consistently and repetitively said that Excel is just too intimidating to teachers.

7. Discussion and research opportunities

Our study's three parts provide complementary views that together answer our initial questions.

Q1. What features of spreadsheet editors are people actually using in practice? Our spreadsheet analysis indicates that people largely use spreadsheets for storing numerical or textual information, often in a tabular form similar to a database. Only the simplest features in Excel are needed for organizing and editing such data. Further analysis indicated that few spreadsheets have charts, and almost all charts are simple line, bar or pie charts. Within these charts, the automatically-generated data labels are one of the most frequently-used sub-features, with all other kinds of objects being much less frequently inserted by users.

This, of course, is not to say that spreadsheets are always simple, that only text editing features are needed, that the basic charts will always suffice, or that labels are the only chart sub-feature of value. Indeed, a huge variety of different features and sub-features are reflected in users' spreadsheets, but at less frequent rates of use.

Moreover, there seems to be an infrequent but steady need for features for manually tweaking or tailoring spreadsheets at many levels. Charts sometimes need to be customized through manual insertion of objects. Users request exotic charts that can only be constructed by discovering and tuning multiple features. Teachers call for entire spreadsheets that are tailored for obtaining targeted education outcomes in specific topics and grade levels.

Q2. What problems do people encounter when using features? The primary problem seems to be wading through the complexity of Excel to find needed features. In prior work aimed at studying learning barriers of novice Visual Basic programmers, this sort of problem was labeled a "selection barrier" [8]. In that study, only 10% of barriers fell into this category. In ours, most "Charting" subforum threads related to locating features, as did nearly 20% of all "General Excel" threads. These statistics help to explain why so many teachers feel intimidated by Excel.

Another difficulty observed in prior work was "design barriers" related to decomposition of a problem, amounting to only 3% of programmer problems in that prior work [8]. This category is analogous to our *Problem setup* code, which accounted for approximately 20% of "Charting" and "General" forum threads. Since these problems are "inherent cognitive difficulties of a programming problem, separate from the notation used to represent a solution" [8], they probably do not reflect a limitation in Excel except insofar as the editor does not attempt to provide support for walking people through the process of problem decomposition.

The remaining difficulties relate to understanding, using, coordinating, and evaluating features [8]. Our chart-specific results do not appear to differ much in these areas from the prior study of Visual Basic programmers. For example, coordination barriers accounted for 19% of difficulties in that work, while *Novel charts* accounted for 15% of our "Charting" subforum threads (though, of course, these two statistics are not precisely comparable since *Novel charts* also required discovering several features, which is a composite form of a feature selection barrier). We also observed that 12% of charts were unreadable due to critical errors in usage, and multiple interviewees expressed fears of making mistakes when using features.

Q3. How could spreadsheet editors be improved through addition, enhancement or elimination of features? The results above suggest several avenues for refining spreadsheet editors to better assist users.

Temporarily hiding features

The Pareto principle is alive and well in Excel: most spreadsheets could be created by invoking relatively few features. This suggests that most features could be temporarily hidden except in user-controlled modes when they are needed. Excel already does this to a very limited extent; for example, features in the data analysis "pack" are turned off by default. However, it is impossible to hide most features (including most features that we rarely saw in use). For example, there is no way to turn off formula editing (which might be desirable when creating a *Database* spreadsheet) or form widgets (which might be desirable when creating anything other than a *Data form* spreadsheet).

If teachers could turn off all features except for those needed in a particular lesson, then they might not have such a strong need to master Excel in its entirety before using it in class. As students (and teachers) gain proficiency, features could be turned on until students experience an editor essentially identical to what would be encountered after graduation. This is not possible with "stripped down" tools like the Fathom editor[1] that simply omit features entirely.

Improving support for spreadsheet reuse

Interviewees called for spreadsheets matched to course- and grade-specific educational outcomes. Given the limitations of researchers' time and domain knowledge, there is no way that we can provide spreadsheets tailored to all teachers everywhere. Only teachers themselves have the manpower and the domain knowledge needed to provide such a broad range of spreadsheets. Yet given the range of skills levels possessed by teachers, it is also clear that not all teachers are ready to create the needed materials for themselves.

One approach for resolving this challenge might be to help teachers across one or more states to share, find and reuse one another's spreadsheets. For example, a skilled teacher could create a spreadsheet suitable for teaching 9th graders what it means to solve a quadratic equation, and then publish it to a repository where other teachers could download and use it. These other teachers would need some method for finding relevant materials when they want to teach a certain concept, and it is questionable whether a standard keyword-based search engine would provide the needed level of search specificity and effectiveness. We are aware that there has been work in the area of spreadsheet repositories for over a decade (e.g., [20]), but we are not aware of any system specialized to the needs of teachers. Developing an approach that works well for educators might also inspire additional enhancements for systems oriented towards other specialized user populations.

Helping users to overcome learning barriers

Posting a request for help in an online forum is a slow and unreliable way to find features, since obtaining an answer may take days or weeks, if it ever comes at all. Users need a much more effective way of matching goals to specific spreadsheet features, including those for charting. Moreover, given the frequency of *Novel chart* problems in the "Charting" subforum, users need help with mapping goals to *multiple* features and using them together correctly. These barriers might be overcome by adapting techniques that have been explored elsewhere (see [8] for a discussion of approaches). However, achieving this will be particularly hard if some features are actually hidden from users as we suggest above. In that case, novel interaction techniques might be needed to help people find, briefly unhide, and try out features, then return the editor to a "safe" mode if features do not work out as intended.

Finally, after discovering and using features, people need ways to evaluate the results, then discover and correct errors. As we have seen, these errors can be obvious and might cause charts to display no data at all, or they can be subtle usability problems such as light grey text in a small font on a white background. Detecting such errors, and helping people to correct them, will likely require new features. We hope, of course, that these features can be designed in a way that does not further perplex and intimidate the user.

9. Acknowledgements

We thank Maggie Niess for feedback on our interview design and on early drafts of this paper, as well as Ronald Metoyer for interview design suggestions.

10. References

[1] J. Caulkins, E. Morrison, and T. Weidemann. Spreadsheet errors and decision making: Evidence from field interviews. *J. Organizational and End User Computing, 19*, 3, 2007, 1-23.

[2] Y. Chan and V. Storey. The use of spreadsheets in organizations: Determinants and consequences. *Information & Management, 31*, 3, 1996, 119-134.

[3] P. Cragg and M. King. Spreadsheet modelling abuse: An opportunity for OR? *J. Operational Research Society, 44*, 8, 1993, 743-752.

[4] M. Fisher and G. Rothermel. The EUSES Spreadsheet Corpus: A shared resource for supporting experimentation with spreadsheet dependability mechanisms. *Workshop on End-User Software Eng.*, 2005, 47-51.

[5] B. Glaser and A. Strauss. *The Discovery of Grounded Theory: Strategies for Qualitative Research,* Aldine Publishers, 1977.

[6] M. Hall. A risk and control-oriented study of the practices of spreadsheet application developers. *Hawaii Intl. Conf. Sys. Sciences*, 1996, 364-373.

[7] D. Ibrahim. Teaching Science and Mathematics Subjects Using the Excel Spreadsheet Package, *Symp. Frontiers in Science Education Research*, 2009.

[8] A. Ko, B. Myers, and H. Aung. Six learning barriers in end-user programming systems. *Symp. Visual Lang. and Human-Centric Computing*, 2004, 199-206.

[9] B. Lawson, et al. A comparison of spreadsheet users with different levels of experience. *Omega, 37*, 3, 2009, 579-590.

[10] Mathematical Sciences Education Board, National Research Council. *High School Mathematics at Work: Essays and Examples for the Education of All Students*, National Academies Press, 1998.

[11] National Council of Teachers of Mathematics, *Data Analysis and Probability Standard: Principles and Standards for School Mathematics*, 2004.

[12] M. Niess, et al. Mathematics teacher TPACK standards and development model. *Contemporary Issues in Technology and Teacher Education, 9*, 1, 2009, 4-24.

[13] M. Niess. Scaffolding math learning with spreadsheets. *Learning & Leading with Technology, 32*, 5, 2005, 24-27.

[14] M. Oliveira and S. Nápoles. Using a spreadsheet to study the oscillatory movement of a mass-spring system, *J. Spreadsheets in Education, 3*, 2, 2010.

[15] R. Panko. What we know about spreadsheet errors. *J. End User Computing, 10*, 2, 1998, 15-21.

[16] S. Powell, K. Baker, and B. Lawson. Errors in operational spreadsheets. *J. Organizational and End User Computing, 21*, 3, 2009, 24-36.

[17] K. Rajalingham, D. Chadwick, and B. Knight. Classification of spreadsheet errors. *Symp. European Spreadsheet Risks Interest Group*, 2001.

[18] C. Scaffidi, et al. Dimensions characterizing programming feature usage by information workers. *Symp. Visual Lang. and Human-Centric Computing*, 2006, 59-62.

[19] C. Scaffidi, M. Shaw, and B. Myers. Estimating the Numbers of End Users and End User Programmers. *Symp. Visual Lang. and Human-Centric Computing*, 2005, 207-214.

[20] R. Walpole and M. Burnett. Supporting reuse of evolving visual code. *Symp. Visual Lang. and Human-Centric Computing*, 1997, 68-75.