

Excel as a Qualitative Data Analysis Tool

DANIEL Z. MEYER

Illinois Institute of Technology

LEANNE M. AVERY

State University of New York College at Oneonta

Qualitative research seeks to examine the interconnections in rich, complex data sources. The statistical tools of quantitative methods separate out pieces of data in a manner that defeats the purpose. But, like quantitative researchers, qualitative researchers often still find themselves overwhelmed by the amount of data and equally in need of tools to extend their human senses. This has led the development of a number of software packages designed for this purpose. An often overlooked option, however, is Microsoft Excel. Excel is generally considered a number cruncher. However, its structure and data manipulation and display features can be utilized for qualitative analysis. In this article, the authors discuss data preparation, analysis, and presentation, including discussion of lesser known features of Excel.

Keywords: *qualitative methods; data analysis; constant comparative method*

Researchers using qualitative data often find themselves lost in a sea of data. Although it is the very richness and interconnectiveness that we find appealing, the data can also be overwhelming. Some sense must be made of them while preserving their complexity. We therefore conceptualize tracking as a central hurdle in qualitative data analysis: We often need to be able to connect one bit of qualitative data to another bit. This need to track has resulted in a significant market for qualitative data analysis software tools that utilize the power of modern computing to augment our own human senses.

Excel is often viewed as a number cruncher and is therefore associated with quantitative data analysis, but we have also found it useful as a qualitative tool. It can handle large amounts of data, provide multiple attributes, and allow for a variety of display techniques. In this article, we demonstrate the use of Excel as a qualitative data analysis tool. We will cover data preparation, analysis, and presentation, paying particular attention to less

commonly used features of Excel such as conditional formatting. We end with a discussion of the nature of Excel's technical structure.

DIFFERENCE BETWEEN QUANTITATIVE AND QUALITATIVE NEEDS

A brief (and oversimplified) conceptual distinction between quantitative and qualitative data analysis will be useful in highlighting what a qualitative data analysis tool should provide. Quantitative researchers are faced with an overwhelming amount of data—too much to see the patterns with unaided human senses. So their approach is to filter out the noise and synthesize the relevant data into something that can be interpreted by a human. Calculating a mean is a simple example of this: It is hard to make any sense of a raw list of 1,000 test scores, but a mean gives a sense of the population. Qualitative researchers are in a similar position of being overwhelmed by the data. However, they are often interested in precisely the connections and nuances that are frequently lost when filtering and synthesizing. Qualitative research, therefore, is better seen as a tracking problem. Researchers need some way of saying that this event over here has some relationship to that event over there. Like quantitative researchers, our ability to do that unaided is quite limited. This has led to interest in computer software as a way to facilitate qualitative analysis.

A CAVEAT AND ASSUMPTION

There is an important caveat that we should state before proceeding. All research projects (and *researchers*) are not the same. What works for one project may not be best for another. Furthermore, beyond the technical ability to perform a given operation, the *way* in which a system implements that operation may be the deciding factor in choosing one over another. The capabilities of a particular tool and the usefulness of those capabilities for a particular project are two separate issues. This article is about the capabilities of Excel. Not every technique shown here will be useful for every project. We also do not intend this article to be a tutorial—there are better options for direct training. Rather, our goal is to engender a conception of Excel that includes handling nonnumerical data.

Furthermore, because we cannot attend to every possible qualitative data circumstance, we will make some assumptions about the nature of the data.

We will focus on data that consist of some sort of transcript. We have used Excel as a database for tracking mixed data sources. However, as this represents a more straightforward use of Excel and transcription analysis represents such a significant occurrence in qualitative data analysis, the focus on transcript data is more productive.

END GOAL

To provide some orientation, we begin with a glimpse of something closer to the end. Figure 1 shows a portion of a coded transcript. Column E contains the actual talk—the heart of our data.¹ To the left are identification codes and to the right are analytical codes. We will discuss how and why this format was constructed and what can be done once the data are in this form.

PREPARING DATA

We begin with a discussion of preparing a transcript, in part because it involves analytical decisions that will have consequences down the road. Although we hope to provide a convincing case for Excel as a data analysis tool, it is *not* an effective tool for transcription. In particular, quickly moving the cursor around a transcript—something important to efficient transcribing—is easier in a text editor than Excel. Using the arrow keys, for example, would actually select the entire cell, resulting in the entire cell contents being replaced when new text is typed. These may seem like minute differences, but in the context of transcription, every small effort to save time or energy can help.

When creating the transcript in another application,² it is important to have a sense of how Excel will be importing the file. The file will be important as a tab-delimited file.³ This means that as Excel reads the file, tabs will indicate that it should move to the next horizontal cell, and carriage returns will indicate it should move to the beginning of the next row. Note that this means that blank cells need to be accounted for by inserting additional tabs.

Figures 2–4⁴ show the same transcribed text but with three different notations used in the text editor that result in three different configurations in Excel. Figures 2 and 3 are two legitimate alternatives for what will become the codable unit (which we will address next), while Figure 4 illustrates a mistake. In Figure 2, the intention is to make the entire turn of each speaker one row in Excel. Therefore, for each turn, there is a label for the speaker, followed by a

FIGURE I
Coded Transcript

	A	B	C	D	E	H	I	J	K	L	N	O	P	Q	R	S	T	U	V
	Group	Date	File	Speaker	Talk	speechg content1	speechg content2	speechg content3	speechg content4	artifa ctacti on									
1	sewage	46 0074	Darrin		Well it seems like having kids play with chlorine, and microbes are not gonna really, isn't gonna really be realistic.	content activ													
81																			
82	sewage	46 0075	Megan		mm														
83	sewage	46 0076	Darrin		So we may have to have them kind of, design only on paper.	content activ													
84	sewage	46 0077	Darrin		Or in a system like, Stella for example, or	content activ													
85	sewage	46 0078	Megan		mmm, yeah														
86	sewage	46 0079	Darrin		Like Stella might be a really good way to, model this.	activ													
87	sewage	46 0080	Darrin		that adds a whole nother completely serious element to it.	content													
88	sewage	46 0081	Megan		yeah														
89	sewage	46 0082	Darrin		we'll stick with the,														
90	sewage	46 0083	Darrin		the basics first.														
91	sewage	46 0084	Kevin		So, are we gonna I guess, thinking that we want to do a design challenge	goal													
92	sewage	46 0085	Megan		Can I see the-wastewater treatment thing?	examnat													
93	sewage	46 0086	Darrin		What's that?	***													
94	sewage	46 0087	Kevin		Do we want to do a design challenge?	goal activ													
95	sewage	46 0088	Darrin		I think the cool part about a design challenge, is that you can give them, um,	content activ													
96	sewage	46 0089	Darrin		what you can give them as parameters, essentially.	content concep pedog t													

FIGURE 2
Turn as Unit

D	→ I mean all this stuff has to get filtered out. So, these are the kinds- we can do a day where we introduce- what's there, ¶
M	→ Kinda setting it up for what their gonna have to actually- ¶
D	→ right ¶
M	→ filter out- like, take care of when they build the thing. ¶
D	→ mmm, and then we can do issues like, um, you know, how can we- And then we can just let them essentially take it, and just grapple with these ideas for a day or two and try to figure out- how their gonna address these pieces. You can't take it all- can't do it all at once. ¶
D	I mean all this stuff has to get filtered out. So, these are the kinds- we can do a day where we introduce- what's there,
M	Kinda setting it up for what their gonna have to actually-
D	right
M	filter out- like, take care of when they build the thing.
D	mmm, and then we can do issues like, um, you know, how can we- And then we can just let them essentially take it, and just grapple with these ideas for a day or two and try to figure out- how their gonna address these pieces. You can't take it all- can't do it all at once.

tab (indicated by →), followed by the speech, and finally a carriage return (indicated by ¶). Note that there is a difference between the text wrapping to another line and a carriage return making another line. The former is meaningless for Excel. So in Figure 2, there is only a carriage return between the change in speaker, producing a change in the Excel row at that point. In Figure 3, the intention is for each sentence⁵ to be one row in Excel. Therefore, for the first sentence of a particular speaker, there is the label for the speaker, followed by a tab, followed by the first sentence, followed by a carriage return. For each additional sentence until a new speaker, there is a tab, followed by the sentence, followed by a carriage return. The tab is there so that the additional sentences will be in the same column as the first.⁶ In Figure 4, we see that those tabs are missing, leading to a format in Excel we do not want.

The difference between Figure 2 and Figure 3 illustrates a crucial point to consider at this stage (or earlier!). What they represent is a difference in the choice of codable unit—in this case, turn versus sentence. *What should be used is an analytical decision.* So what seems like only a technical chore of preparing the data for Excel can actually have profound implications for your study. You must decide what unit has the most meaning for your particular research. You may also want to consider what approach preserves the most flexibility. By choosing the sentence as the unit, there are some ways in which the turn can be reconstructed at a later time (which we will illustrate as a later example).

Finally, it is helpful to have a good sense of how a text editor's search and replace features work. We will give two examples. In the examples

FIGURE 3
Sentence as Unit

D	→ I mean all this stuff has to get filtered out.
	→ So, these are the kinds-
	→ we can do a day where we introduce- what's there,
M	→ Kinda setting it up for what their gonna have to actually-
D	→ right
M	→ filter out- like, take care of when they build the thing.
D	→ mmm, and then we can do issues like, um, you know, how can we-
	→ And then we can just let them essentially take it, and just grapple with these ideas for a day or two and try to figure out- how their gonna address these pieces.
D	→ You can't take it all- can't do it all at once.
D	I mean all this stuff has to get filtered out.
	So, these are the kinds-
	we can do a day where we introduce- what's there,
M	Kinda setting it up for what their gonna have to actually-
D	right
M	filter out- like, take care of when they build the thing.
D	mmm, and then we can do issues like, um, you know, how can we-
	And then we can just let them essentially take it, and just grapple with these ideas for a day or two and try to figure out- how their gonna address these pieces.
	You can't take it all- can't do it all at once.

above, note that we only used an initial for each speaker’s name. That can now be replaced with the full name. However, if you simply ask the computer to replace “D” with “Darrin,” it will do just that—including all the places in the speech!⁷ But the fact that each speaker label “D” has a tab next to it can be used to uniquely find the ones you want to change: Search and Replace “D<tab>” to “Darrin<tab>.”

Search and Replace is also often helpful when working with a preexisting transcript. A different demarcation—such as a colon and space—may have been used to separate the speaker from the speech. Also, double carriage returns are often used between speech units. These can be replaced as a batch to prepare for importing to Excel. What these two examples show is that using Search and Replace requires being very careful about distinguishing uniquely what you want to change from what you do not want to change.

IDENTIFICATION CODES

Now the tab-delineated transcript can be imported into Excel. Each row will define a codable unit, and each column will define an attribute of that unit

FIGURE 4
Sentence as Unit with Mistakes

D → I mean all this stuff has to get filtered out. ¶
 So, these are the kinds- ¶
 we can do a day where we introduce- what's there, ¶
 M → Kinda setting it up for what their gonna have to actually- ¶
 D → right ¶
 M → filter out- like, take care of when they build the thing. ¶
 D → mmm, and then we can do issues like, um, you know, how can we- ¶
 And then we can just let them essentially take it, and just grapple with these ideas for a
 day or two and try to figure out- how their gonna address these pieces. ¶
 You can't take it all- can't do it all at once. ¶

D	I mean all this stuff has to get filtered out.
So, these are the kinds-	
we can do a day where we introduc e- what's there,	
M	Kinda setting it up for what their gonna have to actually-
D	right
M	filter out- like, take care of when they build the thing.
D	mmm, and then we can do issues like, um, you know, how can we-

(including the original data itself), as can be seen in Figure 5. In comparison with a database file, each row can be seen as equivalent to a record and each column as a field. At this point, there will probably be nothing more than the speaker label and the speech. There will likely be some simple identification codes where entire sections have the same code. These can be added in large batches using the Fill→Down command. This is also the way blank speaker codes can be filled in. Line numbers can be added using the Fill→Series command. Figure 5 shows a portion of transcript set up with additional identification attributes.

Notice that the identification codes include “group” (there were two groups in this study) and “date” (each group met ten to twelve times). What is included in the same Excel Worksheet,⁸ versus different worksheets in the same Excel Workbook versus a different workbook altogether, is another analytical decision for you to make (although changing this at a later date is easier than the unit question). An argument for including more on the same sheet would be that

FIGURE 5
Sample Transcript with Identification Attributes

◇	A	B	C	D	E
1	Group	Date	Line	Speaker	Talk
100	sewage	4/5/05	0093	Darrin	I mean all this stuff has to get filtered out.
101	sewage	4/5/05	0094	Darrin	So, these are the kinds-
102	sewage	4/5/05	0095	Darrin	we can do a day where we introduce- what's there,
103	sewage	4/5/05	0096	Megan	Kinda setting it up for what their gonna have to actually-
104	sewage	4/5/05	0097	Darrin	right
105	sewage	4/5/05	0098	Megan	filter out- like, take care of when they build the thing.
106	sewage	4/5/05	0099	Darrin	mmm, and then we can do issues like, um, you know, how can we-
107	sewage	4/5/05	0100	Darrin	And then we can just let them essentially take it, and just grapple with these ideas for a day or two and try to figure out- how their gonna address these pieces.
108	sewage	4/5/05	0101	Darrin	You can't take it all- can't do it all at once.
109	sewage	4/5/05	0102	Darrin	You can't- filter and clean, and all these things in the same ?? (cooking) pot.
110	sewage	4/5/05	0103	Kevin	mm
111	sewage	4/5/05	0104	Darrin	You have to do it in stages, so that's the, that's the challenge.

less repetition of work is needed when identical analysis is performed. An argument against it would be that the computer would be asked to do more at once. In this case, each group had its own workbook, and each date was arranged linearly in chronological order. It might seem, therefore, that including the column for group was superfluous, as every line in this file would have the same group attribute. However, it is advisable to include the attribute with each line. That way, if lines of data are taken out of context—perhaps shown with data from other sets—the information will remain with each line.

ANALYTICAL CODING

With the basic transcript in place, analytical coding can now proceed. Additional columns can be used to house whatever coding is desired.

However, it is important to note that there are a couple of options as to how exactly to do this. Figure 6 shows a variety of coding structures in columns H–V. We will discuss each set in turn.

Single Value Code

Column H contains a code called *speechaction*. This was meant to describe the type of comment, as opposed to the content of the comment. So, for example, line 0115 has been coded as a positive judgment (judgpe), lines 0116 and 0117 have been coded as suggestions (suggest), line 0118 has been coded as a negative judgment (judgen), and 0121 has been coded as meta-statement about the design process (metacurdes). *Speechaction* is single value code—that is, each utterance can only be coded with one value. Therefore, this code is represented by only one column.

Multiple Value Code

Columns I, J, and K contain codes for *speechcontent*. This was used to describe the content of the comment. In contrast to the *speechaction* code, this code is not exclusive. That is, a comment can be coded with multiple values. This is why three columns were used.⁹ For example, line 0116 has been coded for curriculum structure (struct), content (content) and activity (activ). Although multiple codes could have been listed in a single cell, using multiple columns, where each cell would have a single unambiguous code, is preferable for subsequent operations.

Columns O–S also code for a multiple value code. This code—*issues*—was used to track the various topics that arose over the course of the group's discussions. Because the list of topics was long and the descriptions did not lend themselves well to abbreviations, numbers were used in contrast to the previous codes. For example, lines 0117 and 0118 have been coded for issue 33, which refers to specific details in the nature of the design challenge.

Two Variables

Columns T–W are used for a coding scheme that incorporates two variables or dimensions. A theme that emerged was how group members could judge an idea through different lenses. Specifically, they may judge it for the logistics, pedagogy, content, or concepts (which were then referred to collectively as *curricular contexts*). So these four columns were used to indicate one dimension—whether the utterance reflected one or more of those perspectives. But, in addition, we wanted to indicate a positive, negative, or neutral position with respect to those perspectives. By having a column for

FIGURE 6
Coding Structures

C	D	E	H	I	J	K	O	P	Q	R	S	T	U	V	W
Line	Speaker	Talk	speechaction	speech content1	speech content2	speech content3	issue1	issue2	issue3	issue4	issue5	logistic	Pedog	content	concept
0115	Darlin	Uh, it would be great.	judgpe												
0116	Darlin	I mean I feel like if we could model things some how, in a way that was, consistent.	suggest	struct	content	activ						x		x	
0117	Darlin	For example, what we could-write things or, or uh, score things on a scale of, how cheap it was, how stable it is, uh, how effective it is.	suggest	content	concep		03							x	x
0118	Darlin	But again those things are gonna be hard to determine without actually creating, ??	judgen	activ			11	03						n	n
0119	Darlin	???													
0120	Kevin	Yeah, that's what-	empty				11								
0121	Kevin	I guess I just thought that this was like, you come up with the guidelines, and then, you give it to them, and then they bring it, and then come here for a day, and then, like run their model, and then, they would do that.	metacurdes	logistic	activ		11					x		x	

each perspective and then entering a p, n, or x to indicate the second dimension, all of this information was included in a relatively compact form.

One could consider combinations of one dimensional codes to form multidimensional codes. The *speechaction* and *speechcontent* codes together describe the nature of the utterance (and in a very early form were initially one code). What distinguishes a coding scheme such as the curricular context coding is that both dimensions are represented by the same set of cells. Why did we not use a similar schema to show the *speechaction/speechcontent* combination? The reason to choose one versus the other has to do with the number of possible values. The context portion of the curricular context code could only have four different values and the position portion only three. This means that it was done with four columns and three possible values to put in those columns. In contrast, *speechaction* and *speechcontent* have far more possible values, requiring far more columns. It is also worth noting that in the curricular context coding, context is represented by the columns and position by the value within those columns—rather than the reverse—because the context portion is not exclusive. We would have had to have as many columns as *speechcontent* codes (more than twenty) and as many values to enter in the columns as *speechaction* (more than thirty).

Logical Formula Coding

We are used to thinking of Excel's computation capabilities as being numerical in nature. However, its formula functions are really logical and can operate on nonnumerical data as well. Functions such as *IF/THEN*, *LOOKUP*, and *CONCATENATE* can all be utilized with text. We will show two examples.

Recall the discussion above about the choice of turn versus sentence as the codable unit. Suppose, having used the smaller choice of sentence, we also want to flag where new turns occur. Figure 7 shows this being done. The formula in the right column achieves this. If the speaker label for the current line is equal to the previous line (meaning this is not a new turn), the value of the cell will be 0. Otherwise, it will be 1.

Figure 8 shows an instance where we want to replicate data from the original data worksheet in another location. In this case, we wanted to compare and contrast incidents of a certain type of event. Had they only consisted of a single utterance, the filter features might be used to highlight the events. However, the events we wanted to look at were responses to other speech, meaning the events contained a starting utterance, ending utterance, and a varying number of utterances in between. We wanted to be able to list both the beginning and ending utterances as

FIGURE 7
Logical Formula Coding

0103	Kevin	mm		1
0104	Darrin	You have to do it in stages, so that't the, that's the challenge.	=IF(D111=D110,0,1)	
0105	Kevin	But isn't it like when you do the design challenge,		1
0106	Kevin	I don't know if I understand something.		0
0107	Kevin	Do you like send this out to like schools, and then like,		0
0108	Darrin	Well you can compare it, that's the beauty of this- well that's-		1
0109	Darrin	One of the, the great things about this Environmental Inquiry thing is that- there's a system that's there to, to compare, to kind of compete.		0
0110	Darrin	But, um, what that system has been in the past, has been al- has been the peer review process, but also has been these like conferences where they come here, and show that		0
0111	Kevin	What I'm saying like- say we design this design		1

attributes of a single codable unit. Thus, this represented a significant change in the unit of analysis from what has been discussed so far.

The LOOKUP function could be used to automate much of this process. The LOOKUP function operates as a directory: Given a value to look up, it will return a match in a list of ordered pairs (such as two columns). Figure 9 shows a representation of how the function works: The lookup value is found in the lookup vector, and the corresponding value is returned from the result vector.

In Figure 8, what is seen in columns G and H, respectively, is automatically replicated using the LOOKUP function. In this case, the lookup criterion was a unique identification code made by combining the date and line number and listed in columns E and F, respectively. The formula shown in the last cell in column H reads LOOKUP (F6,data!\$C:\$C,data!\$G:\$G). It is saying, "Take the value in F6; look for the same value in column C on the 'data' Worksheet; take the row you find that value in and look at column G; put the value of that cell in this cell." From a user point of view, the only information that needed to be manually entered were the responded to line number, coded line number, and date for each incident.

FIGURE 8
Using the LOOKUP Function

A	B	C	D	E	F	G	H
Date	Report Code	Code	RespID	CodeId	Reponse to Talk	Coded Talk	
4/4/01	96	93	35523-0096	35523-0093	I don't know if you guys are aware but there's a big- big debate of this in eastern New York state about a PCB plant.	I don't know if this is feasible.	
4/4/01	140	147	35523-0140	35523-0147	They have like fish tank cleaners, which are just siphons.	they may not know, what they are?	
4/4/01	169	170	35523-0169	35523-0170	One more thing as far as the Hudson river example	which we probably won't follow because, we won't be [??] in New York state,	
4/4/01	191	193	35523-0191	35523-0193	And you are saying bioaccumulation. Bioaccumulation is a result of pollution, not really a how it get's polluted.	it's not very urban	
4/4/01	229	229	35523-0229	35523-0229			=LOOKUP(F6,data!\$C:\$C,data!\$G:\$G)

FIGURE 9
The LOOKUP Function

LOOKUP(lookup_value,lookup_vector,result_vector)

<u>Name list</u>	<u>Number list</u>
Bob	555-2718
Jane	555-3141
Joe	555-1059
Sue	555-6174

LOOKUP(Jane,Name list,Number list) → 555-3141

<u>Name list</u>	<u>Number list</u>
Bob	555-2718
Jane	555-3141
Joe	555-1059
Sue	555-6174

TRACKING

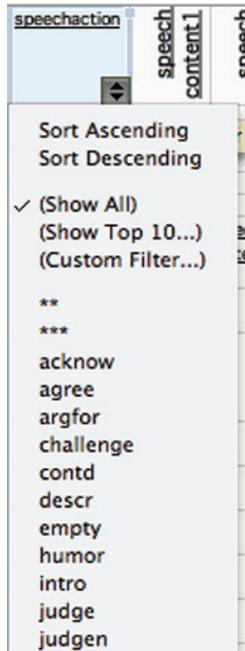
With coding in place, the tracking inherent in qualitative research can occur. Excel contains a number of different tools for aiding this process. The simplest would be the Find feature. Searches can be limited to a particular column or involve wildcards.

A more sophisticated tool is the filter. Excel has two filter modes. The AutoFilter can be applied to a particular column. This will create a pop-up menu that will filter rows based on the values in that particular column. The filter values that the pop-up menu will offer are the ones found in that column. Figure 10 shows the AutoFilter pop-up menu applied to the speechaction column. Figure 11 shows some of the rows resulting from filtering with the value “acknow” (acknowledgment).

Note, however, that the AutoFilter tool can only filter based on a single value. For more complex filtering, Advanced Filter is needed.

There are two simple but beneficial display options that deserve mention. Columns (or rows, for that matter) can be hidden. In addition, the Excel window can be split into two panes (actually two or four panes, because the split can be made vertically, horizontally, or both). This allows for rows or columns that are not near one another to be placed side by side. This can be seen in Figure 5 and others, where a pane has been created to show the header row (row 1) at the top no matter how far down one goes.

FIGURE 10
AutoFilter Pop-up Menu



CONTEXTUAL FORMATTING

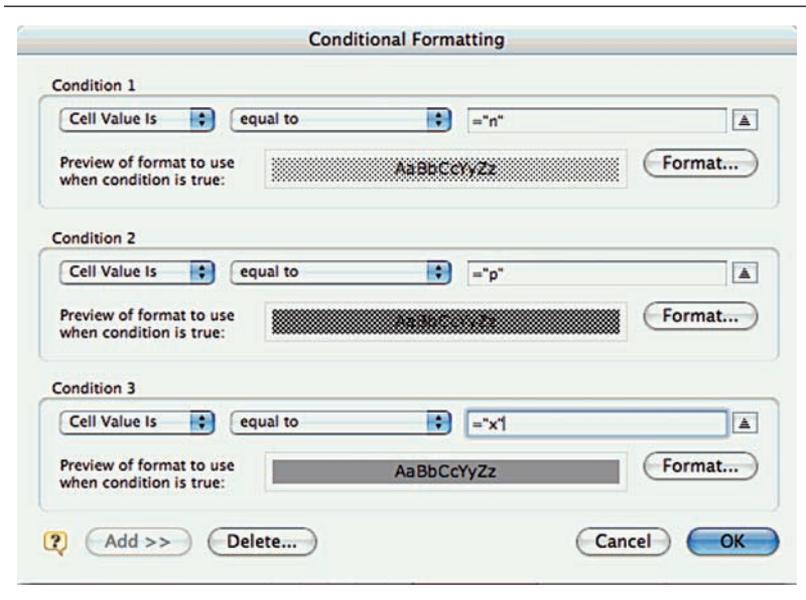
We now come to what is perhaps the most useful and least known feature in Excel. The formatting of a cell can, of course, be changed. Font, background color, border, and so forth, can all be changed at will. But they can also be changed automatically through a feature called Conditional Formatting.

Conditional Formatting allows the format of a cell to be determined by logical criteria. Figure 12 shows the dialog box to establish Conditional Formatting. The example shown tells Excel to shade the background of the cell blue if the value of the cell is “x,” green if the value is “p,” and red if the value is “n.” As can be seen in Figure 13, this has been applied to the curricular context coding described above.¹⁰ Conditional Formatting has also been applied to the issues coding to highlight two particular issues—numbers 11 and 33.

FIGURE 11
Filtered Rows

Line	Speaker	Talk	Note	speechaction
0003	Darrin	I can see that.		1 acknow
0004	Megan	Wow.		1 acknow
0020	Darrin	mhm [affirmative]		1 acknow
0022	Unknown F	yeah		1 acknow
0036	Darrin	Right		1 acknow
0045	Darrin	right		1 acknow

FIGURE 12
Conditional Formatting Dialog Box



PUTTING IT ALL TOGETHER

We will end with two examples that combine several of the techniques we have shown.

Although the transcript format shown so far, with speaker names in one column and speech in another, is most common, sometimes other formats are

FIGURE 13
Applications of Conditional Formatting

Group	Date	Line	Speaker	Talk	issue1	issue2	issue3	issue4	issue5	logistic	pedog	content	concep
sewage	4/5	0115	Darrin	Uh, it would be great.									
sewage	4/5	0116	Darrin	I mean I feel like if we could model things some how, in a way that was, consistent.						x		x	
sewage	4/5	0117	Darrin	For example, what we could-write things or, or uh, score things on a scale of, how cheap it was, how stable it is, uh, how effective it is,								x	x
sewage	4/5	0118	Darrin	But again those things are gonna be hard to determine without actually creating, ??								x	x
sewage	4/5	0119	Darrin	???									
sewage	4/5	0120	Kevin	Yeah, that's what-									
sewage	4/5	0121	Kevin	I guess I just thought that this was like, you come up with the guidelines, and then, you give it to them, and then they bring it, and then come here for a day, and then, like run their model, and then, they would do that.						x	x		

preferred. Ochs (1979) notes that physical layout can influence our interpretations of transcribed speech. For example, the standard top-down linear format can imply a causal relationship between sequential utterances that might not be the case with small children. A side-by-side format might be more appropriate.¹¹ Excel can be used to automatically generate one from the other.

In Figure 14, the last three columns show a side-by-side format that has been generated automatically. The formula seen in row 120 demonstrates how. If the value of the speaker column for a given row is equal to the value in the header row for the given column, then that cell is made equal to the speech cell. Otherwise, it is simply blank.

Notice also that some of the cells have different bordering. This is made using Conditional Formatting and is dependent on the *artifactaction* code shown in column M. This was a coding scheme that grew out of the *speechaction* code mentioned earlier. Hence, through the side-by-side formatting and the Conditional Formatting, we have a more visual depiction of the flow of the conversation.

The second example also is aimed at a visual depiction. The image in Figure 15 is actually part of an Excel Worksheet. Each row has been made only a single pixel high. Several of the coding schemes are shown across the top, and in each (very short) row, Conditional Formatting has been used to gray a cell if that particular code is present. We therefore get a sort of global view on the work. Figure 15 has just under 500 lines, representing about half an hour of conversation. Figure 16 shows a further application, adding in the filter function. In this case, a similar image has been produced, but the data

FIGURE 14
Generating Side-by-Side Transcript

Line	Speaker	Talk	artifact cition	artifact result	Darrin	Megan	Kevin
0117	Darin	For example, what we could-write things or, or uh, score things on a scale of, how cheap it was, how stable it is, uh, how effective it is.	add	Scoring on price	add: Scoring on price		
0118	Darin	But again those things are gonna be hard to determine without actually creating, ??	chal	Hard to do things if not	chal: Hard to do things if not		
0119	Darin	???					
0120	Kevin	Yeah, that's what-					
0121	Kevin	I guess I just thought that this was like, you come up with the guidelines, and then, you give it to them, and then they bring it, and then come here for a day, and then, like run their model, and then, they would do that.	sum m	thought this was come up w/ guidelines, give it to		=IF(\$D127= X\$7,\$L127&"&*\$SM127,")	summ: thought this was come up w/ guidelines, give it to them.
0122	Darrin	I think it'd be great if we could do that.	sup	Would like it	sup: Would like it		
0123	Darrin	But I don't thing it's possible with-					
0124	Darrin	I'm not sure about this.					
0125	Darrin	But I'm, I'm having a hard time seeing, um, students, put their model's, into, into motion.	chal	Doubt students can put models in motion	chal: Doubt students can put models in motion		
0126	Darrin	Just because it takes a long time to, filter these things, and for the microbes or the chemicals to kill whatever it is.	chal	Takes long time for microbes to kill	chal: Takes long time for microbes to kill		
0127	Darrin	And its pretty complex too, like in, uh-	chal	Complex	chal: Complex		
0128	Darrin	Unless we can design some kind of a parrallel system.					
0129	Megan	This, this one uses, um, like micro-organisms, um, kinds of coarse (media), like stones or plastic, and um, ??.	intro	Use of micro-organisms and coarse media		Intro: Use of micro-organisms and coarse media	

have been filtered to show only rows coded with issue 17. With this, we can focus on how issue 17 was discussed over the course of the project.

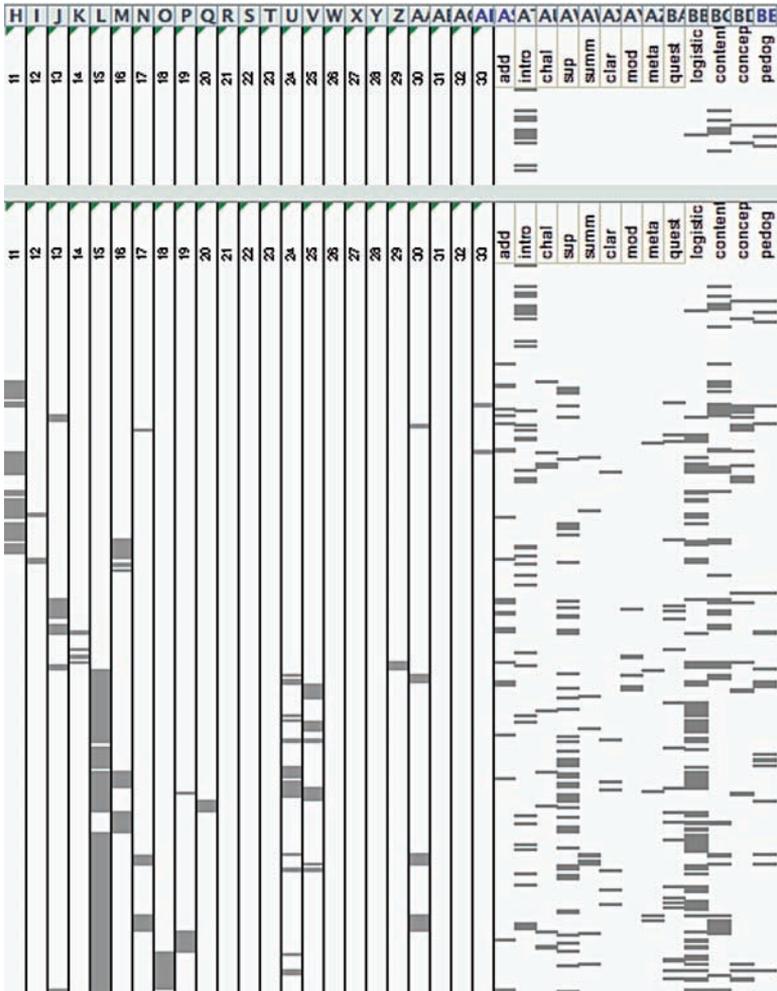
UNIT OF ANALYSIS

Having illustrated the use of Excel with qualitative data, we end with a brief discussion of how the structure of Excel impacts the unit of analysis and the pros and cons that result.

An Excel spreadsheet is essentially a database, with each row being a record. This presumes to some extent that the data can be broken up into such records in a regular manner before substantive analysis begins. Put another way, it means having a situation where a definable, regular unit of analysis is possible from the beginning. We have shown examples of effecting a change later in the process, but it does tend to be difficult.

There is one slight caveat to the constraint of having a regular unit of analysis. Although having the codable unit be variable is more difficult, having the *result* of coding regular units be sections of various sizes is quite

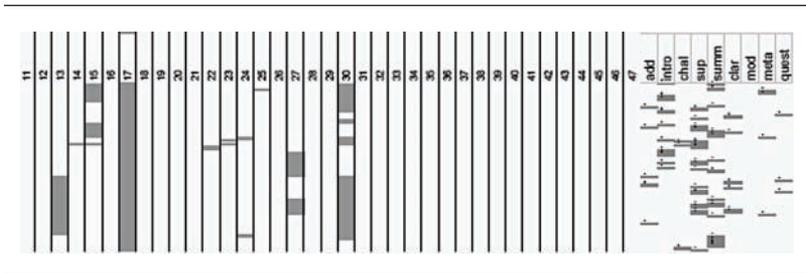
FIGURE 15
Global View



natural. The sections of data that are graphically shown in Figures 13, 15, and 16 are cases of this.

This leads in to how the unit of analysis structure of Excel can be seen as either a positive or a negative. In general, the difficulty in flexibility would seem to be an obvious con, and this is where analysis will begin with a very

FIGURE 16
Filtered for Issue 17



clear type of event in mind. However, for analysis that is more emergent from the data, such as Constant Comparative Analysis (Glaser and Strauss 1967), being forced to code an independently defined unit is quite helpful.

It should be emphasized that this should not only be a logistical decision. It is important that the data fit Excel's technical format, but it is also important that *making them fit* is in line with the theoretical and conceptual frameworks underlying the study. Choosing a particular unit of analysis is not just a matter of choosing what is convenient for storage. How the data are displayed and worked with will affect how they are interpreted. Demarcating the data into pieces that will be analyzed as a unit is itself an analytical act.

CONCLUSION

To reiterate the caveat we gave at the beginning, our aim has not been to argue for Excel's superiority or for a particular procedure for qualitative analysis. Rather, we have wanted to highlight potential features of Excel that may be overlooked, given Excel's predominant use for quantitative applications. Excel certainly is a number cruncher. However, its capabilities, as we have demonstrated, can extend to qualitative analysis applications as well. Conceptually, we summarize them as two abilities. First, it can simply house information—quantitative or qualitative. This includes the ability to organize data in meaningful ways. Second, Excel's "crunching" ability is not limited to numerical calculations. Rather, its logical functions can provide significant aid in qualitative analysis.

NOTES

1. All Excel images were made using the 2004 Mac version. All versions prior to Vista have a very similar “look and feel.” However, Vista represents a major redesign. Although all the functionality should be maintained, the location of features in the Vista version will be significantly different.

2. Our illustrations will be in Word, though they will not (and should not) utilize any of the advanced features that might be missing from a more basic text editor. A possible exception to this is the ability to show invisible formatting makers, which is very helpful but not necessary.

3. Excel actually has more flexibility in how it imports data, but tab-delineated files are a pretty clear choice. For example, you could use a comma-delineated file, but your transcript might include commas.

4. All of the examples in this article are from a study of science teacher education students collaborating to design innovative science curricula (Meyer 2003). The data shown are from recordings of group planning sessions. For another study making use of Excel, see Avery (2003). For a discussion of the analysis in both studies, see Avery and Meyer (2007).

5. Defining a sentence is not quite as easy as this implies but is beyond the scope of this article. In most of the examples we will show, we will refer to the sentence being used as the unit as a shorthand, even though the actual transcription rules that were used were more complicated.

6. We could have also included the label for the speaker—making the format identical to the first sentence—but there is a batch way to do this in Excel that we will note later. Again, anything that reduces the workload during transcription is generally worth it.

7. This is a good time to note the saying “Garbage in—garbage out.” The computer does what you tell it to do, not what you want it to do.

8. Each Excel file is referred to as a *workbook*. Each workbook can include multiple worksheets, which can be seen as parallel pages.

9. There was nothing special about three—more were simply never needed. However, it is certainly conceivable to have a circumstance in which there is a reason to limit the number.

10. Note that the gray coloring that also occurs was done through the normal, “manual” formatting process. This was done simply as a visual aid to guard against making a notation in the wrong column. The Contextual Formatting then overrides this if a condition is met.

11. Note that in the study these examples came from, this was really not the case. The traditional format was appropriate and preferred. This is an instance where we wish to emphasize what *can* be done rather than what *should* be done.

REFERENCES

- Avery, L. A., and D. Z. Meyer. 2007. Organizing and analyzing qualitative data. In *Teachers taking action: A comprehensive guide to teacher research*, ed. C. A. Lassonde and S. E. Israel, 73–88. Newark, DE: International Reading Association.
- Avery, L. M. 2003. Knowledge, identity, and teachers’ communities of practice. PhD. diss., Department of Education, Cornell University, Ithaca, New York.
- Glaser, B. G., and A. L. Strauss. 1967. *The discovery of grounded theory*. Hawthorne, NY: Aldine.

- Meyer, D. Z. 2003. Social engagement in curriculum design. PhD diss., Cornell University, Ithaca, New York.
- Ochs, E. 1979. Transcription as theory. In *Developmental pragmatics*, ed. E. Ochs and B. B. Schieffelen, 43–72. New York: Academic Press.

DANIEL Z. MEYER is an assistant professor in the Department of Mathematics and Science Education at Illinois Institute of Technology. His teaching interests focus on assisting science teachers in developing and implementing inquiry-based learning activities. His research interests focus on the development of teacher pedagogical content knowledge as a social process and the establishment of communities of practice. He recently coauthored, with Leanne Avery, the chapter “Organizing and Analyzing Qualitative Data” in Teachers Taking Action: A Comprehensive Guide to Teacher Research (International Reading Association, 2007).

LEANNE M. AVERY is an assistant professor of science education in the Department of Elementary Education and Reading at SUNY College at Oneonta. Her teaching interests focus on helping elementary teachers overcome their “science baggage” by using inquiry-based pedagogies to enhance their self-efficacy for teaching science. Her research focuses on rural children’s local science and engineering knowledge and place-based teacher professional development as a means of valuing and utilizing local rural knowledge in classroom practice. She recently coauthored, with Daniel Meyer, the chapter “Organizing and Analyzing Qualitative Data” in Teachers Taking Action: A Comprehensive Guide to Teacher Research (International Reading Association, 2007).