Analysis of Thin Online Interview Data

Toward a Sequential Hierarchical Language-Based Approach

Richard J. Kitto
John Barnett
University of Western Ontario

Abstract: Despite the best of intentions, qualitative researchers can be faced, in some circumstances, with having to make meaning from thin, or less than optimal, data. Using a real study as context, the authors describe the ways that they made sense of their thin data on teachers' perceptions of a large-scale evaluation instrument. They propose a rigorous, sequential approach to e-mail interview analysis based on three hierarchical language levels—lexical, semantic, and pragmatic—to save, by analysis, meanings that might otherwise have been lost because of limited data availability. Researchers are, of course, cautioned to avoid making claims that cannot be substantiated by the data.

Keywords: *online interviews; qualitative analysis; thin data*

Evaluators using qualitative methods can sometimes find themselves in situations in which they need to make conclusions based on less than a sufficient amount of data and/or context within that data; data that are important nevertheless. For example, they might be restricted in the number of visits to various sites being evaluated. They might be working within a politically charged atmosphere in which few participants are willing to communicate. They might even be evaluating programs over such a wide geographical area that they need to use electronic tools to find and contact participants and collect data. Finally, they might have to come to conclusions and make recommendations within a limited time frame that does not permit restarting the data collection. This article describes an approach to help evaluators deal with such problems.

The authors were faced with a situation in which they had small amounts of important, text-only, qualitative data obtained by e-mail about teachers' views of a large-scale assessment instrument. They could discard the data or try to find a method to analyze it, one in which they could have confidence.

The e-mail data were considered "thin" for two reasons. As qualitative researchers, the authors were aware that they did not have any first-hand knowledge of the individual situations—schools and classrooms in which the teachers were working—from the participants' e-mail messages. The researchers had used e-mail because they had not been able to visit and conduct face-to-face interviews. The size of the jurisdiction (all of the province of Ontario, Canada) meant that

Richard J. Kitto, Faculty of Education, University of Western Ontario, 1137 Western Road, London, ON, Canada N6G 1G7; phone: (519) 661-2111; fax: (519) 661-3833; e-mail: rjkitton@golden.net.

American Journal of Evaluation, Vol. 28 No. 3, September 2007 356-368
DOI: 10.1177/1098214007304536

participants were scattered over an extremely large geographical area (approximately 1.5 times the area of the state of Texas), and e-mail was the best data they could obtain. Second, they finished their data collection with a limited data set from a relatively small number of volunteer participants who were then teaching in a contentious and politically charged provincial atmosphere that might have discouraged participation.

Such thin data can produce problems of analysis for qualitative researchers because they do not provide enough evidence to support results (LeCompte & Preissle, 1993). Unlike quantitative data that can employ particular statistical methods to account and correct for very small samples of participants, thin qualitative interview data are difficult to analyze because qualitative analyses generally rely on what Geertz (1973) calls "thick description"—masses of data with large amounts of contextual information from multiple sources. This "thick" data helps to ensure that the understandings developed by researchers are credible.

Where there is a lack of sufficient data evaluators can—and should—go back into the field to attempt to collect more data. However, this strategy might not always work. For example, the subject matter might limit participation, or all the relevant data might have been collected. In other situations, the context of the research might have changed unalterably, or simple financial factors might limit participation (e.g., when potential participants are scattered over a large area). At the end of the day, evaluators have to make a choice to analyze the thin data they have or end the research project. This article is concerned with the process of making as much sense of thin data as possible.

Online Data

Despite their obvious shortcomings, e-mail interviews can provide an efficient mechanism for gathering data online. Researchers can exchange, in a relatively short amount of time, series of e-mail messages with participants in the same way interviewers who are physically present ask series of interview questions. Each exchange of e-mail messages is equivalent to a question and answer. Evaluators ask pertinent questions and analyze the responses in the same way that they would analyze interview data that were collected in person.

In this case, the evaluators started with a series of interview questions and sent them one at a time to each participant. The evaluators waited for a response before sending the next question. They analyzed each response and, when necessary, asked follow-up questions pertaining to each individual's answer. The ability to interact with participants' individual responses differentiates e-mail interviews from answers entered by participants in text boxes in online surveys.

Such a process not only is efficient for data collection, but the format of the data may provide additional benefits for analysis. First, online interview data might be more accurate because there is no loss or inaccuracy in transcription. Second, and more significantly, such interviews are not necessarily constrained by time (Roberts & Woods, 2000). Finally, the responses may represent participants' more considered judgment because they have had the opportunity to examine their answers, reflect on and rethink their positions, and edit the products. Consequently, the data might be richer than a simple count of lines of transcript would suggest. Indeed, there has been much work on using computer-mediated communication in qualitative research (Mann & Stewart, 2002).

However, when there is only a small amount of such data, new analytical approaches are needed because qualitative evaluators may have to conduct their research project without having been able to immerse themselves in their participants' environments. Therefore, the meanings they ascribe to participants and the theoretical understandings they generate could potentially lead to untrustworthy conclusions. For example, participants have both time and

opportunity to craft coherent positions without displaying discrepant behavioral cues such as facial expressions incongruent with their spoken words. Cues such as those might be discernable to a trained interviewer in a face-to-face interview, and they could then be used as the basis for probing deeper into a participant's thinking.

The authors assumed that e-mail participants would not be more likely to produce fraudulent answers than participants who were physically present. Rather, they held a view that some participants might attempt to frame their answers in more rational, reasonable, and coherent ways than if they had spoken the words aloud. After all, e-mail participants have the time and opportunity to edit out key utterances that they consider inconsistent with their overall view prior to sending their answers. So researchers might be faced with data that have lost some of their meaning through editing. Without an overwhelming mass of data to check for inconsistencies, coming to dependable conclusions could be very difficult.

The authors propose an approach to making as much trustworthy meaning as possible from such interviews: deep analysis based on general linguistic-like categories (Johnson, 1986; Kvale, 1989) in which key terms are counted and interpreted both alone and in context. The conclusions can then be used in triangulation with other data-collection methods. In other words, evaluators can test their conclusions against data from other sources. Although beyond the scope of this article, the evaluators did compare their conclusions from the e-mail interviews against their conclusions from data obtained in an online survey and against data from a pilot study that had been conducted earlier.

The technique they created will be described using data gathered in a project about teachers' perceptions of a large-scale secondary-school mathematics evaluation instrument.

Context

In the late 1990s, the government of Ontario, Canada, introduced provincewide testing for ninth grade mathematics students with an evaluation instrument developed and administered by the Education Quality and Accountability Office (EQAO), an arm's-length agency of the Ontario government. Without a long tradition of large-scale assessment in the province, the tests were not greeted well by teachers, teachers unions, or school districts, all of whom were concerned about the potential results. The authors attempted to uncover how this instrument affected teachers' classroom teaching and assessment practices. For various reasons, such as the geographical size of the jurisdiction and availability of online tools such as e-mail and cost, all data were gathered electronically.

The authors framed their analysis around Messick's (1989) model of validity. Messick proposed two aspects of construct validity in testing: test interpretation and test use. Test interpretation provides evidence about the performance of a person or deduces something about the person who has taken the test. Test use, however, involves making decisions that relate to the future of the person who took the test, and that potential result can make such tests high-stakes endeavors. For example, the results from standardized tests such as the LSAT, MCAT, or SAT can have profound effects on the life of the test taker; consequently, it is important that they be as valid as possible. Therefore, it is important that test interpretation and test use be based both on the evidence and on the consequences. Together, these facets of validity serve as the justification for making fair decisions based on the results of the assessment.

The authors were interested in the evidential and consequential bases of test interpretation (Messick, 1989) and asked research questions such as "Do teachers view the EQAO assessments as examining significant constructs in grade nine mathematics?" and "Do the EQAO materials help teachers employ new or better ways to assess their students?"

The original evaluators' study employed an online survey to collect both quantitative data (from Likert-scale questions) and qualitative data (from free-form optional comments). Unfortunately, only 72 participants responded. However, the data were augmented by data from multiple, sequential e-mail interviews with 10 teacher volunteers. This article solely deals with the e-mail data set consisting of only 30 pages of text.

Interviews, Transcripts, and Language Levels

Before beginning data analysis, a number of methodological approaches were examined. Most traditional interview analysis methodology (Creswell, 1998; Gubrium & Holstein, 2002) was not useful because it assumes that an interviewer can see, hear, and challenge the participants during the interview process. For example, Mishler (1986) discusses the joint construction of meaning by the researcher and participant, and Charmaz (2002) discusses grounded theory. However, analysis of an interactive, face-to-face process in real time is much different than that needed when the data are asynchronous and text based. Furthermore, e-mail interviewers are rendered passive (Fontana & Frey, 2000) after asking their questions because participants compose their responses in privacy in many different locales.

Discourse analysis, a method that investigates the ways that communication occurs between and among people (Berry, 1981; Wetherell, Taylor, & Yates, 2001), was also rejected because its main focus is on the structures that enable information to be transmitted, the surrounding text, and the sociocultural environment (Riggenbach, 2002). The structure of the e-mail interaction in this study was both very simple (researcher initiation–participant response) and identical for each interview. Even though the discourse structure of electronic texts has its own well-developed literature (Yates, 2001), the ways that language or electronic communication empower or disempower participants was not of concern. The sole interest was to uncover teachers' views.

The authors used the constant comparative method (Glaser & Strauss, 1967; Miles & Huberman, 1984, 1994). Specific units of data were classified (i.e., coded) inductively, and the codes were continuously compared to prior codes as data analysis proceeded and were refined as necessary while new data were still being added.

Many times, evaluators do not start with a preconceived notion of how to code the data or with a theoretically derived framework of codes. In other words, as the data are coded, evaluators continually create new codes when necessary and delete old codes that do not fit the developing framework in the researchers' minds. When necessary, evaluators recode formerly coded data so that the system of codes remains coherent as it develops. This process of theory generation was called grounded theory by Glaser and Strauss (1967). However, the term *emergent* (Strauss & Corbin, 1998) was used to better reflect the process of having codes arise from constant reinterpretation. Emergent theory was also consistent with the epistemology (Elgin, 1996) of constructivism and allowed for understandings to be established with fewer preconceptions. However, additional tools were needed because of the thin data.

Language Analysis

Elements of linguistics were very useful in the analysis, but the authors needed to use terms in ways that somewhat differed from the standards of formal linguistic analysis. For example, at a simplistic level, the term *token* rather than *type* was used to stand for each unique word in a text. The number of tokens was therefore a subset of all the words in that text. This was not a linguistic analysis of the text per se; rather, it was an analysis of the words

to support or challenge their developing theory. Thus, the authors thought of their method as language analysis rather than linguistic analysis.

They employed verbal analysis, "a methodology for quantifying the subjective or qualitative coding of the *contents* of verbal utterances" (Chi, 1997, p. 2) as their general approach. In verbal analysis, the evaluator literally counts different kinds of words to reduce the subjectiveness of coding. This is a quantitative–qualitative approach because it quantifies an evaluator's "subjective impression by coding the verbal evidence for that impression and comparing the frequencies of the codes quantitatively" (Chi, 1997, p. 7). To implement the approach, overarching analytical levels had to be invented that could be used as the organizing principles for the analysis.

Two widely disparate frameworks suggested similar analytical approaches. First, Jensen (1989, p. 100), proposed a four-level structure: phonemic, grammatical, semantic, and pragmatic. The phonemic level (i.e., sounds) was applicable only to face-to-face interactions because Jensen's model was applied to oral interactions, and there were no such data, but the other levels appeared appropriate.

> The *grammatical* form of sentences, for one thing, can point to some general assumption that a respondent is making, for example implying a particular self-image. (p. 100)

> The *semantic* analysis of meaning . . . can also serve to identify a particular conceptual organization of and orientation towards social reality. A complete listing of the terms which a respondent associates with the individuals, groups or institutions talked about in the interview is a simple but powerful first step in such an analysis. (p. 100)

> The *pragmatic* level . . . [means that] in discourses of interviewing, textual forms carry general modes of consciousness as well as specific worldviews. It is at this general level of discourse, then, that it becomes possible to interpret what is the "message" of the interview. (p. 103)

Computer science provided a second lens. Johnson (1986), in writing about artificial intelligence, described five levels of performance that a computer-based language translator would need.

> Together, these five kinds of knowledge—pragmatic, semantic, syntactic, lexical, and phonological—form a hierarchy of abstraction, a ladder that ascends from the specific to the general. Beginning on the bottom rung with the low-level, highly specific knowledge of phonemes, we climb upward, to knowledge of words, grammar, meaning, and finally to general, pragmatic knowledge about life. (p. 127)

Johnson's framework corresponded remarkably well with Jensen's (1989) model. Just as Jensen's phonemic level was not needed in the analysis, so too Johnson's phonological level was unnecessary to deal with the transcripts. However, by combining their models and deleting the phonemic and phonological levels, four levels of analysis could be created (see Table 1).

However, only three levels, lexical, semantic, and pragmatic, were fruitful. The syntactic level was not fruitful for analysis because the inherent structure of participants' answers to e-mail questions such as "To what extent do you agree with the expressed goals of EQAO regarding instructional practices and assessment strategies?" was always in the active voice. Thus, there was no differentiation. It was the same for first person and third person use. Teachers tended to write about themselves in the first person and about others in the third person. Therefore, no additional meaning was clear.

Table 1
Language Levels Used in Transcript Analysis

Level	Nature	Typical Analysis Level	Main Goal of Level
Lexical	Words	Individual words	Establish tone
Syntactic	Structure	Active or passive verbs; personal or impersonal pronouns	Confirm tone
Semantic	Meaning	One- or two-sentence blocks	Find basic arguments
Pragmatic	Use in context	Whole answers	Combine longer arguments and narratives

The three levels that were used (lexical, semantic, and pragmatic) were like three different sizes of "meaning filters," much like nets of different-sized mesh. Through sequential use of these filters, smaller filters could trap the finest grain of meanings, whereas the larger filters would catch bigger ideas. Used together, they should have captured meanings at different levels that could then be compared for consistency.

At the lexical level, the individual word was the unit of analysis. Data consisted primarily of declarative statements with distinct nouns and adjectives. For the semantic level, the sentence was the unit of analysis. Occasionally, two or more sentences had to be combined to more clearly demonstrate an individual thought. Finally, one or more paragraphs were used as the analytical unit for the pragmatic level. At this level, meaning could be made about the ways the participants were using their answers to create overarching frames of reference.

Description of Transcript Analysis

Each transcript was analyzed at lexical, semantic, and pragmatic levels. The lexical analysis involved generating concordances for each interview question by participant, combining the concordances into a spreadsheet file for each participant, and leading finally to lexical data reduction.

The process started with a transcript of an e-mail question and a response.

> Evaluator: Are you personally influenced by the evaluation scores reported by EQAO? (December 9, 2004)
>
> Participant SV: As a teacher of grade nine, no, I taught at [high socioeconomic status] High School, a very rich school, and now I teach at [low socioeconomic status] Collegiate Institute (ranked as the 11th most neediest school in [my city]).
>
> I know my teaching has not changed, and there is a great difference in the school scores. . . . There are so many external factors that can effect the EQAO scores that they are not valid. I would like to see a mathematical analysis of EQAO scores relating scores and family income, new immigrants, etc.
>
> There are many very good teachers who are identified as poor teachers because their students did not perform on EQAO. A judge of a good teacher is not only their ability to get the concepts of their subject across to the students, but to teach the students a multitude of topics outside their subject area, like being a functioning member of society, etc. (December 12, 2004)
>
> Evaluator: To what extent have you changed your assessment strategies to be more consistent with the recommendations from EQAO? (December 16, 2004)
>
> Participant SV: Not much, parents do not understand rubrics in mathematics. They understand right and wrong, they do not understand projects, and other forms of assessment. . . . They base their child's achievement on tests, quizzes, and exams. Even if you educate them otherwise, they do not believe the value. (December 18, 2004)

The transcripts were first analyzed at the lexical level.

Figure 1
Words and Frequencies From the Interview Extract

Home > Freq List Builder > Frequency Text Input > **Freq. List Output**

Text: **Untitled**
Date: **3/16/2007 11:1**
Tokens: **197**
Types: **115**
Ratio: **0.5838**
Sort: **descending**

RANK	FREQ	COVERAGE individual	COVERAGE cumulative	WORD
1.	8	4.06%	4.06%	A
2.	8	4.06%	8.12%	NOT
3.	7	3.55%	11.67%	OF
4.	7	3.55%	15.22%	THE
5.	6	3.05%	18.27%	AND
6.	5	2.54%	20.81%	THEIR
7.	5	2.54%	23.35%	THEY
8.	4	2.03%	25.38%	ARE
9.	4	2.03%	27.41%	I
10.	4	2.03%	29.44%	SCHOOL
11.	4	2.03%	31.47%	SCORES
12.	4	2.03%	33.50%	TO
13.	3	1.52%	35.02%	AS
14.	3	1.52%	36.54%	DO
15.	3	1.52%	38.06%	EQAO
16.	3	1.52%	39.58%	IN
17.	3	1.52%	41.10%	STUDENTS
18.	3	1.52%	42.62%	THERE
19.	3	1.52%	44.14%	UNDERSTAND
20.	2	1.02%	45.16%	AT
21.	2	1.02%	46.18%	ETC

Step 1: Lexical Analysis—Generating a Concordance

Each interview was analyzed using a Web site called the Compleat Lexical Tutor (Cobb, 2004). The Web site contains a variety of tools that allow researchers to input text and generate output in a wide variety of formats such as concordances, word strings, statistics, and cloze tests. The Web site was used to create a complete concordance of tokens (i.e., unique words) for each interview. Although the process was completed for all interviews, part of a sample result (see Figure 1) illustrates the type of analysis carried out using the extracts of text previously shown in this article.

Step 2: Lexical Analysis—Spreadsheet Word Count

In the actual evaluation, the words used by each participant in each e-mail answer were copied to columns in a spreadsheet (see Figure 2), resulting in 44 sets of words, 6 from participants who answered the basic four questions and 4 from participants who answered the basic four questions and also provided additional thoughts as an answer to a fifth question.

Figure 2
A Sample of Lexical Counts for One Interviewee

	A	B	C	D	E	F	G	H	I	J
1		**Q1**		**Q2**		**Q3**		**Q4**		**Q5**
2		461	words	84	words	26	words	79	words	634
3		231	tokens	63	tokens	25	tokens	56	tokens	291
4		**Participant # ...**								
5		A 5		A 2		A 1		A 2		A 11
6		ACCURATE 1		ALWAYS 1		ALL 1		ADVANTAGES 1		ABOUT 3
7		ACTUALLY 1		AM 1		AS 2		AFFECT 1		ACTIVLY 1
8		ADVANTAGED 1		AN 1		AT 1		ALWAYS 1		AFFECT 1
9		AGAIN 1		AND 4		BIG 1		AND 4		AFRAID 1
10		ALIGN 1		APPROACH 1		BOARD 1		APPROACH 2		AGAIN 2
11		ALIGNED 1		AS 1		BY 1		AS 1		AGREE 2
12		ALL 5		ASSESSMENT 1		CULMINATING 1		AT 1		ALL 2
13		ALMOST 1		CHANGE 1		DO 1		CATEGORIES 1		AM 1
14		ALSO 2		CURRICULUM 1		DOES 1		CONTEMPT 1		AMAZED 1
15		AM 2		DEVELOPING 1		IT 1		CURRICULUM 3		AMBIGUITY 2
16		AMBIGUOUS 1		DIFFICULT 1		NO 1		DELIVERING 1		AMOUNT 1
17		AMOUNT 1		DO 1		NOT 1		DIFFERENT 2		AN 1

Table 2
Negative or Positive Word and Unique Word Counts

	Words		Tokens	
Negative or Positive Data Reduction	Negative	Positive	Negative	Positive
All interview content	12,381		1,953[a]	
Significant words or tokens	2,649		294	
Apparently emotive words or tokens	189		89	
	136	53	69	20
Emotive words or tokens after context check[b]	95		52	
	82	13	46	6

a. Distinct count of tokens for all interviews collectively.
b. Checking the words in context revealed that many were neither negative nor positive or were not related to the assessments.

Next, the tokens were examined in each of these 44 sets of words. Some tokens appeared to be highly emotive, whereas others were less so. The highly emotive tokens appeared to be either positive or negative by their apparent affective denotation. These were the researchers' emergent categories.

Step 3: Lexical Data Reduction

The two evaluators independently examined the lists of tokens and coded each token as positive (e.g., "amazed"), neutral (e.g., "categories"), negative (e.g., "contempt"), or insignificant (e.g., "as"), without regard for their part of speech. Then they met and compared their sets of codes. Whenever a difference existed, the evaluators negotiated the meaning of that token until both could agree on the coding for each token. Methodologically, this process was meant to provide a foundation for further analysis of the tokens at the lexical level.

The total numbers of words and tokens were then analyzed for all interview questions for all participants. To avoid the problem of redundancy, interview transcripts were compared, and the duplicate tokens found in the texts of all the interviews were removed. The sum of the tokens prior to reduction was 1,953. After the duplicate and insignificant tokens were eliminated, the process resulted in a list of 294 significant tokens (as shown in Table 2). Of those, 89 tokens appeared to be emotive, with 69 negative (e.g., "afraid," "deficiencies," "discourage," "useless") and 20 positive (e.g., "amazed," "brilliant," "delighted," "excited"). A similar process was employed with words as a double check.

However, many of the tokens were not what they seemed when examined in context and had to be recoded. There were four possible kinds of changes in this process: positive to negative or vice versa and positive or negative to neutral. Examples of changes from positive to negative included "incredible" ("can't deal with the incredible amount of content") and "utopian" ("unfortunately they are set for a utopian world"). An example of the opposite was "dislike" ("Some colleagues dislike EQAO testing. . . . I believe that EQAO type questions do align themselves nicely with what we are asked to do."). Similarly, many apparently emotive tokens were benign in context, for example, "encouraged" ("the different learning styles that we are being encouraged to address"). When a term signified a meaning other than its apparent meaning, it was recoded in the appropriate category. A quick check of the tone of the responses was provided by the ratio of negative to positive words (82:13) and tokens (46:6). By either measure, negative terms heavily outweighed the positive ones.

Table 3
Semantic Code Frequencies (Sentences)

Interpretation	Frequency	Use	Frequency
Evidence			
Test fit (concepts, curriculum, and assessment)	19	Prediction of 10th grade performance	14
Test validity	41	Use of EQAO score in school grading	20
EQAO score and class generated grade	17		
Consequences			
Test relation to teaching goals	39	Test effect on classroom practice	22
Politics, accountability, or cost	22	Test effect on teacher grading	37
Suggestion for improvement	14	Personal consequences	24
General issues			
Curriculum concerns	29		
School management issues	7		

Note: EQAO = Education Quality and Accountability Office.

Step 4: Interview Semantic Analysis—Interview Coding in Atlas.ti

At the semantic level, the data corresponded fairly well to Messick's (1989) notion of test validity. Participants' answers could generally be placed in four quadrants based on their interpretation or use of the test compared to the evidential or consequential nature of their views. However, there were some general issues brought up by the participants that did not fit within the model. The numbers in Table 3 represent the number of sentences.

The use of qualitative codes and counts of the instances of each code simultaneously provided two levels of information. It provided a guide to the reasons for the feelings. For example, the impact of the assessments on teacher marking (37 sentences) was slightly more important than the concern with prediction of 10th grade achievement (14 sentences). Second, summary statements generated from the interviews could be compared to the lexical analysis results. Most of the comments were related to the negative social consequences of the tests themselves. The second largest proportion came from those who felt the content of the test was not valid (another negative feeling).

Step 5: Pragmatic Analysis

All coded interview data were sorted by code and then by participant and saved as one worksheet; data then were resorted by participant first and then by code and saved as a second worksheet. Both worksheets were then exported to a word processor, producing two searchable 50-page documents. Relating all quotations to a particular code (in the first document) generated the overall flavor of the responses to particular questions in the interviews. The second document organized all of the quotations by author, then by code, and was useful for organizing the data for member checking as described below.

As at the other levels, a large number of negative feelings were expressed toward the test. The main theme that emerged was the lack of connection that teachers felt between their classroom assessments and those of EQAO. Interestingly, even though the pragmatic analyses tended to strongly support the meanings made at other levels, teachers themselves tended to differ in their overall views. Eight out of the 10 seemed to be very negative toward the test, whereas 2 were neutral or supportive of the test. Member checks confirmed that the conclusion to which the researchers had come and the summaries they had prepared were at least fair representations of participants' views. These meanings were congruent with the online survey results.

Member Checking

To confirm that the interviewees agreed with the analysis of their answers, the information for each participant was copied into a Word table that grouped all of the comments within a category in the left-hand column with a space in the right-hand column for a summary (Table 4 illustrates the process).

Next, the summaries for each participant were merged into a set of Word documents that also contained brief descriptions of the research and the four categories of questions. These statements were e-mailed to the participants. During the next 2 weeks, most participants replied that they agreed with the summary of their views (three of them enthusiastically), and none of the others suggested any changes when explicitly asked. No participants reported any disagreement, and none asked to withdraw their comments. This suggested that the meanings we had made and the summaries we had prepared were at least reasonably accurate representations of their views.

Discussion

The analysis of thin qualitative data is difficult. Without adequate thick description, there is the danger that evaluators using qualitative data could make meanings that are not supported or supportable. As a result, some evaluators discard data that are not sufficiently "thick." Unfortunately, the meanings contained within that data are always lost when they are abandoned. Of course, thin data will always contain less meaning than thick data; there is less substance to confirm the meanings.

Most standard approaches to qualitative data analysis do not work well because they require thick data, phonological or behavioral observation, or direct, real-time interaction with the evaluator. This article has described a method for analyzing thin data with more confidence. However, evaluators using this method have to take great care they do not use methods that create artificial meanings or make more meaning than is contained in the data.

A language-based method that analyzes data using a number of different filters can produce meanings at three different levels: lexical, syntactic, and pragmatic. The finest-grained filter can look for lexical meanings at the individual word level. The next filter can help to analyze the syntactic meanings of sentences. The third, or pragmatic, can more holistically analyze meanings.

The lexical analysis of e-mail interviews provides a look at the words chosen, although many words belonging to emergent categories such as positive and negative have to be checked in context to verify the interpretation. Although the semantic analysis is primarily qualitative, counting specific instances of codes helps to construct an overall explanatory structure that can be compared with the meanings at the pragmatic level.

The meanings that researchers make also have to be carefully checked for inconsistencies at the various levels. When the meanings made at the three levels are reasonably coherent, researchers can be fairly confident that they have retained as much meaning as is contained in those data. Conversely, when the meanings are inconsistent and the data analysis does not produce a clear picture, researchers might choose to rethink their conclusions. Perhaps other meanings need to be explored until all the various levels of analysis produce a result consistent with a single theoretical framework. In such cases, it may even mean that the researchers have to reconsider, recategorize, and perhaps even reconceptualize the evidence itself. Indeed, even at the risk of alienating former participants, researchers might need to ask them for clarification and/or seek responses from new participants until the level of coherence reaches the point

Table 4
Data Sorted by Author Then by Code for Member Checking

Selected Quotations From Interview With "SV"	Researchers' Summary for Member Checking
We tend to teach to EQAO the final days before writing it.	You are frustrated by some of the "side effects of the EQAO assessments. Teachers were expected to make changes in their teaching style without knowing if it was a good idea to do so, and the changes to be made were not all that clear. Also, teachers get judged on how their students performed on EQAO regardless of other factors. Finally, parents don't understand how the EQAO assessment works, and don't understand the value of alternative methods of assessment such as projects. In addition, you have identified some consequences for students, particularly test anxiety, and, for the applied-level students, an even greater loss of self esteem.
Forced a change in style of teaching mathematics. Teachers were expected to embrace this style/method (which wasn't articulated to them) which was touted as being better for the student.	
Not at all, no recommendations were made specifically, I had to determine those for myself and I don't know if the changes I made were beneficial or not.	
Time to finish it in depth. . . . Making my marks a little lower. . . . No one asks, did I have enough time to finish the course on EQAO.	
Not much, parents do not understand rubrics in mathematics. They understand right and wrong, they do not understand projects, and other forms of assessment. . . . They base their child's achievement on tests, quizzes and exams. Even if you educate them otherwise they do not believe the value.	
I always tried to teach mathematics using manipulatives (making me different from others) this means I have always assessed using different techniques. . . . Sometimes frowned upon by my test and quiz colleagues.	
Good teachers who are identified as poor teachers because their students did not perform on EQAO. A judge of a good teacher is not only their ability to get the concepts of their subject across to the students, but to teach the students a multitude of topics outside their subject area, like being a functioning member of society etc.	
A lot more anxiety from the students about over testing.	
A lot more poor self esteem for the students in the applied program look at the stats there.	

Note: EQAO = Education Quality and Accountability Office.

that the theoretical conclusions explain most, if not all, of the data and that no new meanings can be created to effectively explain it.

Multiple-level analyses of data also need to be compared to other data sources. In fact, without good member checking and adequate triangulation from other sources, the method described might still produce incorrect meanings. Thus, these levels of analyses do not replace triangulation, in part because the same data are being used at all levels.

Finally, even though this method can be effective and feasible for evaluation studies, it is a very labor-intensive method of extracting meaning. Hence, the final caveat is that evaluators have to decide if the problem to be studied is significant and worthwhile enough on its own merit before undertaking this type of analysis. Overall, the method described might be a useful one for evaluators confronted with relatively small amounts of important data that lack context.

The interplay of these different levels of analysis produces more material to allow researchers to offer their "considered judgment" (Elgin, 1996) with respect to their conclusions.

References

Berry, M. (1981). Systemic linguistics and discourse analysis: A multi-layered approach to exchange structure. In M. Coulthard & M. Montgomery (Eds.), *Studies in discourse analysis* (pp. viii-198). London: Routledge & Kegan Paul.

Charmaz, K. (2002). Qualitative interviewing and grounded theory analysis. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context & methods* (pp. 603-628). Thousand Oaks, CA: Sage.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences, 6*(3), 271-315.

Cobb, T. (2004). *The compleat lexical tutor*. Retrieved March 16, 2007, from http://www.lextutor.ca/

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.

Elgin, C. Z. (1996). *Considered judgment*. Princeton, NJ: Princeton University Press.

Fontana, A., & Frey, J. H. (2000). The interview: From structured questions to negotiated text. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 645-672). Thousand Oaks, CA: Sage.

Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In *The interpretation of cultures: Selected essays* (pp. ix-470). New York: Basic Books.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine.

Gubrium, J. F., & Holstein, J. A. (2002). *Handbook of interview research: Context & method*. Thousand Oaks, CA: Sage.

Jensen, K. B. (1989). Discourses of interviewing: Validating qualitative research findings through textual analysis. In S. Kvale (Ed.), *Issues of validity in qualitative research* (pp. 93-108). Lund, Sweden: Studentlitteratur.

Johnson, G. (1986). *Machinery of the mind*. New York: Times Books.

Kvale, S. (1989). Introduction to issues of validity in qualitative research. In S. Kvale (Ed.), *Issues of validity in qualitative research* (pp. 7-12). Lund, Sweden: Studentlitteratur.

LeCompte, M. D., & Preissle, J. (1993). *Ethnography and qualitative design in educational research* (2nd ed.). San Diego, CA: Academic Press.

Mann, C., & Stewart, F. (2002). Internet interviewing. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context & method* (pp. 603-627). Thousand Oaks, CA: Sage.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-104). New York: Macmillan.

Miles, M. B., & Huberman, M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills, CA: Sage.

Miles, M. B., & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Mishler, E. G. (1986). *Research interviewing: Context and narrative*. Cambridge, MA: Harvard University Press.

Riggenbach, H. (2002). *Discourse analysis in the language classroom. The spoken language* (Vol. 1). Ann Arbor: University of Michigan Press.

Roberts, P., & Woods, L. (2000). Alternative methods of gathering and handling data: Maximising the use of modern technology. *Nurse Researcher, 8*(2), 84.

Strauss, A. L., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.

Wetherell, M., Taylor, S., & Yates, S. J. (2001). *Discourse as data: A guide for analysis*. London: Sage.

Yates, S. J. (2001). Researching Internet interaction: Sociolinguistics and corpus analysis. In M. Wetherell, S. Taylor, & S. J. Yates (Eds.), *Discourse as data: A guide for analysis* (pp. 93-146). London: Sage.