

Analyzing Interview Data: The Development and Evolution of a Coding System

Cynthia Weston,¹ Terry Gandell, Jacinthe Beauchamp, Lynn McAlpine, Carol Wiseman, and Cathy Beauchamp

This paper describes the process used by a research team to develop a coding system for analyzing data from interview transcripts and situates the process within approaches to qualitative analysis. Successive versions of the coding scheme illustrate its development over several years; the role of team members and verification in this evolution are discussed. Several lessons emerge from our experience: a) coding is not what happens before analysis, but constitutes an important part of the analysis; b) a team builds codes and coding builds a team through the creation of a shared understanding of the phenomenon; and c) collaborative qualitative research requires a kind of rigor that an independent researcher might not be aware of or need.

KEY WORDS: coding; interview data; qualitative research; research methodology; data analysis.

INTRODUCTION

Much is written about methods for coding interview data: ways to think about it, how to do it, how to document it, and how to report it. We have coded mountains of verbal data in the form of interview transcripts in various contexts during the past decade, and have reported the results of our investigations (e.g., Weston et al. 1997; McAlpine et al. 1999a; McAlpine et al. 1999b). However, these reports are the tip of the iceberg. We have not written about what lies below the surface—the enormous task of developing coding systems and coding the transcripts that provide the very basis for reporting the results of our investigations. Our research team (the authors of this article) has worked together since 1994 to develop a

¹Correspondence should be directed to Cynthia Weston, Department of Educational and Counselling Psychology and the Center for University Teaching and Learning, McGill University, Montreal, Quebec, Canada H3A 1Y2; e-mail: weston@education.mcgill.ca.

coding system to analyze and understand our current data. We have learned a great deal about coding during this time that may be of interest to other researchers who deal with similar data (interview transcripts), use similar methods (coding), and who work or imagine working collaboratively in qualitative research.

Maxwell (1990) notes that there has been a great deal of abstract debate about different paradigms in education, but little analysis of actual examples of research that combine different methods to determine what works and why. In this article we provide an example of an approach to coding verbal data that we believe represents a combination of methods. For example, we used a priori theory to frame our questions, drive our interview protocols, and structure the initial levels of the coding scheme. We moved into a more grounded approach as we discovered codes working through the transcripts.

Influenced by Berliner's (1992) notion of telling stories, we share this story of how our coding system was developed and how it continues to evolve. We describe our context, the process we used to develop the coding system, how we went about verifying codes, and our codebook. We believe there are several lessons in this story which are woven through the account. First, coding is not what happens before analysis, but comes to constitute an important part of the analysis. Second, a research team builds codes and coding builds a team through the creation of shared interpretation and understanding of the phenomenon being studied. Third, collaboration in qualitative research requires a kind of rigor that a lone or independent researcher might not be aware of or need. In this account, we try to "keep a sense of voice present" (Eisner 1998, p. 2) so readers will know the people behind the words, naming team members as relevant to explicate their roles in the complex work of developing a coding system for analyzing interview data. We recognize that some issues we discuss are more relevant for collaborative team projects, while others will be relevant to anyone coding qualitative data.

OUR CONTEXT

Several aspects of context have influenced our thinking about research methods and provide a basis for understanding the development of our particular coding scheme: our discourse community, our research perspective, the nature of collaborative team research, the phenomenon of the study, and preexisting conceptual frameworks we brought to the analysis. We think it useful to describe these briefly before moving into a discussion of the coding system.

Community

First, it is important to recognize the influence of our discourse community. We are in a department of educational psychology whose members use research

methods that span a continuum from quantitative to qualitative, with many using methods from both traditions. The principal professors in the research team, Lynn McAlpine and Cynthia Weston, have been actively engaged in this community for fifteen to twenty years. It is a community of scholars, as Bereiter and Scardamalia (1989) discuss it, in that we learn from and influence each other's growth in knowledge, which is reflected in our thinking about methods. Second, our major funding agencies still tend to prefer somewhat more quantitative approaches. Although we might like our approaches to be independent of such constraints, we are sensitive to these preferences in a competitive funding environment.

Research Perspective

Overall, we consider ourselves within the interpretivist tradition (e.g., Erickson 1986; Miller, Nelson, and Moore 1998), although some may see objectivism in our methods. We are uncomfortable with, and, in fact, resist using a particular label to name our approach, seeing in our work aspects of methodology that are frequently associated with phenomenology, grounded theory, case study, and other perspectives. In part, this is because research team members come from different traditions, educational technology and educational psychology, with differing experience and preference for qualitative and quantitative methods. In addition, Cynthia Weston, Lynn McAlpine, and Terry Gandell have years of professional and personal experience as members of the Center for University Teaching and Learning, which focuses, in part, on working with professors to enhance their teaching. These personal histories have influenced our research perspectives and our choice and range of methods.

Collaborative Team Research

The two principal professors, Lynn McAlpine and Cynthia Weston, have favored a team approach to research since the beginning of their collaboration in 1990 as an opportunity to co-construct knowledge, train graduate students, and expedite analysis. Terry Gandell, chosen for her experience and expertise, developed the initial coding scheme. Three graduate students—Jacinthe Beauchamp, Carol Wiseman, and Cathy Beauchamp—did the coding. Since our methods generate large quantities of data, a team approach to analysis helps complete coding and results in richer, more complex interpretations. The evolution of understanding is accelerated exponentially as the group struggles to articulate the characteristics of a phenomenon. We find that a collaborative approach influences just about everything in the research process: how data is collected, how the coding system is developed, applied, and verified, and ultimately how understanding is constructed.

Phenomenon of Study

In 1993 we received funding to explore how experienced professors use reflection to improve their teaching and the nature of their reflective processes. We documented, through interviews, the reflections of six university professors during the planning and delivery of a one-semester course. The hundreds of pages of interviews that resulted were transcribed, were verified by each professor, and became the focus of our research—what we coded and analyzed. Thus, our approach was also influenced by the nature of the research program, which took place over a long period of time and required procedures for managing large amounts of data.

Conceptual Frameworks Influencing the Coding

Our approach to the collection and analysis of interview data was influenced by a number of conceptual frameworks that come out of our context and communities. On one hand, how researchers see data and the meaning attributed to it is what makes data useful, interesting, and a contribution to knowledge. On the other hand, our biases and perspectives influence interpretation throughout analysis—from how codes are developed to how results are interpreted (e.g., Conostas 1992; Erickson 1986; Webb 1997; Howe and Eisenhart 1990; Kirk and Miller 1986). We have tried to recognize conceptual frameworks that have influenced our interpretations and analysis, although some may be so implicit that we can't even see them (much less report on them). We are very aware that these frameworks are a context for understanding our approach to coding interview data.

First, existing theories helped to frame our research questions, influenced the structure of data collection, and influenced our coding system. Creswell (1998) would likely call this an *a priori* theoretical orientation and Charmaz (1990) a logical deductive approach, in contrast to, for instance, grounded theory, where the questions would become clarified during data analysis. In developing the research proposal, we examined the literature from several theoretical perspectives, searching for constructs that would allow a deeper and more sophisticated understanding of the processes that are linked to improvements in teaching. This corresponds with suggestions that researchers develop tentative theories of what is happening with the phenomenon being studied and why (e.g., Maxwell 1996). To do this, we drew from the literature to build a tentative model of reflection (Fig. 1). This initial model influenced the development of our research questions, interview protocols, and subsequent data analysis. In other words, it led us to prestructure many aspects of our design and methods (e.g., Maxwell 1996). For instance, we were interested in knowing what professors were paying attention to in the classroom as they were teaching, and what changes they made in their plans. We created semistructured interview protocols that included questions related to the constructs in the model; the structuring helped to ensure consistency across interviewers and across

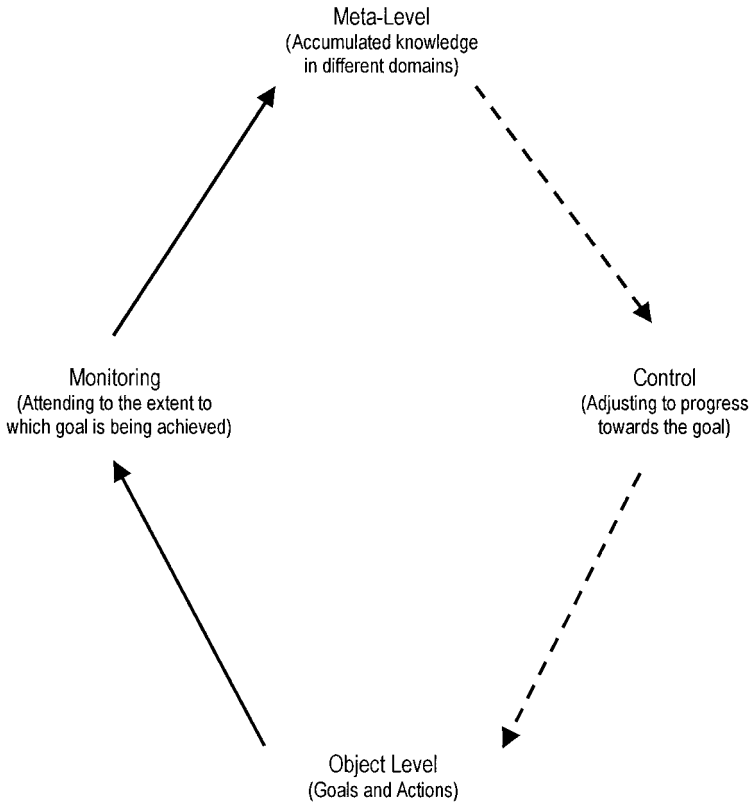


Fig. 1. Model of Reflection (1995).

data sources (e.g., Maxwell 1996). The use of a priori theory at the outset of the study was very helpful in narrowing the range of possibilities for interpretation. Using existing constructs made an incredibly ill-defined problem into a moderately ill-defined one, but left us open to change and discovery.

Other literature relating to the analysis of verbal data using protocol and discourse analysis (e.g., Bracewell and Breuleux 1994; Breuleux 1991; Frederiksen 1986; Newell and Simon 1972), which we had used extensively in previous research, also influenced our approach to coding. These methods are qualitative in nature but draw on structuralist methods. Our familiarity with these methods played an important role in the codes we developed for this research.

Lastly, our academic training influenced our coding. Several team members have a strong background in instructional design, and we tend to deconstruct many issues related to instruction using common terms from instructional design. Quite automatically, we began to see and name constructs from this field in professors' transcripts.

DEVELOPING THE CODING SYSTEM

We have been developing the notion that research communities predispose us to approach research with certain conceptual frameworks. If we were in another field, our frameworks would undoubtedly be different. Recognizing these, we now describe how we approached the development of the coding system. Although there are many aspects that could be discussed, we have chosen to focus on the dialectical and recursive process the team used to develop the codes, the structure of the coding system, and the unit of analysis.

The Dialectical Process

We used a recursive, iterative process in developing the codes and increasing our understanding of the phenomenon. We are pleased with the outcome, but the process took much longer than any of us had imagined. We attempt to provide a realistic description of the protracted and complex process, which may sound familiar to those who have worked in a team context, and may be revealing for qualitative researchers who are planning to move towards more collaborative work.

Conceptualizing the Coding: Lynn, Cynthia, and Terry. Work on developing codes that would capture the process of reflection began in the summer of 1994 when Terry undertook the task of developing the initial coding system. She met with Lynn and Cynthia to discuss the relationship between the research questions and the coding system. We wanted to pursue several constructs that were explicit in the research questions: what professors were monitoring while teaching and how they controlled or adjusted their actions to achieve their goals. Although we saw each professor as a case, we also wanted to see what patterns might emerge across interviews and professors. Thus, the coding system should allow us to build a rich understanding of the nature of each professor's reflection on teaching, and allow us to aggregate the data to search for patterns.

Generating the Codes: Terry Working Alone. In the fall of 1994, with the research questions as a frame of reference, Terry read through a selection of transcripts and noted when professors made evaluative statements or judgments about their teaching or student learning ("monitoring"), and any action taken by the professors to change or modify their teaching or student learning ("control"). These became the first two categories of the coding system. Using these categories, she then read through selections of each professor's transcripts, noting themes that emerged more than twice in relation to each of the two categories. She developed a draft lexicon to help demarcate each theme by identifying words that tended to be markers of *monitor* and *control*.

Four codes were tentatively identified. Two codes related to monitoring: *cues* and *how monitored*; two codes related to control: *what changed* and *how changed*. In addition, two subcodes for *how monitored* emerged as she recognized that a

Table I. First Coding System: October 1994

Cues	Monitor	Control	
	How monitored	What changed	How changed
	Positive	Teaching method	
	Negative	Evaluation method	

professor’s monitoring tended to be *positive* or *negative*. Three subcodes for *what changed* also emerged as she recognized that professors were making changes to *goals*, *teaching method*, or *evaluation method* as a result of their reflection.

This was the beginning of a codebook, which we needed to document our decisions and as a tool for communication among team members. Initial definitions and rules were developed and relevant examples from transcripts were documented in the codebook to help with consistency of coding. Terry did sample coding using these few codes, and tried to determine whether the codes would answer our questions. This phase took about two months.

The First Coding System: Team Feedback. Terry brought an initial draft of the first coding system (Table I) to a meeting of the entire research team in October 1994. Although we did not yet understand the data well, we wanted to determine if these codes were addressing the research questions and to assess the robustness of codes. Each team member attempted to work from the codebook to apply codes to selected segments of transcripts. After doing individual coding, the group reconvened to share individual results and experiences and check for level of agreement. Lively discussion ensued as we shared our perceptions, tried to come to a common understanding of the meaning of each code, and looked for evidence to support and challenge the suggested codes.

Generating the First Codebook: Terry Alone Again. Terry took our comments and went off to do further refinement and development of the codes and the first codebook. She also revisited the transcripts and began to see finer discriminations within the four codes that had survived the team feedback. Each of these was subsequently broken down into subcodes that provided finer distinctions of higher order codes as themes emerged. After these were integrated, a revised codebook was produced.

A Continuing Cycle of Tryout and Revision: Team Feedback. Each time a revised codebook was brought back to the team, we would test it by sample-coding portions of transcripts. Through this process of tryout and the ensuing social construction of team dialogue, new codes were continuously discovered. If agreement among team members was high, codes were considered acceptable; if not, then after each meeting Terry would reexamine the data and refine codes, their definitions, and the codebook for the next meeting. This process was repeated as each new category or set of subcodes was identified. Since we were also developing a team of three graduate student coders (Jacinthe, Cathy, and Carol), each of whom

Table II. Coding System in February 1995

Monitor		Control		Rationale
Cues—External	How monitored	What changed	How changed	
Student verbal	Positive	Goal	Add	Content knowledge
Student non-verbal	Negative	Method	Delete	Pedagogical
Student written	Neutral	Materials	Substitute	knowledge
Environment	Preclass	Evaluation	Maintain	Pedagogical content
Time	Concurrent	Content	Uncertain	knowledge
Cues—Internal	Postclass	Housekeeping	Move	Experiential knowledge
Feeling		Other	Other	Unknown
Experience				Other
Other				

Note: Unshaded areas = new codes. Shaded areas = codes previously recognized.

would be responsible for a different case, it was important to co-construct a shared conception of reflection, as represented in codes, that could be used consistently.

It took us about six months of meetings and hours of animated and laborious discussion, using a constant comparison approach to labeling phenomena and saturating categories with repeated supporting evidence (e.g., Charmaz 1990), until we felt our coding system was sufficiently robust to capture the phenomenon of reflection. As shown in Table II, by February 1995 the coding system had evolved. Several important concepts and codes had emerged from the process we were using to explore the data. For instance, we recognized two categories of cues that professors were monitoring during reflection: external cues (five types) and internal cues (three types). We found as well that professors provided a rationale for their reflection based on six types of knowledge they hold.

The Coding Begins: Jacinthe, Cathy, and Carol. Coding of transcripts began in the fall of 1995. Each coder took responsibility for analyzing the postclass transcripts of two of the six professors. Each became an expert in the idiosyncratic speech patterns of their professors, which led to questions about the interpretation of words, phrases, and passages. Coders tried to clarify these concerns through ongoing team discussions to avoid interpretations that might lead to miscoding the data. They developed a remarkable sensitivity to the data through the analytical process itself: As they analyzed, their understanding of professors increased, and as their sensitivity increased, so did their insight and understanding about the phenomenon of reflection (e.g., Strauss and Corbin 1990). This interaction resulted in the discovery and creation of codes (e.g., Charmaz 1990).

Table III presents the coding system as of December 1996, two years into the project. It displays crucial codes that the coders brought to our attention: they saw that the process of reflection revolves around two categories of goals that professors track, teaching goals (seven types) and learning goals (six types). Table III also displays a new level of codes which indicate three distinct spheres in which professors reflect. Most often their reflections focused on immediate issues in the

Table III. Coding System in December 1996

Time of reflection	Monitor		Goals		Control		Rationale
	Cues—external	How evaluated	Teaching goals		What changed	How changed	
			Objectives	Learning goals			
Non-stimulated concurrent	Student verbal	Positive	Objectives		Objective	Add	Content knowledge
Stimulated concurrent	Student non-verbal	Negative	Method		Performance	Delete	Pedagogical knowledge
Non-stimulated recall	Student written	Neutral	Materials		Method	Substitute	Pedagogical knowledge
	Student general	Mixed	Evaluation		Materials	Maintain	Pedagogical content
	Environment		Content		Evaluation of learning	Uncertain	content knowledge
	Time		Performance		Content	Move	knowledge
	Management		Housekeeping		Housekeeping	Pacing	Experimental knowledge
	Instructional materials		Learning goals		Other	Other	Student knowledge
	Other individuals		Student learning		Other	Long-term planning	Experience
	Research project		Student ability				Feeling
	Other		Student knowledge				Unknown
	Cues-Internal		Student participation				Other
	Feeling		Student positive affect				
	Experience		Other				

Note: Unshaded areas = new codes; Light shaded areas = codes previously recognized. Dark shaded areas = codes moved or modified.

classroom, which we called the *practical* sphere of reflection. Sometimes their reflection focused beyond the classroom, which we called *strategic* and *epistemic* reflection. At this stage, it can be seen that we are also refining codes and moving them around. For example, we once considered *feelings and experiences* to be *internal cues* for reflection; however, through our discussions we came to regard these as their *rationale*, or among the reasons they initiated reflection. Thus these codes have been moved.

While the front end of the study was framed by a priori theory as a way of initially understanding the phenomenon, through our dialectical process this evolved into a more grounded approach as other codes/concepts were discovered that further refined and explained these major constructs (e.g., Strauss and Corbin 1990).

Codes Continued to Evolve Over the Years: The Team, Participants, and Colleagues. During the next two years, the coding scheme and understanding of the phenomenon of reflection continued to evolve and expand through the ongoing discussions of the research team, and through presentations of our work to the participating professors and colleagues.

Within the team, we continued to confer until eventually few new or revised codes surfaced. The interconnectedness of the development and application of the coding system created some challenges. On one hand, openness to evolution of codes allowed team members and participants to feel ownership in the coding system and allowed a richer understanding of reflection to emerge. On the other hand, whenever our thinking about reflection or the codes resulted in revisions to the coding system, we had to do three things: first, communicate our thinking to other members of the team so that changes could be integrated into everyone's thinking; second, document the changes in our codebook so that the evolution would be captured; and third, go back and recode transcripts completed earlier.

At a two-day symposium in 1997 with participating professors, we presented our results and representation of reflection as a kind of member check (e.g., Guba and Lincoln 1989; Maxwell 1996) to ensure that our description of the process of reflection rang true for them. They raised questions and suggested several changes that made us rethink our codes and our thinking about reflection. However, since this verification was not done until after much coding had been completed and initial reports produced, we were not able to integrate their input into some of our reports. Since we found the interaction with participants to be particularly important for our learning (McAlpine et al. 1999c), in the future we would do member checks with participants much earlier in the development of the coding scheme. We believe this is relevant for any qualitative researcher, whether working independently or collaboratively.

We presented our work to colleagues in several venues: within our own department, at learned conferences, and at other institutions. Questions and comments from colleagues (e.g., McAlpine and Weston 1996; McAlpine et al. 1997; Weston

and McAlpine 1998) helped us see when codes were not clearly communicating the meaning of the data.

The team often wondered about the value of ceasing development to produce a final coding system that would be applied consistently across all transcripts. It seemed desirable to have firm categories to count and compare. Although a stable coding system would have facilitated analysis and reporting, it also would have cut short evolution of understanding of the phenomenon. As we continued to code, we gained new insights about the phenomenon of reflection, which would shed new light on the data which generated the insight. This insight, more often than not, required rethinking what had already been done. Thus, we have come to accept that the process of developing codes is never finished. Any presentation of our work or team discussion may result in an evolution of thinking about the phenomenon and the coding, and consequently transcripts may require recoding.

We have also come to understand that, when reporting, our goal is to provide a snapshot of the coding and results to provide a “best” representation of our thinking about the phenomenon at a particular time. Usually, by the time a report is finished—even before it is presented or published—our thinking has changed. In this article we share snapshots of several different moments to give a sense of the evolution that has occurred.

Structure of the Coding Scheme

Table IV presents the coding system in January 1998, three-and-one-half years into the project. By this time the codes were fairly stable, with just a few additions and changes since December 1996.

The product of our dialectical process was a hierarchical system comprising four tiers of codes. Tier 1 represents a superordinate level of codes that we came to recognize later in the analysis—three different spheres of reflection. Tiers 2–4 capture the general process of reflection on teaching, regardless of sphere. Tier 2 codes came primarily from the literature and represent the major constructs represented in the model of reflection: *monitor* and *control*, *goals* and *knowledge*. Tier 3 and 4 codes emerged from the data and further define what is happening in Tier 2, for example, that there are two aspects of monitoring (Tier 2): cues the professors are evaluating and how they are evaluating them (Tier 3). Further, cues (Tier 3) can be categorized in eleven subcodes (Tier 4). Definitions and examples of all codes are in the codebook, which is available upon request.

The Unit of Analysis

A task relevant for any qualitative researcher developing a coding system is to establish parameters for the unit of analysis. In our case, this was done

Table IV. Coding System in January 1998
Spheres of reflection: Practical; Strategic; Epistemic

Tier	Spheres of reflection: Practical; Strategic; Epistemic						
	Monitor		Goals		Control		
Tier 1	Time of reflection	External cues	How evaluated	Teaching goals	When changed	What changed	
Tier 2							
Tier 3							
Tier 4							
	Non-stimulated previous	Student verbal	Positive	Objectives	Previous	Objective	Content
	Stimulated previous	Student non-verbal	Negative	Method	Concurrent	Performance	knowledge
	Non-stimulated concurrent	Student written	Neutral	Materials	Post	Method	Pedagogical knowledge
	Stimulated concurrent	Student general environment	Mixed	Evaluation	Long-term plan	Materials	Pedagogical content
	Non-stimulated Recall	Time		Content		Evaluation of learning	knowledge
	Stimulated recall	Management		Performance		Housekeeping	knowledge of students
		Instructional materials		Housekeeping		Other	Experience
		Other individuals		Learning goals		Long-term Planning	Feeling
		Research project		Student learning		Other	Unknown
		Other		Student ability			Other
				Student knowledge			
				Student participation			
				Student affect			
				Other			

Note: Unshaded areas = new codes. Light shaded areas = codes previously recognized. Dark shaded areas = codes moved or modified.

as an integrated part of the dialectical process described. We address this issue separately because identifying and defining the unit of analysis was as difficult and time-consuming as developing the codes themselves.

We realized that not all text was pertinent, and wanted to concentrate on analyzing places in the transcripts where professors were demonstrating reflection. We wanted to find the phenomenon within the data, and markers of where the phenomenon began and ended. We spent a great deal of time considering a number of possibilities such as line segments, meaning units, and speech bursts. None of these seemed to capture in context an entire event of reflection. The *episode* of reflection was finally chosen as the unit of analysis. Definitions were tried out and improved until intercoder agreement on recognizing episodes was consistently high. Episodes ranged in length from one line to several pages. Within an episode we used the coding system to analyze by word and phrase, which allowed us to be attentive to possible meanings assumed or intended by the speaker (e.g., Strauss and Corbin 1990).

While the episode was a conceptually strong unit of analysis, it posed analytical problems. Episodes were coded simultaneously with four tiers of codes. When we tried to review or summarize particular codes, such as *student nonverbal cues*, entire episodes had to be retrieved and codes extracted from within the larger context. We thus learned to attach codes to words and phrases within episodes to facilitate retrieval.

VERIFYING THE CODING SYSTEM

We recognize that verification is an area where a split in qualitative traditions exists (e.g., Maxwell 1992; Mishler 1990). This split exists even within our research team. Had we each been working independently, we might not have had the same need for a rigorous approach to verification. The collaborative nature of team research requires greater attention to rigor, and it has been an area of concern and challenge for us. Recognizing the impact that assumptions can have on the analysis and interpretation of data, we designed in verification checks to try to minimize any inappropriate influence of context and assumptions on the coding system. We tried to respect our differences by using a variety of approaches—those that verify through the social construction of knowledge as well as those that might be called more quantitative procedures. We will discuss verification in terms of validity and reliability.

Validity

Our strategies to minimize threats to validity address most of the items included in Maxwell's (1996) validity checklist. We searched for discrepant evidence and negative cases constantly during the dialectical process of developing

codes. We used triangulation through collecting data from six professors teaching in different settings (large and small classes) in different areas (math and math educators) and two different universities. We solicited feedback about our research and coding from colleagues at our own university, at other universities, and at learned conferences. We used member checks in two ways with our participating professors to rule out the possibility of misinterpretation of their meaning. The transcripts were sent to the professors for verification before they were coded to ensure that their thinking was represented accurately. In addition, the symposium with the professors allowed them to respond to the outcomes of our work and suggest revisions to better represent their process of reflection. We collected "rich" data in the form of verbatim transcripts of all of our interviews with each professor, which provided sufficient information to test our developing theories. We used descriptive statistics, frequencies, to investigate the amount of evidence in our data to test and support our claims. We used comparisons between cases (professors) and between types of professors (mathematicians and math educators) to help us identify the crucial factors. And, we looked for internal generalizability (e.g., Maxwell 1996) of the model of reflection within the group of professors to ensure that what applied to one also applied to the others.

We also approached validation through a number of avenues of discourse suggested by Mishler (1990). The dialectical process and social construction of knowledge during our team meetings, as well as the coders' co-construction during analysis, were essential for developing our understanding of the phenomenon. Presentations of our work to groups of professors interested in reflection and to our colleagues in higher education were also important. Questions posed in these settings caused us to consider implicit assumptions that may have influenced the codes. For instance, one colleague asked if our coding could capture intensity as opposed to frequency of professors' attention to cues. This made us realize that, although we code the types of cues professors attend to, we do not code the weight, importance, or intensity of attention accorded to each type of cue. The data in the form of full transcripts and tapes are available to other researchers, as are the methods used to make links between data, findings, and interpretations.

Reliability

We realize that many qualitative researchers would fundamentally disagree with the notion of reliability. We used procedures to establish reliability or agreement among team members and among coders for several reasons. First, we saw it as a means for the team members, and multiple coders, to develop a shared perception of the phenomenon. A better term for this might be drawn from Maxwell's notion of interpretive validity (1992); reliability was our way of establishing a shared interpretive validity. We also saw it as a way to train coders and ensure that interpretation of codes was applied as consistently as possible across professors (cases).

During the process of developing the coding system, the team frequently did coding checks to develop a shared conception of reflection. This also allowed us to establish the robustness of a code across interview transcripts. Our approach was to apply the newly developed codes to ten percent of the transcripts from each of the six professors. This was followed by informal group discussion where we compared our coding. Where agreement was fairly high (e.g., eighty percent agreement among five coders) we often concurred that codes were acceptable. If agreement was low, then data were reexamined and codes refined.

During the actual coding of the transcripts, reliability checks were a means for ongoing conversations among the three coders to share their evolving thinking and build shared perceptions about the phenomenon. It also allowed for establishing consistency before analyzing and aggregating the coded interview data. To establish intercoder agreement, an average of five pages (or ten percent of an interview transcript's length) were selected at random from each participant's interviews and coded by all three coders. As the coding scheme included four different levels of codes, reliability was established for each level using the formula from Miles and Huberman (1994). This process was repeated throughout the period during which the coding occurred, each time analysis for a different set of transcripts or different participant was initiated.

Lincoln (1990a) talks about the letting go of an insistence on reality as a singular tangible entity. Some might see the notion of reliability as trying to establish a single reality. We see it rather as an approach for developing a shared understanding that can also establish consistency among coders. Just like the blind men and the elephant, although each holds a different piece, there is an elephant, a phenomenon (e.g., Maxwell 1990). So it is with the phenomenon we are studying. Although team members may feel or see different parts of it, it is by consolidating the different pieces that we can co-construct the elephant. We believe that verification through validity and reliability assists us in accomplishing this goal.

THE CODEBOOK

The development of the codes has been presented in a series of tables, however, without a codebook these codes are difficult to interpret and apply. A codebook is a tool for the development and evolution of a coding system and is an important means for documenting the codes and the procedures for applying them. We wanted our codebook to be clear enough to guide coders in marking the transcripts regardless of their experience and familiarity with context. However, the codebook does not stand alone; coders must be specifically trained to use it effectively.

Our codebook provides a detailed outline of the steps and rules for coding, definitions of each code, and representative examples of quotes from interviews. The first section describes a four-step procedure for coding: how to identify the unit of analysis (an episode), how to identify the type of reflection, how to code monitor

Table V. Example from Codebook: Monitoring-Type of Evaluation Statements

Type of Evaluation Statement: The nature of the professor's evaluation of the teaching and learning		
Code	Definition	Key words or phrases from transcripts
POS	Positive. Any comment indicating evaluation of any aspect of the teaching and learning which is explicitly stated as positive.	Key words: good, well, okay Positive descriptors: relaxed, happy "I thought they were pretty happy."
NEG	Negative. Any comment indicating evaluation of any negative aspect of the teaching and learning which is explicitly stated as negative.	Key words: bad, terrible "I'm a bad model: I lose stuff." "I felt rushed." "I have too much stuff here." "... not as good as I wanted."
N	Neutral. Neutral comments that are neither positive nor negative. This includes descriptive comments about the teaching and learning that do not include an evaluation.	"This is exactly what I planned to say."
MX	Mixed. The evaluation keeps going back and forth. It could be positive, then negative, and then back to positive. It could be presenting both sides and then balancing the good and bad.	"There were good parts and bad parts." "It went okay. But then I started to talk too much. But it was better than the last time."

statements, and how to code control statements. The second section is a complete list of all the codes, each accompanied by a definition and sample words, phrases, or linguistic markers drawn from the interview transcripts. Table V presents an excerpt from the codebook, where the four types of evaluation statements are defined.

The dialectical process which began in the fall of 1994 continued through seven versions of the codebook and continues to this day. In writing this article we closely examined the most recent version and saw that our current thinking had not yet been incorporated. Its outdatedness was explained by one of the coders, who said that at a certain point they had internalized the codes so much that the codebook seemed almost "a document for others, not us." As time went on, the team used the codebook less as a guide and more as a frame of reference, and thus gave less attention to updating the codebook.

CONCLUSIONS

In this article we have told the story of the development and evolution of a coding scheme which may be useful for researchers who use coding to analyze interview transcripts. We think there are several lessons that emerge from our experience.

First, coding is not what happens before analysis, but comes to constitute an important part of the analysis. There is a reciprocal relationship between the development of a coding system and the evolution of understanding a phenomenon. As the codes and codebook evolved, our understanding of the phenomenon was also transformed. This evolution became evident in representations of our model of reflection, which matured as a direct result of the increased understanding that emerged through the analysis of transcripts and development of codes. The model recontextualizes the coding system as a mechanism for coming to understand the phenomenon of the process of reflection on teaching. We liken the process to continually zooming in and out. One begins with the big picture, an overall conception of the phenomenon, moves in to focus on details through coding, and moves out again to see how the details might have changed the way we interpret the larger picture. Thus, the development of a coding system is a critical analysis tool in that it leads to an ongoing evolution in understanding the phenomenon. Unlike grounded theory, we did not use emerging categories to drive further data collection, but they had a significant role in the development of our coding system and in our understanding of the phenomenon of reflection, and have influenced approaches to analysis and coding on subsequent research projects.

Second, a team builds codes and coding builds a team through the creation of shared interpretation and understanding of the phenomenon being studied. The process is long and the experiences of individual team members, such as coding, giving presentations, and writing papers, contribute to different progressions in thinking. It is important to find mechanisms for documenting and sharing thinking. We found that revision of the codebook and the model were effective ways of accomplishing this.

We wish we had been more assiduous in documenting successive versions of the codebook. Keeping the codebook up to date through reviewing and revising is an important mechanism for sharing evolving understanding and keeping team members informed of changes in thinking about the phenomenon. Appropriate times for us to make revisions would have been after reliability checks among coders or after presentations of our work, both of which usually resulted in revision. Often, when we got around to making the necessary revisions, our thinking was ahead of the codebook, so team members were holding different conceptions of the coding and the phenomenon. Also, the codebook is a means of preserving the collective memory of the team. We now see it as an important artifact for reminding us of how our thinking about the phenomenon has evolved. Right now we have some gaps in our history where we cannot recapture how and why a change occurred. For example, documenting where codes originated and where names came from (e.g., Conostas 1992) would have been an excellent way of creating a record of the many decisions made over several years. Finally, the codebook continues to serve as an essential tool for training new coders joining the team. When we recently updated it for new coders, we realized how many of the procedures for applying the codes

had become shared tacit knowledge. Team members spent hours reconstructing our thinking to make these steps explicit.

In the end, the team has a sense of shared authenticity with the results from the phenomenon being repeatedly challenged and tested. When individual team members speak of the research, they feel ownership and are comfortable with the conclusions. Through this process we learned to be ready to modify our thinking based on new feedback; with so many team members there were always new ideas. Consequently, when we share our work, we see opportunities for new perspectives in an ongoing process.

Third, collaboration in qualitative research requires a kind of rigor that a lone or independent researcher might not be aware of or need. Our process and concerns may be revealing for qualitative researchers planning to move toward collaborative work. In our research team, individuals came from different research traditions, so we were intentional in respecting differences by incorporating a variety of approaches in our research design that would support individual comfort levels. Further, collaborative team research requires building a shared conception of a phenomenon. A critical tool for us in developing this shared conception were the continuing conversations about our thinking provided by regular reliability or intercoder agreement checks. These conversations and negotiations of interpretation were essential prerequisites to developing a codebook that could be understood and used by all. Overall, the use of multiple approaches to verification not only supported the differences among team members and enhanced the evolution of team understanding of the phenomenon, but also increased the acceptability of results within our discourse community whose members come from diverse traditions.

As noted earlier, Lincoln (1990b) suggests that combining methods may not result in a coherent set of findings. Alternately, Maxwell suggests that it is important to explore combinations of methods to understand what works and whether inconsistencies between approaches lead, for example, to inappropriate design decisions or invalid conclusions. Our experience may serve as an example of a qualitative study that has used a combination of methods. We would say that using a combination of methods strengthened both our process and results. The power of a tool of analysis derived in this manner is that now the emergent codes are becoming the *a priori* framework for our further work. As our understanding of the phenomenon becomes increasingly rich, we see new research questions derived from the analysis which will further refine our understanding and representation of the phenomenon. Without the codebook we feel impotent, since it and the model provide the conceptual frame that drive the research questions we are now pursuing.

ACKNOWLEDGMENTS

This work is supported in part by a grant from the Social Sciences and Humanities Research Council of Canada. The authors also wish to thank the editor and the reviewers for their contributions to this manuscript.

REFERENCES

- Bereiter, C., and Scardamalia, M. (1989). Intentional learning as goal of instruction. In L. B. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser* (pp. 361–392). Hillsdale, NJ: Lawrence Erlbaum.
- Berliner, D. C. (1992). Telling the stories of educational psychology. *Educational psychologist*, 27, 143–161.
- Bracewell, R. J., and Breuleux, A. (1994). Substance and romance in the analysis of think-aloud protocols. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 55–88). Newbury Park, CA: Sage Publishing.
- Breuleux, A. (1991). The analysis of writers' think-aloud protocols: Developing a principled coding scheme for ill-structured tasks. In G. Denhière & J.-P. Rossi (Eds.), *Texts and text processing* (pp. 333–362). Amsterdam: North-Holland.
- Charmaz, K. (1990). "Discovering" chronic illness: Using grounded theory. *Social Science and Medicine*, 30, 1161–1172.
- Constas, M. A. (1992). Qualitative analysis as a public event: The documentation of category development procedures. *American Education Research Journal*, 29, 253–266.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks: Sage.
- Eisner, E. W. (1998). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. Upper Saddle River, NJ: Prentice-Hall.
- Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.) (pp. 119–161). New York: Macmillan.
- Frederiksen, C. H. (1986). Cognitive models of discourse analysis. In C. R. Cooper & S. Greenbaum (Eds.), *Written communication annual: Vol. 1. Linguistic approaches to the study of written discourse* (pp. 227–267). Beverly Hills, CA: Sage Publishing.
- Guba, E. G., & Lincoln, Y. S. (1989). *Effective evaluation: Improving usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco: Jossey-Bass.
- Howe, K., & Eisenhart, M. (1990). Standards for qualitative (and quantitative) research: A prolegomenon. *Educational Researcher*, 19, 2–9.
- Kirk, J., & Miller, M. L. (1986). *Reliability and validity in qualitative research*. Beverly Hills, CA: Sage.
- Lincoln, Y. S. (1990a). Campbell's retrospective and a constructivist's perspective. Review of D. T. Campbell, *Review of Methodology and Epistemology for Social Science*. In J. A. Maxwell & Y. S. Lincoln, Methodology and epistemology: A dialogue. *Harvard Educational Review*, 60, 501–504.
- Lincoln, Y. S. (1990b). Response to "Up from positivism." In J. A. Maxwell & Y. S. Lincoln, Methodology and epistemology: A dialogue. *Harvard Educational Review*, 60, 508–512.
- Maxwell, J. A. (1990). Response to "Campbell's retrospective and a constructivist's perspective." In J. A. Maxwell & Y. S. Lincoln, Methodology and epistemology: A dialogue. *Harvard Educational Review*, 60, 504–508.
- Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62, 279–300.
- Maxwell, J. A. (1996). *Qualitative research design: An interactive approach*. Thousand Oaks: Sage.
- McAlpine, L., & Weston, C. (1996). Researching reflection: A metacognitive model. American Educational Research Association, New York, April.
- McAlpine, L., Weston, C., Beauchamp, J., Beauchamp, C., & Wiseman, C. (1997). Building a model of reflection on teaching: New findings and insights. American Educational Research Association, Chicago, March.
- McAlpine, L., Weston, C., Beauchamp, J., Wiseman, C., & Beauchamp, C. (1999a). Building a metacognitive model of reflection. *Higher Education*, 37, 105–131.
- McAlpine, L., Weston, C., Beauchamp, C., Wiseman, C., & Beauchamp, J. (1999b). Monitoring student cues: Tracking student behavior to improve instruction in higher education. *Canadian Journal of Higher Education*, 29–2, 113–144.
- McAlpine, L., Weston, C., Beauchamp, C., Wiseman, C., & Beauchamp, J. (1999c). Reciprocity in the research process: Learning opportunities for both participants and researchers. American Education Research Association, Montreal, April.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks: Sage.

- Miller, S. M., Nelson, M. W., & Moore, M. (1998). Caught in the paradigm gap: Qualitative researchers' lived experience and the politics of epistemology. *American Educational Research Journal*, 35, 377–416.
- Mishler, E. G. (1990). Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard Educational Review*, 60, 415–442.
- Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory, procedures and techniques*. Newbury Park, CA: Sage.
- Webb, G. (1997). Deconstructing deep and surface: Towards a critique of phenomenography. *Higher Education*, 33, 195–212.
- Weston, C., LeMaistre, C., McAlpine, L., & Bordonaro, T. (1997). The influence of participants in formative evaluation on the improvement of learning from instructional materials. *Instructional Science*, 25, 369–386.
- Weston, C., & McAlpine, L. (1998). Improving university teaching: Understanding and applying a cognitive model of reflection. American Educational Research Association, San Diego, April.