CHAPTER THREE

Methodology

This chapter discusses the procedures for gathering and analyzing the data. The first section tells *who* the data was collected from, providing a more general description of the research design as well a rationale for *why* methodological components were chosen. The second section offers more detail about the data gathering, including *what* data was collected and *how*. The third section illustrates the specific way that data was analyzed to create a richer framework for understanding EPSTs' conceptions of variation.

Research Design

Subjects were chosen from a section of a course I'll refer to as *Mathematics for Elementary Teachers 2* (MET 2). The course was the second of a two-course sequence at a University located in a metropolitan area of the Pacific Northwest. Intended for prospective teachers, the two math courses (MET 1 and 2) are required for those wanting to enroll in the Graduate Teacher Education Program (GTEP). Completion of GTEP leads to issuance of an Initial Teaching License. To give a sense of the environment common to the subjects, a general overview of the content and pedagogy for the typical MET 2 course is next described. Within the context of the typical MET 2 course, specific features of the particular section used for my research are articulated.

Content and Pedagogy of MET 2

The content for MET 2 includes geometry, probability, and statistics. Instructors usually divide the ten-week course into two parts of roughly equal length, with one part dealing with geometry and the other part dealing with stochastics. The probability component includes single-stage and multistage experiments, and students investigate both theoretical and experimental probabilities. Theoretical probabilities are obtained by deriving the sample space and considering the possible outcomes of a specific event, while experimental probabilities are obtained by simulation. Students consider disjoint events as well as independent and dependent events. The statistics component includes descriptive measures of central tendency (mean, median, and mode) and spread (range, interquartile range, and standard deviation). Students also analyze data using a variety of graphs, such as boxplots, line plots, bar charts, histograms, pie graphs, and scatterplots. Themes of sampling, such as random sampling, stratified sampling, and making predictions based on sampling, are also a part of the statistics component.

The pedagogy for MET 2 varies with the instructor, but there are some common pedagogical themes. One theme is that students participate in activities, both as a class and in smaller groups. The MET 2 classroom is arranged so that students sit in groups, which can be as large as six students per group. Once a problem has been posed, or an activity given, students typically will work singly, then together in their groups, and finally share ideas with each other in a class-wide discussion. The sharing of ideas points to a second common theme, which is communication. Students are expected to communicate what they are thinking about and to ask questions of one another so as to understand each other's reasoning. A third common theme is that the teacher acts more as an inquiring guide than as a lecturer. Ideally, the teacher facilitates discussion by asking questions, encouraging other students to ask questions, and generally guiding the class to consensus where possible.

Multiple sections of MET 2 are typically offered during any given quarter. During the Winter Quarter when my research took place, classes were held two days a week for ten weeks. Each class session was scheduled to last two hours and fifty minutes. The section I chose to use for my research was led by Steve, an experienced instructor for MET 1, 2 and other courses for teachers and preservice teachers of mathematics. Steve's plan for the curriculum reflected the components described earlier, but he modified his previous practice of doing all the geometry in the first part of the course, and all the stochastics in the second part. Instead, he devoted the first four weeks to geometry, and then gave the next four weeks to stochastics, followed by a last installment of geometry. The modified schedule was done to help accommodate my research plan.

One reason I chose Steve's section is because of Steve's skill in modeling the pedagogical themes mentioned earlier. He lectures less and he poses problems, guides activities, and facilitates discussions. A second reason for choosing his section is because Steve and I had worked well together as colleagues for more than four years, and our philosophies of teaching and learning were similar. The familiarity with each other's styles and congruence of philosophies helped make working together easy, particularly when we were co-teaching some of the lessons. A third reason is because Steve was willing to modify the sequence of geometry and probability and statistics so that my out-of-class interviews could take place before and after the stochastics portion of the class sessions.

Student Characteristics

The urban setting of the University fosters a wide range of student backgrounds, and it is difficult to describe typical demographic characteristics of the students who take MET 2. Past students in my classes have ranged in age from the early twenties to the middle fifties. Some are undergraduates and some are graduate students. Some want to become teachers and occasionally some are just filling a University math requirement.

The majority of MET 2 students have taken MET 1 at the University, a course taught in a similar style to MET 2. The content of MET 1 includes whole-number arithmetic, number theory, fractions, decimals, and ratios. Other than MET 1, most students have taken few, if any, post-secondary math classes. It may have been as many as twenty or more years since they have had *any* mathematics class. At the outset of MET 1 students often write a "mathography," which entails a description of their past math classes and their feelings about past math experiences. Most students describe themselves as not having been very good at math in the past, and most MET 2 students share negative memories of precollege mathematics. Some students say they feared or hated their math classes, and others say they were bored. The attitude of most beginning MET 1 students is that mathematics is a rule-oriented discipline. The role of the teacher is to reveal the rules, and the role of the students is to memorize and apply the rules.

The MET 1 experience helps to give most students a new and different vision of what “doing mathematics” entails. Because the pedagogy in MET 1 is similar to that in MET 2, most students come to expect a learning environment in MET 2 where they will be active participants and where their thinking strategies are validated. They have been enculturated to the process of problem-solving and communicating their reasoning. Although not all MET 2 students are comfortable with the learning environment, those who have come through MET 1 at least know what to expect.

A total of thirty students completed Steve’s section of the MET 2 course. A profile of the background for these students is offered with respect to the following attributes: Class level at the university, gender, when and where the students had taken MET 1, and any prior probability or statistics courses taken. These attributes are summarized in Table 1.

Table 1. *MET 2 Class Profile*

Attribute	Categories of Attribute	Number of Students	Total Students
[1] Class Level at the University	Undergraduate	9	30
	Graduate	21	
[2] Gender of the Student	Male	6	30
	Female	24	
[3] Where MET 1 was Taken	At PSU	26	30
	Not at PSU	4	
[4] When MET 1 was Taken	Within Last Year	24	27
	More Than a Year Ago	3	
[5] Any Prior Probability or Statistics Courses	Yes	12	27
	No or Unsure	15	

The student total in Table 1 is not always thirty because not all data could be gathered for every student. For example, information on the first three attributes was gleaned from the university's information system, and the last two attributes were informed by the PreSurvey which was completed by 27 students during the first week of the course. Appendix B shows the entire PreSurvey instrument.

Almost two-thirds of the class (19 out of 30 students) were continuing on with the same instructor, Steve, for consecutive quarters. This fact is noteworthy because it helps explain the general disposition of the class with respect to their attitudes about studying mathematics. Of the twelve students who could recall having had prior formal education in probability or statistics, eight expressed favourable attitudes on the PreSurvey when asked "How did you feel about Probability and Statistics at that time?" For example, one student wrote that it had been an "interesting class, while another commented that it had been "fairly easy to understand." The four other students with prior formal exposure expressed unfavourable attitudes, such as: "It was my least favorite class in all of college."

When asked how they felt about learning probability and statistics now, only 5 of the 27 respondents expressed explicitly negative thoughts, such as LT, who wrote: "I feel very scared. It makes me nervous." Sixteen students put something that was explicitly positive, such as

DM: "Very excited, looking forward to it."
CS: "Pretty comfortable, ready and excited"
EM: "I'm interested to learn more"

The remaining 6 students had responses that were somewhat neutral in character, as the comments by GP illustrate: “I’m open to it, but not really excited.” It seems reasonable to assume that, since 19 students did have Steve during the previous quarter, there would be some influence on the students’ attitudes and expectations. As one such student, MM, put it, “I’m excited about this class because I enjoyed [MET 1].”

Overview of Research Design

The two main sources of data for my study were written instruments given to everyone in Steve’s class and individual interviews conducted with six students selected from the class as my case studies. A third source of information was class observations and videotapes made during the stochastics portion of the MET 2 course to help record the learning environment that the students experienced.

Table 2 summarizes the overall research design, the type of information gathered, who it was gathered from, and when it was gathered during the 10-week quarter. The table also shows the contexts (sampling, probability, or data and graphs) emphasized by each instrument or activity.

Prior to doing any class activities in stochastics, during the first week of the academic quarter, baseline information was collected from the MET 2 students in Steve’s section. The information was collected via a written survey that addressed prior mathematical experience and contained a range of questions about probability and statistics. Over the next couple of weeks, while Steve taught the geometry portion of the MET 2 course, I interviewed six students outside of regular class time. The interviews were videotaped and lasted about 45 minutes on average. The initial survey and interview were named PreSurvey and PreInterview because they occurred prior to the stochastics portion of the MET 2 course.

Table 2. *Overall Research Design*

When	What	Contexts	Type of Data	Who
1st week	PreSurvey	All	Written (In class)	Classwide (n=27)
2nd & 3rd week	PreInterview	All	Videotaped (Out of class)	Six cases (n=6)
5th & 6th week	Class Intervention #1 (Four Questions & Body Measurements)	Data & Graphs & Sampling	Videotaped (In class)	Classwide (Varied)
6th week	PostSurvey #1	Data & Graphs	Written (Out of class)	Classwide (n=28)
7th week	Class Intervention #2 (Known & Unknown Mixtures)	Sampling & Graphs	Videotaped (In class)	Classwide (Varied)
7th week	PostSurvey #2	Sampling & Graphs	Written (Out of class)	Classwide (n=30)
6th & 8th week	Class Intervention #3 (Cereal Boxes & River Crossing Game)	Probability & Graphs	Videotaped (In class)	Classwide (Varied)
8th week	PostSurvey #3	Probability & Graphs	Written (Out of class)	Classwide (n=29)
9th & 10th week	PostInterview	All	Videotaped (Out of class)	Six cases (n=6)

In the 5th week of the course, Steve made the transition from geometry to statistics and probability, and I began attending each class session. In addition to making observations, I also videotaped portions of class activities and discussions. Of the many class activities that took place over weeks 5 through 8, six activities were designed as interventions about variation. There were two activities in each of three interventions, with one intervention focused on each of the contexts, data and graphs, sampling, and probability. The three class interventions are listed in Table 2, and are further described later in this chapter.

After each one of the three interventions had been conducted in class, a take-home assignment was given, which I have called PostSurveys because they occurred after the entire intervention for that context had been completed. The PostSurveys are differentiated according to their context: For instance, the first PostSurvey corresponded to the context of data and graphs.

After the last intervention had happened in class, and the PostSurvey (Probability) had been administered, a second interview was conducted with the same six students as those interviewed for the PreInterview. The second interview was called a PostInterview because it took place after the interventions occurred in class. Like the PreInterviews, the PostInterviews were videotaped and lasted about 45 minutes on average. All PostInterviews occurred in the last two weeks of the quarter.

I used all the data collected to help inform my research questions. The written documents, observations, and interviews formed a corpus of data that I analyzed with grounded theory techniques to describe components of a conceptual framework characterizing EPSTs' thinking about variation. I also used the entire corpus of data to consider comparisons of EPST's thinking before and after the instructional interventions and to consider which tasks were most useful in examining conceptions of variation. In presenting my results, I used a case study format to profile the thinking of the six cases who participated in the interviews.

Rationale for Design

This is a qualitative study that combines two traditions of qualitative inquiry: grounded theory and case studies. I'll justify the choice of these two traditions by describing how they helped answer my research questions, and then articulate the three types of data collected: written documents, observations, and interviews.

Grounded Theory: Strauss and Corbin (1994) said that "grounded theory is a *general methodology* for developing theory that is grounded in data systematically gathered and analyzed" (p. 273, italics in original). I used three techniques from grounded theory - open coding, axial coding, and constant comparison – to theorize an "evolving framework" for characterizing EPSTs' thinking about variation. This evolving framework, which addresses my first research question about finding components of a conceptual framework, is described in the next chapter. I'll next offer a brief description of the techniques and terminology of grounded theory.

Grounded theory begins with *open coding*, a process of describing and building categories of similar phenomena, the dimensions of which are defined by their conceptual properties. Adding to the power of open coding is *axial coding,* defined as "the process of relating categories to their subcategories [or properties], termed 'axial' because coding occurs around the axis of the category, linking categories at the level of properties and dimensions" (Strauss & Corbin, 1998, p. 123). Microanalysis is the combined approach of open and axial coding, often using a line-by-line analysis, to "generate initial categories (with their properties and dimensions) and to suggest relationships among categories" (p. 57). Emergent tentative hypotheses suggest links between categories and properties (Patton, 2001; Merriam, 1998). As data are iteratively compared to the emerging categories, the categories themselves are refined in light of reviewing the data (Strauss & Corbin, 1994). "This process of taking information from data collection and comparing it to emerging categories is called the *constant comparative* method of data analysis" (Creswell, 1998, p. 57, italics in original). Patton (2001) calls this comparative analysis "a central feature of grounded theory development" (p. 490). Later in this chapter, I'll demonstrate how I used the techniques of grounded theory to derive the evolving framework.

Case Studies: Case studies are often associated with ethnographies, grounded theories, or exploratory research. Creswell (1998) defines "a *case study* [as] an exploration of a 'bounded system' or case (or multiple cases) over time through detailed, in-depth data collection involving multiple sources of information rich in context" (p. 61). Stake (1994) agrees that a case study should represent "a specific, unique, bounded system" (p. 237), and the boundaries can be defined by time and place. The MET 2 course, conducted in the same location over ten weeks with the same students and instructor, represented the kind of bounded system needed to conduct a case study. I chose six students from the class to serve as my cases (how I chose them is described in the next section), and I used the evolving framework to compare their conceptions of variation from before to after the instructional interventions. Thus, case studies helped me address my second research question about comparing EPST's conceptions over the duration of the research.

There are two specific reasons why case studies worked well together with a grounded theory approach for the purposes of answering my research questions. One reason is because case studies, like grounded theory, allow theory to be generated via descriptive data. Case studies encourage the use of descriptive data "to develop conceptual categories or to illustrate, support, or challenge theoretical assumptions held prior to the data gathering" (Merriam, 1998, p. 38). McMillan and Schumacher (1997) wrote that "case studies are appropriate for exploratory and discovery-oriented research" (p. 395). A second reason is because both case studies and grounded theory encourage the use of multiple sources of data (Stake, 1995; Strauss & Corbin, 1998). The use of different data sources is referred to as *triangulation*, which serves to "clarify meaning by identifying different ways the phenomenon is being seen" (Stake, 1994, p. 241). Triangulation strengthened my research by letting me get information in several forms and at different times, so that my findings were "consistent with the data collected" (Merriam, 1998, p. 206). My study was triangulated by three methods of data gathering: written documents to review (via the surveys), classroom observations during the instructional interventions, and individual interviews with my cases.

Written Documents: The use of written documents as a method for collecting data in a case study is well regarded (Stake, 1995; Merriam, 1998; Patton, 2001). Written documents can supplement observational data, and "quite often, documents serve as substitutes for records of activity that the researcher could not observe directly" (Stake, 1995 p. 68). In my study, there were two main types of written documents I collected: pre-activity documents (the PreSurvey) and post-activity documents (the PostSurveys for Sampling, Data & Graphs, and Probability). All of the survey instruments are listed in Appendix B. I also collected in-class work that came out of the three interventions. For example, when we began the class intervention on sampling, students wrote about what they thought a "sample" was, who they thought used samples, and why they thought taking samples might be useful. Students also wrote down initial predictions for what they thought would result from thirty samples each of size ten taken from the Known Mixture. Small groups of students produced posters for actual results from the Known and Unknown Mixture activities, and I saved or photographed all the posters.

Observations: The purpose of using observations as a data collection method was to record the overall class contexts in which the activities occurred. I wanted not only to capture the contributions of my six cases, but also to hear what the ideas the rest of the class shared within their small groups and in classwide discussions as they engaged in the activities.

Observations are a common data collection technique in case studies (Stake, 1995; Merriam, 1998). Best and Kahn (1998) write that "when observation is used in qualitative research, it usually consists of *detailed notation* of behaviors, events, and the contexts surrounding the events and behaviors" (p. 253, italics in original). Patton (2001) includes the following three dimensions of concern when conducting observations: the role of the observer, the disclosure of observation, and the recording procedures.

In my research, I had roles as both participant and observer. I was a participant by virtue of co-directing some of the MET 2 activities, and an observer by virtue of recording the class activities. Patton claims that "the participant observer employs multiple and overlapping data collection strategies, being fully engaged in experiencing the setting (participation) while at the same time observing and talking with other participants about whatever is happening" (pp. 265, 266). Regarding disclosure, the participant observation is what Fraenkel and Wallen (2000) called *overt*, because the researcher will be identified and the cases will know they are being observed. For recording procedures, I videotaped the three class interventions with the help of a colleague, Matt. I also videotaped parts of all the other class sessions having to do with probability and statistics, and took notes after each session ended. To "minimize the errors resulting from faulty memory," caution Best and Kahn (1998), "simultaneous recording of observations is recommended" (p. 295). Videorecording of the classroom during the activities showed the context of the learning environment, the overall flow of the class, and the specific contributions of my cases to the class discussion. At multiple times during the three class interventions, the specific tables where my cases sat were videotaped, so that I could capture what they were saying to each other in small-group discussions during the activities.

Notes from my observations were added to the corpus of data that was used to help shape the evolving framework, compare EPSTs' conceptions before and after the interventions, and inform which tasks were useful in examining EPSTs conceptions of variation. Thus, observational data helped supplement my thinking about all three research questions.

Interviews: A semi-structured, task-based interview format was the third method used for gathering data. Interviewing is a common and powerful method of trying to understand how other people think (Fontana & Frey, 1994). Best and Kahn (1995) note that "interviews are used to gather information regarding an individual's experience and knowledge" (p. 255), and Patton (2001) says that the purpose of interviewing "is to allow us to enter into the other person's perspective. Qualitative interviewing begins with the assumption that the perspective of others is meaningful, knowable, and able to be made explicit" (p. 341). By semi-structured, I mean that the interviews were scripted at the outset, but my protocol allowed for a variety of probes depending on the responses of the responses of the interviewees. By task-based, I mean that the subjects were not interacting merely with me as an interviewer, but "with the task environments" (Goldin, 2000, p. 519).

Goldin (2000) goes on to mention the value in task-based interviews, noting that the tasks can be adjusted in wording and content according to the results of previous research. He adds, "Interview contingencies can be decided explicitly and modified when appropriate. In comparison with paper-and-pencil test-based methods, task-based interviews make it possible to focus research attention more directly on the subjects' processes of addressing mathematical tasks" (p. 520).

I conducted the first round of interviews, the PreInterviews, after the PreSurveys had been collected and reviewed but before any of the class sessions in stochastics had begun. The second round of interviews, the PostInterviews, took place during the last two weeks of class, after the three interventions had taken place. For each round of interviews, there were some tasks given which were identical or to tasks given on the survey instruments. Also, the tasks for the PostInterview reflected themes that had been explored in the class sessions. The interviews were all videotaped so that the subjects' explanations and nonverbal communication was recorded. Some of the cases wrote on their copies of the interview scripts, and their written notes were collected when relevant. Their written notes and observations that I made during the interviews became a part of the overall data for the study, along with the transcriptions of the interviews.

In summary, this research was designed to gather data from multiple sources. All of the data was used to build my evolving framework using grounded theory techniques. The entire framework is explicated as the first part of my results in Chapter Four. Although I did use all the data in comparing EPSTs' conceptions, I focused on data from six cases to exemplify some comparisons of EPST's conceptions. I present comparisons of the six cases as the second part of my results in the next chapter, further organizing the presentation around the tasks that were most illustrative for the research. In the next section, I'll describe specific details of the data gathering.

Data Gathering

This section is organized chronologically. The PreSurvey was administered before the PreInterviews were conducted. Then, each of the three class interventions was followed by a corresponding PostSurvey as shown earlier in Table 2. Finally, PostInterviews were conducted. All the instruments are found in Appendix B

PreSurvey

The PreSurvey was given during the second class session of the first week of the quarter, and was completed by 27 students. The students had known that a survey was to be administered, because I had visited the first class session and told them. Also during the first class session, I described the research project and their opportunity to be involved, and distributed the informed consent forms found in Appendix A. All students were willing to have their written work included as part of the collected data, and all eventually gave consent to be videotaped during the class sessions. I also had eleven students volunteer to be interviewed outside of class, so I used the PreSurvey responses to help determine who would make up my final six cases (described further in the next chapter). The average time for completion of the PreSurvey was about 45 minutes. The structure of the nine-page PreSurvey had two parts: The first part was the first page, containing background questions to determine the attributes shared earlier in Table 1, such as what prior experiences in probability and statistics they could recall having. Also on the first page were questions about the meaning of the terms "random" and "variation." The second part of the PreSurvey (comprising the other eight pages) held a total of nine questions, many of which had multiple parts. The specific questions and classwide responses for the PreSurvey are summarized in the next chapter, but the contexts for the questions in second part of the PreSurvey are given in Table 3.

Aside from the background questions, all the rest of the questions on the

PreSurvey were either identical to or very similar to questions asked by other researchers of middle and high school students. In particular, a NSF-funded project (Shaughnessy, 2003) used written surveys with 12 different classes of middle and high school students, and many questions on the PreSurvey came from the NSF surveys, which in turn had been motivated by prior research. For example, on the PreSurvey and the NSF high school survey, two questions asked students for a description of what the terms "random" and "variation" meant to them, and those questions were similar to the ones used by other researchers (e.g., Watson, et. al., 2002).

Table 3. *Contexts for PreSurvey (Part Two) Questions*

Question	Brief Description	Contexts
1	Results are predicted for drawing one, several, and six samples of 10 candies from a jar (60 Red & 40 Yellow)	Sampling
2	Ranges are predicted for drawing six and then thirty samples of candies	Sampling
3	Results are predicted for drawing fifty samples of candies	Sampling
4	A graph is made to show what the results of fifty samples might look like	Sampling & Graphs
5	Test scores are shown for two different classes. The question addresses the two classes' relative performance	Data & Graphs
6	Two graphs, showing the student heights at two different school, are compared to see which shows more variability	Data & Graphs
7	Results are predicted for performing one, two, and six trials of 50 flips of a fair coin.	Probability
8	Chances of winning a game involving two 50-50 (Black-White) spinners are addressed.	Probability
9	Chances of winning a game involving a 50-50 and a 25-75 (Black-White) spinner are addressed.	Probability

A question involving the comparison of graphs used in the PreSurvey (Question #5) also was used by Watson and Moritz (1999). Question #1, using samples of candies from a jar (akin to the Candy Task mentioned in Chapter 2), was used on the NSF surveys as well as in other research (Torok & Watson, 2000; Reading & Shaughnessy, 2000; Shaughnessy et. al., 1999; Shaughnessy & Ciancetta, 2001). Thus, for the PreSurvey, I chose to use questions that have been used by other researchers to obtain an overview of middle and secondary students' thinking about variation.

PreInterview

After the class in which the PreSurvey was completed, I read through all the responses to get a general sense of how much the students had written and what they had to say. I paid particular attention to the surveys from the eleven students who had said they were willing to be interviewed. Of these eleven students, I selected six who I predicted would make good cases because I thought their responses on the PreSurvey were representative of the whole class in the sense that their responses were similar to those of other students in the class. Also, my six cases provided enough written detail to convince me that they would have no problem sharing their thoughts in an interview situation.

Having a total of six cases is largely a pragmatic decision, determined by the resources available in conducting my research. Qualitative inquiry generally involves relatively small samples, and the key in purposeful sampling is to select "information-rich cases whose study will illuminate the questions under study" (Patton, 2002, p. 230). Creswell (1998) discusses the value in purposeful sampling, saying that cases can be chosen to provide different perspectives on the phenomena under study.

Because my research design required pairs of Pre and PostInterviews for the same subjects, I conducted the PreInterview of all eleven of the volunteers over the second and third weeks of the quarter. I interviewed all eleven because although I was planning on six cases, I could not guarantee that those six would still be enrolled or be available at the end of the quarter for the PostInterview, nor could I guarantee that the quality of their information in the interview setting would be as rich as I had expected it to be. It turned out that the six cases I initially had in mind from reading their PreSurveys were in fact very useful for informing my framework and helping me understand their conceptions of variation, and they did complete the postinterviews as well. The other five of the eleven volunteers also completed postinterviews, which I conducted because I had the resources and figured that their contributions gave me additional data with which to work on future research.

The PreInterview subjects met with me outside of the regular classtime. A videocamera was set up to record the interview, and a separate tape recorder also was present because the audio cassettes were useful in transcribing the dialogue. I had a copy of the interview script (see Appendix B) and also gave a copy to the subject, who was encouraged to write on the script as desired. The PreInterview contained 13 multi-part questions. The contexts for the questions are given in Table 4. More details of the PreInterview questions and sample responses are provided in the next chapter.

The first question on the PreInterview was identical to the first question on the PreSurvey (akin to the Candy Task) for two reasons. First, I wanted to see how their verbal responses compared with what they had written, and to see if they had anything more to add to their earlier explanations. Second, I wanted a familiar context to ease them into the interviewing mode. The smoothness of the transitions from question to question were important because experience has shown me that sometimes the scenario described in a question can be confusing to students who had never actually done such activities. For example, telling someone to imagine drawing 6 handfuls of 10 candies each, with replacement, from a jar containing 100 candies in it (40 of which are yellow and 60 of which are red), can seem like an overwhelming amount of different numbers to keep track of for one question.

Table 4. *Contexts for PreInterview Questions*

Question	Brief Description	Contexts
1	Results are predicted for drawing one, several, and six samples of 10 candies from a jar.	Sampling
2	Lists are shown for different outcomes of six trials. Subjects are asked to comment on the likelihood of each list occurring.	Sampling
3	The supposed results of 30 samples are shown. Subjects are asked about the likelihood of the results being real or fake.	Sampling & Graphs
4	The supposed results of 300 samples are shown. Subjects are asked about the likelihood of the results being real or fake.	Sampling & Graphs
5	Three graphs show different ways of portraying the same data set. Subjects are asked how the graphs differ.	Data & Graphs
6	A set of 21 measurements for the duration of a train ride is given. Subjects are asked for reasons why the results are not identical.	Data & Graphs
7	The 21 measurements from Q6 are graphed in two different ways. Subjects are asked to compare the two graphs.	Data & Graphs
8	Two different graphs are shown: Wait-times for eastbound and westbound trains. Subjects are asked to compare the graphs.	Data & Graphs
9	Results are predicted for one sample of sixty tosses of a fair die..	Probability
10	Supposed results are shown for four samples with the die, and subjects are asked which results seem real or fake.	Probability
11	Results from repeated samples with the die are predicted.	Probability
12	A 2:1 (White:Black) spinner is used, and results are discussed in terms of what seems surprising to the subjects.	Probability
13	For the spinner in Q12, one sample is defined as sixty spins. Two graphs showing supposed results from 20 samples are compared	Probability & Graphs

Because the PreInterview took place before the stochastics portion of MET 2, I did not want to assume that the questions would automatically make sense to the subjects. Therefore, just as the PreSurvey was scaffolded to include one sample, then several, then six, so too was the PreInterview modeled to move gradually to ever larger numbers of samples.

The PreInterview script contained specific questions that I used with each subject, but the protocol also allowed me flexibility to follow the subjects' train of thought. Thus, each interview contained common questions as well as unique input from the different cases.

Class Intervention #1

All of the class interventions are described in greater detail in Appendix C, so in this chapter they are only briefly discussed. The two activities comprising the Class Intervention for the context of data and graphs were called "Four Questions" and "Body Measurements".

The "Four Questions" activity was chosen for two reasons. One reason is because Steve and I had each used versions of the activity with other MET 2 classes, and were therefore experienced in how it went and what it offered. The second reason is that it offered a good opportunity to discuss both average and spread in data sets. Steve therefore started the class exploration of statistics in the fifth week by having the entire class gather data from one another in response to four questions:

> How many pets do you have?
> How many years have you lived in Portland (to nearest half-year) ?
> How many people are in your household?
> How much change (in coins) do you have today?

After graphing the data in different ways, the class had a discussion about levels of detail provided by each type of graph and about what were "typical" values for a MET 2 student or for the whole class. The tension between centers and spread of data was one theme to emerge from the discussion over graphs from the "Four Questions" activity.

The second activity in the class intervention that focused on the context of data and graphs was "Body Measurements", which was also selected because it was a well-rehearsed activity for Steve and me. More importantly, a similar activity was to be used for the NSF-sponsored project with middle and high school classes, and comparisons could be made between EPST's and precollege students in the future. As in "Four Questions", we gathered class data for "Body Measurements": Everyone's own armspan, height, handspan, head circumference, and pulse rate per minute were recorded. Also, all students in class measured Matt's armspan, to gather data from a repeated-measurements experiment. Again, we had a class discussion about the data and graphs for the body measurements, this time focusing more on causes of variation.

Data & Graphs PostSurvey

The Data & Graphs PostSurvey, a take-home assignment, was given at the end of the final class session of week 6, and collected the following week. Students had from Thursday to Tuesday to work on the PostSurvey, and they were encouraged to work together on the assignment. I can't be certain who did or did not work in teams outside of class, but traditionally groups of MET 1 and MET 2 students do spend time before or after class working together. The PostSurveys were graded by Steve only as "done" or "not done", similar to other writing assignments given in class. Table 5 summarizes the questions asked on the Data & Graphs PostSurvey.

In creating the Data & Graphs PostSurvey questions, I wanted to examine students' reasoning as they compared or evaluated different graphs. For example, would students refer more to centers or to spreads? Also, I wanted see what kinds of causes for variation they could come up with on their own. In Q1c, I was curious about their ability to reason from the given average of 4 inches of rain to generate a reasonable graph showing appropriate variation for the rainfall during a typical June in Columbus.

Table 5. *Summary of Data & Graphs PostSurvey Questions*

Question	Brief Description
1a	Bar charts are given showing the 30-year average monthly rainfall for Portland and Columbus. Students discuss differences and causes in the rainfall patterns.
1b	Boxplots are given for the same data sets used in Q1a. Again students discuss differences, and also are asked which city they think is rainier, and why.
1c	Assuming that the average June rainfall in Columbus is 4 inches, students are asked to draw a graph showing what each day's rainfall in June might look like.
2a	Dotplots and boxplots are given showing annual traffic death rates for two regions in America, the South and Northeast. Students are asked to compare the rates.
2d	Students are asked to think of factors that might explain the differences in rates between the two regions.

Finally, for my six cases, and the others who would be interviewed a second time, I wanted additional evidence in their reasoning about variation in histograms, dotplots, and boxplots.

Class Intervention #2

In the seventh week of class, the two activities "Known Mixture" and "Unknown Mixture" were done with Steve's students. Matt and I had done this activity at six schools as a part of the NSF-sponsored project. We had seen how effective the activities could be in drawing attention to variation. For example, middle and high school students always commented on the different ranges of the graphs depicting actual results of the sampling activities.

Prior to the Known Mixture, we started with a general discussion of what samples were, who uses samples, and what samples were good for. Then the scenario in Figure 10 was given as a part of a handout. The class discussed initial expectations for this scenario, especially focusing on what would happen if the random draw of 10 names were to be repeated thirty times.

<u>Scenario for Known Mixture Activity</u>

The band at Johnson Middle School has 100 members, 70 females and 30 males. To plan this year's field trip, the band wants to put together a committee of 10 band members. To be fair, they decide to choose the committee members by putting the names of all the band members in a hat and then they randomly draw out 10 names

Figure 10 – Known Mixture Activity

After students talked about predictions for drawing thirty samples each of size ten, we simulated this activity using chips in a jar. Actual data was gathered and graphed. Then we had a discussion about how the graphs of the predicted data compared to one another, how the graphs of the actual data compared to one another, and also how the predicted graphs compared to the actual graphs.

We then made a transition into the second activity in this intervention, the Unknown Mixture. It was made clear that even though we had known what was in the earlier jars, samples still had varied. Now we had larger jars, each containing 1000 chips of yellow and green with the same mixture. However, the exact mixture was not known to the class (it was actually 550 yellows and 450 greens). The students were asked to decide in their groups what sample size they wanted to use (we imposed an upper limit of size twenty for all groups) and how many samples they wanted to draw. Then they were to carry out their plans, do the sampling, graph the results, and make some conjectures about the true mixture in the jar. After the simulation was carried out, we had a class discussion about the different choices made in sampling, the class results, and we tried to forge a class consensus about what the true mixture was.

Sampling PostSurvey

The Sampling PostSurvey take-home assignment was given at the end of the final class session of week 7, and collected the following week. Table 6 summarizes the questions asked on the PostSurvey.

Table 6. *Summary of Sampling PostSurvey Questions*

Question	Brief Description
1a	A boxplot shows the results of 20 samples of size 10 drawn from the smaller candy jar (60 Red & 40 Yellow). Individual sample results are then inferred.
1b	Since the minimum result on Q1a shows 3 Reds in a handful of 10, the number of trials needed to get a 0 or 1 Red are predicted.
2	Results are predicted for drawing one, several, and six samples of 100 candies from a larger jar (600 Red & 400 Yellow)
3	Ranges are predicted for drawing thirty and then three hundred samples of 100 candies
4	Results are predicted for drawing fifty samples of 100 candies

I asked Q1a because it got at the idea explored in class about how boxplots can obscure variation. I was also curious to investigate their ability to work with boxplots, especially working backwards and predicting what the underlying distribution might be which led to that boxplot (as opposed to taking the data and creating the boxplot as is typically done). Q1b was added to draw on our class experience with the ProbSim software and the numbers of samples needed to get extreme results. Questions 2, 3, and 4 all were parallel to questions previously asked in the PreSurvey and PreInterview for the Small Jar (60 Red & 40 Yellow), only now the sampling was done from the Large Jar (600 Red & 400 Yellow). I was curious to see how answers for the Large Jar compared to what students had put for the Small Jar, and I also wanted to have my six cases thinking about the Large Jar because I built additional questions on that sampling scenario into the PostInterview.

Class Intervention#3

There were two activities that made up this intervention, "Cereal Boxes" and "The River Crossing Game". These were chosen specifically because of the probability aspects involved in the activities. Cereal Boxes relies on the use of spinners and River Crossing on the use of dice as random generators, and these two activities were the main ones done in MET 2 involving random devices.

Cereal Boxes actually took place in the first class session of week 2, just before we gathered data for Body Measurements. As explained earlier, there was considerable overlap in the three contexts, and Cereal Boxes is a good example of this overlap. Cereal Boxes is sample-until scenario, assuming that any of five different stickers can be obtained within each box of cereal, and that the five stickers have equal chances of being obtained. The question is, about how many boxes would need to be opened to obtain all five stickers. The situation can be simulated by using an equal-area five-region spinner. Cereal Boxes brings together probability, sampling, and data and graphs in a way that highlights variation.

The second activity for this intervention, the River Crossing Game, involved finding the sum of two dice. Both the Cereal Boxes activity and River Crossing Game are part of the *Math and the Mind's Eye* curriculum (Shaughnessy & Arcidiacono, 1993).

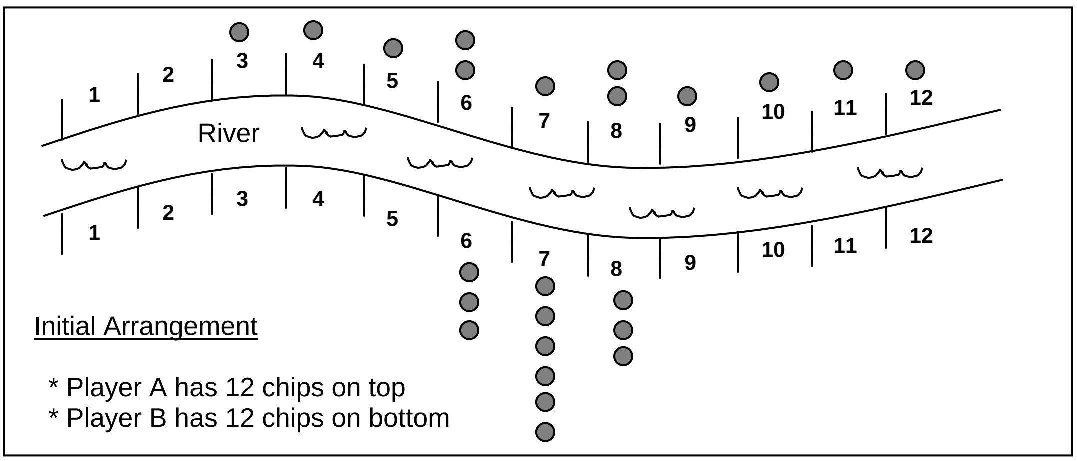

Figure 11 – River Crossing Game

Using two players, each player receives 12 chips to place on their side of a "river", along spaces marked 1 through 12. After configuring their chips in an initial arrangement (see Figure 11 for an example of two players' initial arrangements), players took turns tossing a pair of dice. If either player had any chips on the space showing the total for the dice, one chip could "cross the river" and be removed from the board. The winning player was the first one to remove all the chips on his or her side. For instance, in Figure 11, if the dice resulted in a sum of 10, Player A on top could remove one chip. If the dice showed 8, Player A and B could each remove one chip. As with Cereal Boxes, in the River Crossing Game we made predictions, gathered and graphed data, and discussed results.

Probability PostSurvey

The Probability PostSurvey take-home assignment was given at the end of the final class session of week 8, and collected the following week. Table 7 summarizes the questions asked on the Probability PostSurvey.

Table 7. *Summary of Probability PostSurvey Questions*

Question	Brief Description
1	The number of blacks resulting from 50 spins at a ½ -white & ½ -black spinner make up a single sample. Results are predicted for doing one, two, and six samples of 50 spins.
2	Ranges are predicted for doing thirty and then three hundred samples with the spinner.
3	A graph is made to show what the results of fifty samples (each of 50 spins) of the spinner might look like

Question 1 was crafted to be similar to the sampling questions on the PreSurvey, the PreInterview, and the Sampling PostSurvey, only this time the focus was on probability. Instead of grabbing handfuls of candies, the students were asked to consider samples of fifty spins of a 50:50 (White:Black) spinner. Also, Question 1 was similar to the PreSurvey question on probability that used a sample of 50 flips of a fair coin. Question 2 was similar to what was asked on the PreSurvey and Sampling PostSurvey, and the graph of fifty samples for question 3 was similar to what was asked for fifty samples of the Small Jar in the PreSurvey. I was curious to see how the probability questions with the flips of a coin (in the PreSurvey) compared to the spinner environment (in the PostSurvey), and also how the responses for the probability context compared with those responses for the sampling context. Finally, for my six cases I wanted added familiarity with the spinner scenario, since I based several additional questions in the PostInterview on repeated samples of spinners.

PostInterview

The PostInterviews took place during the 9th and 10th weeks of the quarter, after Steve had gone back to teaching geometry. I followed a very similar protocol with the PostInterviews as I had with the PreInterviews, videotaping each interview as well as making a separate audiorecording for the transcribing process. The PostInterview contained 13 multi-part questions, and the contexts for the questions are given in Table 8. More details of the PostInterview questions and sample responses are provided in the next chapter.

As in the PreInterview, I chose the very first question on the PostInterview to be identical to a question the subjects had already seen (it had been asked in the Sampling PostSurvey. Also, the first four questions were isomorphic to those on the PreInterview, maintaining the population proportion of 60% Red, except that they used the Large Jar (600 Red & 400 Yellow) as opposed to the Small Jar (60 Red & 40 Yellow). Question 5 involved both the Small and the Large jars, the only question in any of the instruments to do so directly. Questions 6, 7, and 9 were asked in part because of their similarity to the MAX train ride questions on the PreInterview. Lastly, question 10 was identical to the first question on the Probability PostSurvey, and set the subjects up for a transition to the last few questions that also involved spinners.

Table 8. *Contexts for PostInterview Questions*

Question	Brief Description	Contexts
1	Results are predicted for drawing one, several, and six samples of 100 candies from a large jar.	Sampling
2	Lists are shown for different outcomes of six samples. Subjects are asked to comment on the likelihood of each list occurring.	Sampling
3	The supposed results of 30 samples are shown. Subjects are asked about the likelihood of the results being real or fake.	Sampling & Graphs
4	The supposed results of 300 samples are shown. Subjects are asked about the likelihood of the results being real or fake.	Sampling & Graphs
5	Two graphs show supposed results of forty samples at the small and also at the large jar. Subjects are asked if graphs are real or fake.	Sampling & Graphs
6	A set of 20 measurements for the weight of a muffin is given. Subjects are asked for reasons why the results are not identical.	Data & Graphs
7	The 20 measurements from Q6 are graphed in two different ways. Subjects are asked to compare the two graphs.	Data & Graphs
8	35 different muffins from the West End bakery are shown. Subjects are asked how much their (36^{th}) muffin might weigh.	Data & Graphs
9	Two different graphs are shown: Muffin weights for East and West End bakeries. Subjects are asked to compare the graphs.	Data & Graphs
10	Results are predicted for doing one, two, and six samples of 50 spins at the 50:50 spinner.	Probability
11	Lists are shown for different outcomes of six samples. Subjects are asked to comment on the likelihood of each list occurring.	Probability
12	A graph shows the supposed results of twenty samples. Arguments from other people about the results are discussed by subjects.	Probability & Graphs
13	Two graphs show supposed results of two classes doing thirty samples at the spinner. Subjects are asked if graphs are real or fake.	Probability & Graphs

It was mentioned earlier how the activities in the interventions were designed to get at variation. As can be seen in Table 8, the activities in the interventions had direct ties to the tasks on the PostInterview questions. For instance, the intervention on data and graphs included different types of graphs and the amounts of variation they showed. Body Measurements got at the ideas behind repeated measurements, as did the muffin weight questions on the PostInterview. The Known and Unknown mixtures had students actually draw chips from a container to experience drawing candies from Large and Small Jars. Cereal Boxes and the River Crossing Game had students use traditional random generators such as spinners and dice to get a sense of what was likely in a probability context. There is one big difference on the PostInterview compared to the PreInterview: PostInterview questions 8, 9, and 12 all included boxplots as well as either dotplots or histograms, but boxplots were not covered on the PreInterview. Thus, several of the tasks involving graphs had two types of graphs, again relying on the experience gained in the class interventions.

Data Analysis

In this section, I'll illustrate the process of using grounded theory techniques to develop what I call my evolving framework, which addresses my primary research question. The framework in its entirety is presented in the next chapter, which focuses on *results* of the study. The *method* by which the framework was derived is described in this chapter. The process of deriving a framework is laborious and quite detailed, so I'll just use one piece of the framework to provide an example of how I applied the three techniques of grounded theory: open coding, axial coding, and constant comparison. Before launching into my illustrative example, a last methodological component to be discussed is the role of computer software in my data analysis.

Role of Computer Software

The creative yet systematic process of theory building, which in my study took the form of fleshing out a conceptual framework about variation, was aided enormously by the NUD•IST software.

Grounded theory techniques allow for the inclusion of a wide scope of data, such as the written responses, transcribed interviews, and observational notes. As the process of theory development moves through cycles of constant comparison, memos suggesting the continual conjecturing and refinement of categories and concepts also become a part of the data. Thus, data management becomes a crucial issue in using grounded theory. Richards (1994) boldly states that "all researchers working in the qualitative mode will clearly be helped by some computer software" (p. 105). The use of qualitative data analysis software facilitates not only the management of data, but "it can offer leaps in productivity for those adept at it" (Patton, 2001, p. 447).

The software used in this study was NUD•IST (Non-numerical Unstructured Data Indexing, Searching, and Theorizing), a theory-building program that aids in data storage, coding, retrieval, and category comparisons and linking (Richards and Richards, 1994; Richards, 1994; Patton, 2001). NUD•IST is well-suited for the analysis techniques of grounded theory, although it cannot be emphasized enough that software only assists in the process – software does *not* analyze data for the researcher (Patton, 2001; Creswell, 1998). The flexibility in coding, categorizing, and revising made the development of theory a dynamic and reflexive process. The single biggest obstacle was in learning what NUD•IST could do for the analysis, and how to get the program to do what I wanted. Once the various ways of categorizing, indexing, and coding were learned – the logistics of the program as well as the potential – NUD•IST became an invaluable tool.

Developing the Evolving Framework

As a tradition of qualitative inquiry, grounded theory allows analysis to begin even in the absence of any initial structure or preconceived ideas of what the data might hold. For my research, I did start with an initial conceptual framework, which was based on previous research as well as my own experience. However, the initial conceptual framework was a rough structure that offered little in the way of specifics; it was mainly used as an overarching guide for designing my tasks. Recall that the initial framework had three aspects: *expecting*, *displaying*, and *interpreting* variation. Each aspect had different dimensions. For example, *expecting* had the two dimensions of *what* was expected and also *why* it was expected. Grounded theory offered a way to expand on the dimensions of the initial conceptual framework. I'll show how this expansion happened with the dimension of *what* was expected.

I'll start with the process of *open coding*, whereby I gathered all responses having to do with *what* was expected and looked for what I called broad common "themes" within the responses. Open coding led to the emergence of three themes for the dimension of *what* was expected: responses concerning the expected value, responses about repeated values, and responses about a range or extreme values. I then used *axial coding* to focus on these three themes in turn to describe them in terms of what I called "characteristics" of the theme. The process of *constant comparison* meant that I iteratively went back and forth from the data to the emerging themes and characteristics, looking for confirming and disconfirming evidence from all my data sources to help me conceptualize the evolving framework as it was being built.

Open Coding: All my observation notes, memos, interviews, and survey data had been transcribed and imported into NUD•IST as text files. I chose to set a line of text as the smallest unit which could be coded (other choices included setting a paragraph or section or the entire document as the unit of analysis). I then went through all the data and coded all the occurrences where I found references to *what* was expected. In NUD•IST the process of coding means highlighting the lines of text and selecting a "node" to code the text at. Every node has a title, and on this first pass at coding I simply titled my node "Describing What is Expected".

The examples I'll use in this section are responses taken from my six cases from Question 10 on the PostInterview. The process I'll be describing was applied to all of my data, but in order to illustrate the depth of analysis that led to the framework, I need to limit the example. Question 10 had three parts (Q10a, Q10b, and Q10c), and imagined a man getting a sample of 50 spins of a half-black and half-white spinner. The question in Q10a was "How many times do you think the arrow might land on black? Why?" Table 9 shows some sample responses that I coded at "Describing What is Expected" for Q10a.

Each line of text is coded at the level of my dimension ("Describing What is Expected"), but a closer examination of the responses shows different themes within the responses. A review of my memos showed that the first thing I noticed was how both DS and EM used the words "close to 50%", and JM used the parallel "approximately 50%". Comparison of other data for these cases showed me that they did know 50% of the 50 spins in this situation was 25 blacks, so when they referred to "50%" it was isomorphic to saying the expected value. That got me thinking that a theme to look for in their responses had to do with what they said about the expected value.

Table 9. *How many times might the arrow land on black? Why?*	
Subject	Response
DS	Like, just close to 50%, but not exactly! Yeah, within 2 or 3.
EM	I think it will land there somewhere close to 50% of the time, I don't think it will always be 50% of the time, I think it will be probably between 40 and 60% of the time. So, 25...between 20 and 30 spins.
JM	JM Well, approximately 50%, but it will be , you know, plus or minus, maybe, 20% of that number – Somewhere in there.
RL	Oh, how about… Somewhere between 21 and 29… I don't say, you know, 18 to 29… I'm going afar from 25 in either direction. It's probably within that [21 to 29] range.
SP	Yeah, so it expect it , like, up from 25, maybe like 30, and 20 … between 20 and 30

I then coded the responses of DS and EM and JM as "Concerning Expected Value", since that was a theme I was hypothesizing. A strength of NUD•IST is that one can code text at as many nodes as one wishes. Now, for example, DS's response has two codes: one code at the dimensional level "Describing What is Expected" and also at the thematic level "Concerning Expected Value." At this point, I wouldn't know much about the characteristics of the theme "Concerning Expected Value" until I'd done axial coding. Something else I noticed in the responses was a focus on range, such as when DS mentioned "within 2 or 3", or EM said "between 20 and 30 spins". Notice how RL gives both a range he is comfortable with (21 to 29) and a range he thinks is too wide (18 to 29). At this point, I created a node for the theme I was hypothesizing, and titled it "Concerning Range or Extremes". JM's response also reflected the theme concerning range ("...plus or minus, maybe, 20% of that number") as does SP's response ("between 20 and 30"). Lastly, I noticed how EM said "I don't think it will always be 50% of the time," and that made me wonder about a possible theme concerning repeated values. That is, I thought maybe I should be on the lookout for responses that specifically mentioned if results would be the same or different from one sample to the next. I had a tentative theme "Concerning Repeated Values" which I carried into the next set of responses with my other themes.

This illustrative example omits many of the other memos and conjectures I had in the first pass at open coding, and makes the process appear more streamlined than it was. The point of microanalysis (open coding combined with axial coding and the method of constant comparison) is that you can begin with broad and multiple categories and properties and then winnow them down, collapsing and combining as you continually re-conceptualize your data. Thus, there are many other themes that had occurred to me in my first pass at looking at the data, but I am only presenting the final ones here.

<u>Axial Coding:</u> After I applied open coding to all my data, axial coding encouraged me to focus on the themes, explicitly looking at the various characteristics of each themes. In Table 10 I show some responses to Q10b, which asked how the results of a second sample of 50 spins would compare to the first sample results from Q10a. I've presented the lines of text along with the coded themes.

The main difference between what I did in open coding and what I did in axial coding comes down to focus. In open coding, I collected all responses that had some information "Describing What was Expected". At the same time, I had many memos and tentative themes that I thought might be emerging.

Table 10. *How do you think his results on the second set of 50 spins will compare with the results of his first set?*				
Subject	Response	Themes		
		Expected Value	Repeated Values	Range or Extremes
DS	I think it'd be close to it, but different.	●	●	
	So, maybe if he got 28 the first time,			●
	he'd get 24 the second time, or 23...			●
EM	Somewhere near 50%, right	●		
JM	Yeah, I think [it'd be] fairly close in	●		
	the sense that it's gonna be	●		
	around the... 25 blacks,	●		
	plus or minus that 10% or so.			●
RL	I think that it's likely to fall in a			●
	same range, similar range.			●
SP	I think the range would still be			●
	somewhere very similar to that one.			●
	There'd be – just different numbers,		●	
	but still somewhere in that range.			●

In axial coding, I went back through the data and specifically focused on potential themes. For instance, in Table 10, I looked for and coded responses as "Concerning Expected Value" I noticed how DS mentioned that the second sample would "...be close to it", and a comparison of what she had said earlier showed me that what she was suggesting was a result that would be close to the expected value (of 25 blacks). EM and JM also have responses coded at the theme "Concerning Expected Value", with their expression of results being "near 50%" and "around the ...25 blacks". Then, I went back through the data and looked specifically for responses "Concerning Repeated Values". In Table 10, DS and SP specifically mentioned that they expected different results. Similarly, for the theme "Concerning Range or Extremes", JM, RL, and SP explicitly used the term "range", while DS suggested possibilities that range from 23 blacks to 28 blacks.

In the axial coding process, I made many memos to record my own thinking about the data. For example, I noticed when DS mentioned possibilities ranging from 23 to 28 blacks that those were reasonable choices and she didn't seem to expect unlikely extreme values such as 5 or 45 blacks. When RL mentioned range, he first said "same range" but then immediately amended this by saying "similar range". By doing a comparison of other places where RL talks about results being "similar", I was able to find out that in his case and also in the case of most of the students, "similar" seemed to imply "similar but different." By thinking deeply about the themes, I initially recorded my thinking as tentative links between the characteristics of the themes, and the links either became stronger with the addition of more and more data or were discarded.

<u>Constant Comparison:</u> This method lets me take new data and compare it to the themes and characteristics even as those themes and characteristics are emerging. Constant comparison is an ongoing process that can occur alongside and after both open and axial coding. In my previous illustrations of open and axial coding, many instances of constant comparison took place as I referred to other similar survey or interview responses for the different subjects. I also made references to class observations. For instance, in Table 9, JM referred to expecting approximately 50% blacks "plus or minus...20% of that number". Then, in Table 10, he mentioned 25 blacks "plus or minus...10%". By looking at what his group did in class on the posters for the Unknown Mixture, and by referring to my notes on class observations, I could see that a 10% margin of error had been discussed. Thus, JM may have been

influenced to think in terms of “plus or minus 10%” even though 20% of 25 is an easier number to work with. Notice how in Table 9 EM and SP gave ranges of 20 to 30 blacks, which corresponds to 25 blacks plus or minus 20% (of 25). Also, EM had said “40 to 60% of the time” which could be thought of as 50% plus or minus 10%. Constant comparisons let me draw from many data sources as I made different connections about my themes and their characteristics.

Constant comparisons also let me see when I had *saturated* my dimensions, themes, and characteristics. Saturation in this example referred to the point where new data is not adding anything new to the themes I had initially envisioned. Table 11 shows the results for Q10c, which asked for a list of what might happen in six samples of 50 spins each, and why. The same three themes as in Table 10 are again shown in Table 11, which helped confirm that these were themes that I could characterize and look for in other data from other questions. I also was able to compare the subjects’ lists for their six samples to the responses they had given earlier. For example, DS had said in Table 9 “within 2 or 3” (of 25 blacks), yet her list reflects 25 plus or minus 4. Notice too how EM had earlier suggested a range of 20 to 30 blacks, but in her own list she went a little wider than that on both sides. JM did a bit of self-reflecting as he considers his range of 21 to 29 and thinks about what he had said about “plus or minus 10%”. RL confirmed my ideas about his earlier response when he explicitly stated “similar, but not identical.”

Saturation does not imply that one stops thinking of new connections that can be made among the themes and characteristics.

Table 11. *Write a list to describe what might happen in six sets of 50 spins. Why did you choose those numbers?*

Subject	Response	Themes		
		Expected Value	Repeated Values	Range or Extremes
DS	[28, 23, 25, 29, 21, 26] I have a few scattered	●		
	close to 25, but not 25. So that...probably on	●	●	
	average it'll be close to 25.	●		
EM	[20, 23, 28, 32, 18, 24] For the most part, it was			●
	generally concentrated between – give or take...			●
	50% of the time, somewhere near there…But...			●
	occasionally it would be even lower than that,			●
	so I... threw the18 in there...And the 32, yeah.			●
GP	[21, 23, 24, 25, 27, 28] They're around the 25.	●		
JM	[21, 23, 25, 26, 27, 29] Well, they're close to	●		
	that 50 percentile, that we're looking for , plus	●		●
	or minus – I'm thinking – 10% or so. Actually,			●
	I'm a little high aren't I, with the 29? But still...			
RL	[21, 23, 22, 27, 28, 29] I did a pair, each equally			●
	far out from the mean in either direction. And...	●		●
	there's no repeats, but they're all sim[ilar] …		●	
	They're similar, but not identical		●	
SP	[20, 23, 24, 26, 28, 29] Just fall within that			●
	range of 20 to 30, that you would expect. Yeah,		●	●
	they could repeat, but I just did a range acro[ss]		●	●
	from 20 to 30, just to choose.			●

There are so many different layers of meaning that can be found in a grounded theory approach to qualitative analysis that one could continually return to the data and find something new. Thus, being guided by my principle research question to come up with a more descriptive framework for understanding EPSTs' conceptions of variation, I concentrated on describing the characteristics of the dominant themes I saw rising out of the data.

This section provides a taste of what characterizes the three themes for "Describing What is Expected". Concerning the expected value, Tables 9 through 11 show how results are thought of as being “close to”, “somewhere near”, or “around” that value. Results may or may not include the expected value. For example, in Table 11 when picking six hypothetical results, three of the cases did include 25 blacks in their list, and the other three cases did not. Concerning repeated values, although multiple samples could repeat, they shouldn’t all be identical. Results might be “similar, but not identical.” Again looking at Table 11, notice how every one of the six cases gave lists that were composed of distinct entries. There is not a single repeated value in any given list. Concerning range or extreme values, subjects sometimes stated explicit numerical ranges which gave clues to the numbers they were comfortable with. In the example of Q10, a range of about 20 to 30 meant that 21 to 29 was considered reasonable, but so too was 18 to 32. Thus, even when stated explicitly, ranges tended to be flexible.

Even with the detailed illustration given within this section, the themes for “Describing What is Expected” (Concerning Expected Value, Repeated Values, and Range or Extreme Values) have not been completely fleshed out just by responses to PostInterview Q10. I have only profiled the six subjects’ responses to this question. My purpose in this section has been to show how I used the grounded theory techniques to do a microanalysis of the data. The complete analysis encompassed six cases’ responses to the interview questions as well as classwide responses to the survey questions. The illustration of microanalysis in this section relied on less than 50 lines of text comprising subjects’ responses in Tables 9 through 11, while the entire research project covered literally thousands of lines of text. Thus, NUD•IST was extremely helpful in managing the codes as they emerged, especially since a key feature of NUD•IST is the ability assign multiple codes to lines of text. As many of the responses in this section showed, a subject could give a sentence or fragments of sentences that contain several distinct meanings. DS gave a succinct example in Table 9, when she expected results "...close to 50%, but not exactly! Yeah, within 2 or 3". She appeals to the expected value, and also gives a range. In the next chapter I'll summarize each theme for each dimension within the aspects comprising my evolving framework.

Summary

Subjects were chosen from the MET 2 class, where the pedagogical style in the class and in its prerequisite, MET 1, encourages students to communicate their thinking both verbally and in writing. Subjects were familiar with and able to openly share their thinking, a necessary condition for this study. The overall design for the research incorporated elements of both grounded theory and case study traditions. In addition to interview data gathered from six cases, classwide data was gathered via a set of written documents (surveys), and the study was also augmented by in-class observations.

The three data gathering methods, observation, document review, and interviews, all were connected to the three class interventions. The three interventions (made up of two activities each) corresponded to the three contexts for looking at conceptions of variation: variation in data sets, variation in sampling, and variation in probability situations. After administering a PreSurvey to the whole class, PreInterviews were held outside of class time. Then, after each class intervention was conducted, a take-home PostSurvey was distributed. Finally, PostInterviews were completed.

To analyze the data, the use of grounded theory techniques allowed for the characterization of the subjects' understanding to naturally emerge within the conceptual framework in the shape of distinct but linked themes. The rudimentary structure posited as an initial conceptual framework was validated and extended by doing a microanalysis of all the data collected. The data analysis was facilitated by the use of the NUD•IST computer software. The final result was a rich description and overall characterization of the conceptions of variation held by elementary preservice teachers, given by the evolving framework in the next chapter.