The Pennsylvania State University

The Graduate School

College of Education

DEVELOPING AN UNDERSTANDING OF VARIATION: AP STATISTICS TEACHERS' PERCEPTIONS AND RECOLLECTIONS OF CRITICAL MOMENTS

A Dissertation in

Curriculum and Instruction

by

Susan A. Peters

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2009

The dissertation of Susan A. Peters was reviewed and approved* by the following:

Rose Mary Zbiek
Professor of Education
Dissertation Advisor
Chair of Committee

Glendon Blume
Professor of Education

M. Kathleen Heid
Distinguished Professor of Education

Thomas Hettmansperger
Professor of Statistics

Laura Simon
Lecturer in Statistics

Glendon Blume
Professor of Education
Coordinator of Graduate Program in Curriculum & Instruction

*Signatures are on file in the Graduate School

ABSTRACT

This phenomenological study investigates conceptions of statistical variation that secondary mathematics teachers who are recognized leaders in AP Statistics exhibit. This study also investigates perceptions and recollections of activities and actions that teachers who exhibited robust understandings of variation suggest contributed to their current understandings of variation. The data include questionnaires, event history calendars, critical incident descriptions, resumes, course syllabi, content-focused interviews, and two learning-context interviews for each teacher. Constant comparative analysis (Glaser & Strauss, 1967) of content-interview data and syllabi yielded three distinct types of teachers' conceptions of variation: Expected but Explainable and Controllable (EEC), Noise in Signal and Noise (NSN), and Expectation and Deviation from Expectation (EDE).

The teachers' responses to variation-related tasks were used in conjunction with the SOLO Model, research results about students' learning related to variation, and expositions on what it means to understand statistical variation to develop a framework for robust understandings of variation. The framework consists of two cycles of levels of reasoning in the formal mode. Robust understanding of variation is indicated from integrated reasoning about variation across three perspectives—design, data-centric, and modeling— in the second cycle of levels. Teachers' understandings of variation were assessed using the framework. Five teachers exhibited reasoning about variation that was consistent with robust understandings of variation.

Analysis of learning experience-related data for these five teachers followed protocol for phenomenological studies. Factors that may have contributed to these five teachers' developments of robust understandings include their interests in the field of statistics, their desires to have an overarching content framework for themselves and for their students, their foundational knowledge upon which they built deeper understandings, their propensities for

critical reflection, and their acting on opportunities to engage in learning activities and rational discourse with more knowledgeable others. The extent to which they embrace these opportunities may distinguish them from other teachers.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am grateful for the many people who supported and encouraged me throughout my six-year journey at Penn State and beyond. First and foremost, I would like to thank my parents: my mother, Esther L. Peters, and father, the late Wayne P. Peters. They worked hard to provide me with educational opportunities that were not available to them and encouraged me to pursue my dreams. My brother, Paul, sister-in-law Cathy, and nephews Michael and Greg provided love and support at the most opportune times and provided much-needed lightness and laughter during their visits with me.

I am particularly indebted to my advisor and friend, Dr. Rose Mary Zbiek, for her patience, encouragement, and unwavering belief in me. Her high expectations and relentless probing challenged me to continually refine my thinking and pushed me beyond what I believed I could achieve. I will be eternally grateful to her for her commitment of countless hours of time and energy to meet with me and read and re-read dissertation drafts to provide constructive feedback on my work. Through her brilliant insights, scholarship, and commitment to and passion for mathematics education, she provides an example of a true teacher and scholar to which I aspire. I also want to thank other members of my committee for their contributions to making these last six years among the most rewarding years of my life. Dr. M. Kathleen Heid introduced me to research and continually pushed me to think deeply about statistical concepts and mathematics education research, and Dr. Glendon Blume provided insights and encouragement throughout my time at Penn State. Collectively, these mathematics education faculty members provided the physical, financial, intellectual, and supportive environment that allowed me to grow personally and professionally. Working with them on MAC projects and on JRME was a privilege and an honor for which I am extremely grateful. Dr. Thomas Hettmansperger and Dr. Laura Simon exemplified teaching statistics for understanding and continually pushed me to make connections among statistical concepts. I benefited from the mentoring and insights contributed by all five committee members.

I owe a debt of gratitude to my mathematics education family at Penn State and to my extended family at Delaware and Maryland through the Mid-Atlantic Center for Mathematics Teaching and Learning. I appreciate the conversations and insights offered by graduate students, faculty members, and staff members. Among others, Linda Haffly and Tracy Scala each lent a sympathetic ear during the toughest times and somehow managed to bring a smile to my face with her humor. They believed in me, encouraged and supported me, and celebrated good times with me throughout the last six years. Dr. AnnaMarie Conner was a constant source of friendship and encouragement during the hours we spent commiserating and studying together both here at Penn State and remotely from Georgia, and she provided meaningful insights as I encountered milestone events in my pursuit of a Ph.D. She is a true friend and scholar. Donna Kinol, Jeanne Shimizu, and Shari Reed created a wonderful support system for me through their much-valued friendships.

I appreciate the continued support I received from my friends and my Twin Valley family. Sue Paull played a pivotal role in my decision to attend Penn State, and Sue, Lori Breton, Kathy Bean, and Elaine Troisi provided steadfast faith and encouragement during my time at Penn State. Bob and Ginny Mountz backed my education as a student at Penn State much as they championed my growth as a teacher at Twin Valley. Anne Carroll, Janice Ricks, and Ruth Carver sacrificed considerable time and energy to help me in my pursuits and continued to make me feel like part of the AP Statistics community. I am grateful to them and many others for all they have done for me during my tenure in State College.

Last, but definitely not least, I thank the 22 teachers who participated in the pre-pilot, pilot, and main dissertation studies. Their willingness to share their knowledge and experiences with me made this dissertation possible. Their commitment and dedication to their students and to statistics education speak volumes about the quality of their characters. I remain in awe of their substantial knowledge of statistics, eagerness to accommodate my visits and calls, selflessness in devoting numerous hours to benefit future teachers, and humility in sharing their views of statistics and learning with me.

This research was supported in part by the National Science Foundation under Grant ESI-0426253 to The Pennsylvania State University and by a research initiation grant from The Pennsylvania State University College of Education Alumni Society. Any opinions, findings, and conclusions are those of the author and do not necessarily reflect the views of the National Science Foundation or the College of Education Alumni Society.

Chapter 1

Rationale

Status of Statistics Education Research

Statistical content currently occupies a prominent position in content recommendations for students in Pre-Kindergarten through grade 12 (e.g., Burrill, Franklin, Godbold, & Young, 2003; Franklin et al., 2007; National Council of Teachers of Mathematics [NCTM], 1989, 2000), and national assessments for elementary and secondary students reflect an increased focus on data analysis (National Assessment Governing Board [NAGB], 2004; Tarr & Shaughnessy, 2007). Analyses of results from large-scale assessments like the National Assessment of Educational Progress (NAEP) point to improved student performance on data analysis items (D'Ambrosio, Kastberg, McDermott, & Saada, 2004; Tarr & Shaughnessy, 2007; Zawojewski & Shaughnessy, 2000), but concerns remain about students' performance on complex tasks that require sophisticated statistical reasoning (Tarr & Shaughnessy, 2007).

Student achievement often is considered in tandem with teacher knowledge, and recent research results support the widely accepted view that teacher knowledge can positively affect student achievement in mathematics (e.g., Hill, Rowan, & Ball, 2005). Some researchers (Reading & Shaughnessy, 2004) speculate that students' understandings of statistical concepts may be connected to their teachers' lack of experiences with the content and posit that "most" Pre-Kindergarten through grade 12 teachers have few statistical experiences (Shaughnessy, 2007).

To develop teachers' understandings of statistics concepts, researchers (Heaton & Mickelson, 2002; McClain, 2005), leaders from professional organizations (Conference Board of the Mathematical Sciences [CBMS], 2001), and curriculum developers (Chance & Rossman, 2006) opine that teachers need opportunities to experience the study of statistics in ways similar to how they are expected to teach the content. Current results from research suggest that while researchers are making progress in uncovering characteristics of experiences that result in teachers' learning of statistics (e.g., Hammerman & Rubin, 2004; Liu & Thompson, 2005; Makar & Confrey, 2002), researchers are just beginning to reveal characteristics that lead to teachers constructing robust understandings of formal statistical concepts.

Examining existing research studies that investigate teachers' conceptions or learning in statistics reveals conceptions or learning for a limited number of concepts, like hypothesis testing (Liu & Thompson, 2005), sampling distribution (Heid, Perkinson, Peters, & Fratto, 2005), or arithmetic mean (Callingham, 1997); for a limited number of problem contexts, such as group comparisons (Hammerman & Rubin, 2004; Makar & Confrey, 2005); or for a limited number of teachers within an exploratory setting, like the setting of a mathematics course for preservice elementary teachers (Canada, 2004) or professional development for inservice middle school teachers (McClain, 2005) designed specifically to promote particular understandings. When the amount of work needed to expand this limited scope of coverage is coupled with the length of time that typically exists before results from studies are disseminated publicly, designing research-based preservice and professional development programs that facilitate teachers' constructions for statistics in general seems to be a goal for the distant future. Current efforts can take years to begin to effect a change in statistics teacher education and may require more time than legislators and the public will tolerate to achieve a statistically literate population of teachers that can educate statistically literate students. Needed is a complementary research path that will provide more immediate results that eventually can be compiled with the outcomes of long-term investigation.

This study, a retrospective study with teachers who have robust understandings, offers an alternative approach intended to be timely as well as viable for uncovering the characteristics of experiences associated with successful learning. Rather than designing and studying a program that may be successful in having teachers construct robust statistical understandings, studying teachers who already have robust understandings eliminates the time required to design, implement, evaluate, redesign, and reevaluate an educational program. Retrospective examination of the characteristics of successful learning experiences for individuals who already have robust understandings occurs almost immediately. Retrospective study also does not require speculation about contexts that may be successful in perturbing individuals towards the construction of robust statistical understandings; by studying individuals who already have robust understandings, the focus shifts to uncovering the varied contexts that may have facilitated the construction of those understandings. Retrospective study, however, does not offer a panacea for all of the limitations of conventional study.

Because of the reliance on individuals' memories and the accuracy of those memories, the collection and analysis of retrospective data raises issues of reliability and validity (Martyn & Belli, 2002). Research results offer strategies that can reduce the impact of recall effects, including the use of instruments like event history calendars (Freedman, Thornton, Camburn, Alwin, & Young-DeMarco, 1988; Martyn & Belli, 2002)[1] and critical incidents descriptions (Brookfield, 1990; Butterfield, Borgen, Amundson, & Maglio, 2005; Flanagan, 1954).[2] Additionally, a single retrospective study realistically cannot investigate individuals' experiences in learning every statistical concept; however, by focusing on a key concept that underlies every area of statistics, characteristics of experiences critical for developing robust understandings of

[1] The format of the event history calendar is a matrix, with columns containing timing cues for recording behaviors and rows containing behaviors—significant activities or events related to the goals of the research—that can help individuals to frame the occurrence of important events (Freedman, Thornton, Camburn, Alwin, & Young-DeMarco, 1988)
[2] Critical incidents are defined to be unique events that evoke emotion at the time of occurrence or events that mark a transition point in life and are significant in the lives of individuals (Brookfield, 1990).

that concept arguably parallels the characteristics of experiences critical for developing robust understandings of statistics in general. Variation *is* one such statistical concept.

Variation and Statistics

The need to think statistically stems from the presence of variation. Statistical thinking encompasses finding ways to deal with variation in order to answer questions probabilistically and to determine the adequacy of the answers based upon considering the context from which the question originates. Many statisticians view the development of statistical thinking as fundamental to statistics education (e.g., Bailar, 1988; Cobb & Moore, 1997; Moore, 1998). In general terms, statistical thinking embodies an understanding of the statistical problem-solving process—that is, understanding both how to engage in the process and why the process is needed—and understanding the fundamental concepts that underlie the process (Ben-Zvi & Garfield, 2004). Variation plays a crucial role throughout the process of statistical investigation (Franklin et al., 2007).

Various facets of variation arise throughout the investigative process (Franklin et al., 2007). Failure to acknowledge variation or to anticipate possible sources of variation can render a statistical study meaningless before data collection begins. Identifying potential sources of variation allows some of those sources to be controlled through the processes chosen to collect data, thereby increasing the likelihood that the effect of or relationship with the factor(s) of interest can be determined. Because variation cannot be controlled completely, variation also plays a central role in the analysis and interpretation of data. Measuring variation and accounting for variability in the selection of a distribution or model to fit data enables determination of whether independent factors are related to or associated with dependent factors in ways beyond chance expectation. Variation prevents deterministic conclusions about relationships between independent and dependent factors from being made, leaving only probabilistically conditioned statements for interpreting results about a population of interest.

The breadth of individuals' reasoning about the facets of variation can be captured by considering their reasoning about variation through the lenses of three perspectives: a design perspective, a data-centric perspective, and a modeling perspective. Researchers have described data-centric and modeling perspectives on distribution (Prodromou & Pratt, 2006) and analyzed students' reasoning about distribution and variation in the same study using the same data to illustrate the connections between variation and distribution (Reading & Reid, 2006; Reid & Reading, 2005). Reading and Reid found that students' consideration of variation provides some indication of their abilities "to identify, understand, and use the key elements of a distribution" (p. 57), but they also found that students' "strong" consideration of variation included a need to "recognize the effect of a change of variation in relation to other concepts" (Reid & Reading, 2005, p. 51), including distribution. This seemingly reflexive relationship between variation and distribution merits consideration of variation from data-centric and modeling perspectives.

In this study, Prodromou and Pratt's (2006) descriptions of the data-centric and modeling perspectives on distribution have been expanded and modified to describe perspectives for reasoning about variation. This study adds the design perspective because reasoning about variation is warranted by the types of thinking associated with reasoning about variation in consideration of study design. General types of thinking associated with design include strategic thinking to plan and anticipate problems within practical constraints and thinking related to seeking explanations (Wild & Pfannkuch, 1999). Types of statistical thinking associated with design include considerations of variation through noticing and acknowledging variation during consideration of and selection of investigative strategies (Wild & Pfannkuch, 1999).

The three perspectives target different ways in which one might view variation. Reasoning about variation from the design perspective entails using context to identify the nature of and potential sources of variation and considering design strategies to control variation from some of those sources. Reasoning from the data-centric perspective includes measuring, describing, and representing variation while exploring characteristics of distributions and using those representations to make informal comparisons about the relationships among data and variables. Reasoning about variation from the modeling perspective incorporates modeling data or modeling characteristics of data to reason about relationships among data and variables for the purposes of making predictions or inferences from data.

Research Questions

With statistical content occupying a prominent position in the content recommendations for students in Pre-Kindergarten through grade 12, teachers' lack of experiences with statistics, and concerns about students' performance on national assessments, research that provides timely information for the eventual design of preservice and inservice teacher education in statistics is sorely needed. Given the centrality of variation to the study of statistics and the consideration of variation needed for statistical thinking, focusing on the characteristics of experiences for which teachers were able to construct robust understandings of variation can provide some needed information for the eventual design of programs that enhance the development of teachers' statistical thinking.

This study investigates some of these issues and in particular answers the following questions.

- What conceptions of statistical variation do secondary mathematics teachers who are recognized leaders in AP Statistics exhibit?

- For those secondary AP Statistics leaders who exhibit robust understandings of variation, what are the activities and actions that contributed to their current understandings of variation as reflected in their perceptions and recollections of experiences?

Answering the first question requires investigation of AP Statistics teacher-leaders' conceptions of variation. To answer the second question, clear explication of behaviors indicative of robust understanding of variation is needed to identify those teachers who provide sufficient evidence of robust understandings. Finally, a response to the second question requires examining the perceived beneficial learning activities and actions of those with robust understandings to look across experiences for common factors related to learning about variation.

Overview of the Study

This study is designed as a phenomenology for which the phenomenon under study is secondary mathematics teachers' development of robust understandings of variation. Primarily through the analysis of task-based, content-focused interviews and course syllabi, teachers' differing conceptions of variation are extracted and described. The Structure of the Observed Learning Outcomes (SOLO) Model (Biggs & Collis, 1982, 1991) is used to frame understanding and to analyze teachers' conceptions of variation. Through the analysis of two interviews focused on learning experience and instruments containing teachers' accounts of and perceptions of learning, factors contributing to the development of robust understandings of variation are extracted and compiled using data from those teachers who exhibit robust understandings. Analysis is guided by the detailed and systematic recommendations for phenomenological studies, as outlined by Moustakas (1994).

Students' developing conceptions of variation have been studied previously (e.g., Reading & Shaughnessy, 2004; Watson, Kelly, Callingham, & Shaughnessy, 2003) and reported in synthesized form (Shaughnessy, 2007). This study builds on the results of this prior research to provide empirical support from data collected from 16 AP Statistics teacher-leaders for the existence of three unique conceptions of variation for advanced knowers of statistics. Analysis of the data provided empirical support for a conceptual framework for robust understandings of variation. The framework differs from previous expository accounts of understanding in that it attempts to illustrate the connections and relationships among elements of the framework rather than provide lists of observable outcomes (e.g., Garfield & Ben-Zvi, 2005) and attempts to extend descriptions of what it means to reason about variation at advanced levels beyond responses to particular tasks (e.g., Watson & Kelly, 2004). Five teachers were found to provide clear evidence of robust understandings of variation, and data from these five teachers was used to extract learning factors that teachers perceive as contributing to developing their robust understandings.

The next chapter describes the empirical grounding of the study from studies described in statistics education literature and teacher education literature. Chapter 3 contains explication of the conceptual and theoretical grounding of the study, and Chapter 4 details the research methods used in this study. The conclusions and limitations in Chapter 8 follow three chapters that present answers to the research questions and articulate the meaning of robust understandings of variation.

Chapter 2

Literature Review

Research About Variation and Related Concepts

Despite the critical role of variation in statistics and the emphasis on statistics in elementary and secondary mathematics education, students' and teachers' conceptions of variation and their developing understandings of variation have not been common topics in research literature. In 1997, Shaughnessy outlined what he perceived to be "missed opportunities in research on the teaching and learning of data and chance" (p. 6), and, in particular, he identified the paucity of reported research on students' reasoning about variation as a "missed opportunity." Since that time, researchers have begun to study and publish results focused on students' and teachers' reasoning with variation and their conceptions of this key concept. Implicit within this literature are suggestions that preservice teacher preparation in statistics may not provide teachers with sufficient opportunities to develop robust understandings of statistical concepts. A comparison between research focused on teachers' reasoning and conceptions and similar work with students leads to the conclusion that, as Shaughnessy (2007) suggests, "teachers have the same difficulties with statistical concepts as the students they teach" (p. 1000).

The body of research that examines both students' and teachers' reasoning about and understanding of variation and related concepts suggests elements and connections needed for robust understandings of variation without providing a holistic image of robust understanding. When considered in conjunction with expository literature that outlines essential aspects and views of variation deemed necessary for deep understandings of variation, a clearer but still incomplete image of robust understanding comes into view.

The totality of research and expository literature about variation suggests that individuals' reasoning about variation can be captured from three perspectives: a design perspective that integrates acknowledgement and anticipation of variability in the design of quantitative studies; a data-centric perspective that integrates the processes of representing, measuring, and describing variation in exploratory data analysis; and a modeling perspective that integrates reasoning for fitting models to patterns of variability in data and statistics, judging the fit of models, and performing data transformations to improve the fit of models to make inferences from data. In addition to being able to reason competently about variation from these three perspectives, individuals should be able to integrate reasoning from the three perspectives while engaging in the statistical problem-solving process.

In this chapter, I consider both studies with teachers and studies with students to articulate what research reveals about conceptions of variation and reasoning with variation. I first explicate what research reveals about individuals' reasoning about variation and related concepts from a design perspective, as statistical problem solving begins with anticipating and acknowledging variation. Discussion of research on students' and teachers' reasoning from data-centric and modeling perspectives follows. Because understanding of variation both is dependent upon understandings of related concepts and is central for the development of understandings of related concepts, I also consider the results of research that examine concepts related to variation.

Students' and Teachers' Reasoning From a Design Perspective

A major focus of statistics is examining a question about a population through the analysis of data collected from a sample of the population. Making valid inferences about a population depends upon using appropriate sampling methods and designs—ones that appropriately anticipate and acknowledge variability for answering questions of interest. Without properly collected data, conclusions drawn from data are meaningless, which is why statisticians consider methods for collecting data that will allow the question of interest to be answered before they collect any data for analysis. At the heart of many of these methods and designs is randomization, which allows statisticians to determine if observed data characteristics like variation are due to chance or are likely to have been caused by some other factor (Franklin et al., 2007). Two important forms of randomization lie at the heart of observational and experimental designs: random sampling and random assignment. As Garfield and Ben-Zvi (2005) indicate, randomization produces data with variation in mind and minimizes bias in sample selection by introducing planned variation to data.

Reasoning About Variation and Sampling Methods

Elementary and middle school students have been observed anticipating variability in samples and acknowledging variability by recognizing benefits from sampling methods that align with generally accepted methods (e.g., Watson & Kelly, 2002a, 2002b). Of primary importance in observational studies is sample selection that typically includes some form of randomization to produce samples representative of the larger population from which they are drawn (e.g., Groth, 2003). Although random and representative samples may alternatively be considered to be fair and unbiased, everyday use of terms can interfere with students' intuitions about samples and fairness, for example. Young students seem to view a sample as a "bit of something" (Watson & Kelly, 2002a, p. 5) but do not necessarily intuit the "bit" as representative of a larger whole. The fifth graders observed by Jacobs (1999) did not associate fairness with individuals' equal probability of selection but rather associated fairness with individuals' *perceptions* of the selection process as fair. These students had no formal instruction in sampling, and practical issues interfered with their abilities to intuit statistically valid methods. In contrast to Jacobs' students, Watson and Kelly (2002a) observed third-graders subsequent to instruction suggest that samples selected through random methods are “fair.” Some students even described “fair” methods as methods that included ideas for producing random and representative samples, although students were not able to fully realize the benefits of randomization. Watson and Kelly (2002b) observed the same increased sophistication in reasoning about sampling methods from fifth graders. Their fifth graders scored significantly higher than and showed significantly greater improvement than third graders subsequent to instruction (Watson & Kelly, 2002b). These studies suggest that instruction may help students evaluate sampling processes and the products of those processes, although students’ reasoning falls short of describing *why* the sampling methods work.

Randomization is an important consideration for producing representative samples when designing studies, yet the importance of the topic has been largely overlooked by researchers. To date, researchers have paid little explicit attention to students’ conceptions of randomization or students’ reasoning about connections among randomization, variation, and sampling. Research suggests that elementary and middle school students benefit from instruction that examines the role of randomization in sampling, but what connections students make to variation are unclear.

Reasoning About Variation and Samples

With appropriate instruction, students are able to develop skills for reasoning about sample variability by making conjectures about reasonable sample compositions for samples selected from populations with known characteristics. Middle and high school students’ reasoning about sample variability has been classified according to three increasingly sophisticated types of reasoning: additive, proportional, and distributional (Shaughnessy, Ciancetta, & Canada, 2004). Students who reason additively focus on frequency counts, whereas those who reason proportionally focus on relative frequencies to make conjectures about samples drawn from a population of known composition. Supporting students' development of proportional reasoning is an overarching goal of middle school mathematics, but the complexity of reasoning proportionally is well documented (e.g., Behr, Harel, Post, & Lesh, 1992) and suggests one reason why individuals might struggle with the notion of a representative sample. Students who are able to reason proportionally may eventually be able to reason distributionally. Distributional reasoning is a more sophisticated type of reasoning about samples that entails reasoning with expected frequencies and reasonable deviation from expectation to consider possible sample compositions (Shaughnessy et al., 2004). In contrast with reasoning about samples from a population with known characteristics, distributional reasoning may be necessary but not sufficient for reasoning about a population from a sample. In their work, Shaughnessy and colleagues (2004) did not have students reason about the latter situation, but the work of Saldanha and Thompson (2002) suggests that reasoning that is more sophisticated than distributional reasoning might be needed to reason from relative frequencies and deviation from expectation for a sample in order to make inferences about the population.

Inferential reasoning seems to require a multiplicative conception of sample and sampling (Saldanha & Thompson, 2002). Students who reason multiplicatively communicate a view of sample as a "quasi-proportional" subset of a population *and* communicate a view of sample statistics in relation to distributions of sample statistics for samples of the same size. Multiplicative conceptions include the notion of comparing a single sample statistic against the population of statistics resulting from statistics for all possible samples of a given size from the population—that is, comparing a sample statistic to a sampling distribution. Multiplicative conceptions seem to be necessary in forming a firm foundation for inferential reasoning. The secondary students observed by Saldanha and Thompson infrequently exhibited multiplicative conceptions of samples and sampling to reason about variability and patterns of variability in sampling distributions. The researchers suggest that students who do display multiplicative conceptions have the support necessary for “building a deep understanding of statistical inference” (p. 268). Multiplicative conceptions seem to provide not only a foundation for understanding statistical inference but also seem to be important for considering variability in samples and sampling distributions.

Sample Representativeness and Sample Variability

Implicit in a multiplicative conception of sample and sampling are the notions of sample representativeness—the idea that a sample will have characteristics similar to those of the population—and sample variability—the idea that samples are not all identical and thus do not match the population exactly. To exhibit multiplicative conceptions, the ideas of sample representativeness and sample variability are balanced, meaning that an individual implicitly acknowledges that a representative sample should produce statistics similar to population parameters and different samples should be composed of values from different observational units and (most likely) have different summary statistics. Balancing notions of sample representativeness with sample variability has been shown to be a nontrivial endeavor. In their work, Rubin, Bruce, and Tenney (1990) noticed that students tend to overly rely on one idea or the other depending upon the problem context. An overreliance on sample representativeness leads to the deterministic belief that a sample tells everything about the population from which the sample is selected, whereas an overreliance on sample variability leads to the deterministic belief that a sample tells nothing about the population. The researchers contend that the two ideas “are contradictory when seen in a deterministic framework” (Rubin, Bruce, & Tenney, 1990, p. 315), with a sample simultaneously revealing everything and nothing about a population. Probabilistic reasoning is needed to balance the two ideas.

As individuals who take multiple mathematics courses during their undergraduate studies, secondary mathematics teachers may have a propensity for deterministic reasoning (Meletiou-Mavrotheris & Stylianou, 2003) and thus may struggle with the notions of sample representativeness and sample variation. Inferential reasoning—reasoning from the variation of samples toward the variation of sampling distributions to determine the likelihood of drawing samples with particular statistics from a population with hypothesized parameters—is dependent on being able to reason probabilistically and having multiplicative conceptions of samples and sampling. Consideration of variation in samples and sampling distributions seems necessary to reason probabilistically.

Reasoning About Variability in Experimental Design

Little research exists to inform one about how students or teachers may reason about variation from a design perspective when designing experiments, but researchers do provide glimpses into how individuals anticipate variability when designing experiments. In a teaching experiment designed to have students consider error, or variation, as arising from multiple sources like measurements, instruments, and replications in experimental design, fourth graders compared rockets with different physical features and explored whether differences in rockets' achieved heights could be attributed to random error or were indicative of systematic error in rocket types (Petrosino, Lehrer, & Schauble, 2003). During their classroom discussions, students were able to use their collected data to suggest systematic error and thus displayed sophisticated reasoning about variation. Not emphasized in the teaching experiment were ways in which to control various sources of error—particularly ways to control random error.

Few students consider random assignment as a strategy to control variation, even when they are enrolled in an introductory course that emphasizes the role of randomization (Derry, Levin, Osana, Jones, & Peterson, 2000). Consideration of methods to control systematic and random variability involves sophisticated reasoning—reasoning that was rarely seen in Groth's (2003) study to investigate secondary students' understanding of experimental design. Although students may struggle to design experiments that control variation from different sources, even students at the elementary level have been observed to recognize sources of error in data (Masnick & Klahr, 2003). Necessary for designing experiments are both consideration of sources of variation and consideration of ways to control variation from those sources.

Coordinating between-group (systematic) and within-group (random) variation involves what Reid and Reading (2005) label as "strong consideration of variation." The researchers see the coordination between systematic and random variation as a first step toward recognizing the link between variation and formal inferential statistics. They suggest that students may need time and instruction beyond an introductory statistics course, even a course focused on variation, to reason with a strong consideration of variation in a wide variety of contexts. While it appears that reasoning about variability from a design perspective is difficult for students and teachers, proper study design requires reasoning about sources that may introduce variability to data if left uncontrolled.

Students' and Teachers' Reasoning From a Data-Centric Perspective

After a study is designed and data are collected, statisticians typically engage in exploratory data analysis to investigate possible patterns of variability in data and relationships among variables. For students' initial explorations in statistics, however, students typically begin with exploratory data analysis rather than design (e.g., Moore, 1999). Much of the research to examine students' and teachers' reasoning about variation falls under the umbrella of reasoning from a data-centric perspective, which includes reasoning about representing, measuring, and describing variability.

Representing Variability

Widespread availability of technology and easy access to software applications has caused a new instructional focus to emerge for creating and interpreting data representations (Friel, 2008). Rather than spending hours to create graphical displays, students and teachers are able to use technology to create data representations easily and to compare information about variation and patterns of variability revealed in or obscured by different representations, among other possible comparisons. Garfield and Ben-Zvi (2005) consider creating and examining multiple representations of data to reveal different aspects of variability as a necessary characteristic for understanding variation, and Wild and Pfannkuch (1999) coined the term *transnumeration* to describe statistical thinking embodied by the "*dynamic* [italics in original] process of changing representations to engender understanding" (p. 227). Research highlights particular behaviors characteristic of transnumeration.

Two broad categories of behavior related to graphical comprehension seem to align with reasoning about variation from the data-centric perspective: translation and interpretation (Friel, Curcio, & Bright, 2001). Translation involves representing and reading data by changing the form of data to extract descriptive information about the data (Curcio, 1987), whereas interpretation includes rearranging data and using additional representations to interpret and identify trends in data and to reason about variability both within and away from the trend (Friel, Curcio, & Bright, 2001). Both behaviors incorporate elements of representing and describing variability, with interpretive behavior aligned more closely with the behavior of statisticians. Behaviors similar to translation and interpretation have been observed in the activity of middle school students (Ben-

Zvi & Friedlander, 1997). Researchers found that students who translate data from what students perceive to be meaningful representations tend not to reorganize data to explore additional patterns or use summary measures of data to interpret results and thus may overlook important characteristics of data. Students who meaningfully handle multiple representations exhibit interpretive behavior and display transnumeration (Wild & Pfannkuch, 1999) in that they use multiple data representations to uncover meaningful characteristics of data.

Computer tools like TinkerPlots™ Dynamic Data Exploration (Konold & Miller, 2004) facilitate the mechanics of transnumeration by enabling quick and easy creation of multiple conventional and unconventional graphical displays of data. Middle and high school mathematics teachers have been observed using TinkerPlots to compare variability in two groups of data by graphically dividing data into equally spaced bins (Hammerman & Rubin, 2004). These teachers represented and handled "variability by [arranging data and] finding subsets of the data about which they [could] make more deterministic claims" (Hammerman & Rubin, 2004, p. 35). Binning supported teachers' propensity to reason deterministically, which supports the view that without proper training, mathematics teachers may not develop the ability to think probabilistically and thus may apply their deterministic beliefs about the nature of mathematics to statistics (Meletiou-Mavrotheris & Stylianou, 2003).

Measuring Variability

Translation and interpretation are behaviors that encompass reasoning about more than only graphical displays of data, as another representation of data exists in summary measures, such as standard deviation, that describe representative global characteristics of data, such as spread. Konold and Pollatsek (2004) note the inseparability of measures of average and variability, and reasoning about average and variability merge in reasoning about the spread of data relative to center—distributional reasoning (Shaughnessy, Canada, & Ciancetta, 2003). Reasoning about spread relative to a center coupled with thoughtful consideration of formal measures of center and spread to reason about data are marks of sophisticated statistical thinking (delMas & Liu, 2005; Groth, 2005) and seen as necessary for deep understandings of variation (Garfield & Ben-Zvi, 2005).

Reasoning About Measures of Variation

Although recent research suggests that school students have intuitive conceptions of variability and are able to reason about the range of data and the spread of data relative to a center (e.g., Reading & Shaughnessy, 2004; Shaughnessy, Ciancetta, Best, & Canada, 2004), little research has been conducted to investigate school students' measuring of variation with measures different from the range. Research has shown that students exhibit improved reasoning about variability as they study ideas related to data and chance throughout their educational years (e.g., Kelly & Watson, 2002; Watson, 2002; Watson, Callingham, & Kelly; 2007; Watson & Kelly, 2002a, 2002b, 2003a, 2003b, 2004a, 2004b, 2005). Despite their improved reasoning, however, students appear to struggle to move beyond intuition and, in particular, encounter difficulties with reasoning about variation using formal measures of variation.

Garfield, delMas, and Chance (2007) incorporated activities specifically designed to advance students' reasoning from informal reasoning about variation to reasoning about variation with formal measures of variation in their college-level introductory courses. When the courses ended, their students were only beginning to consider variation as a measure of spread from center and display advanced understandings of variation. Garfield and colleagues note that their students were not adept at applying their knowledge of variation to novel situations and thus fell short of exhibiting deep understandings of variation. Other college-level introductory statistics students examined agreement among a number of measures of variation to reason about formal measures of variability in group comparison tasks (Lann & Falk, 2003). The students observed by Lann and Falk seemed to look for rules to describe and compare data variation in place of choosing measures for comparison based upon characteristics of data. A similar search for rules was observed as different introductory-level students compared standard deviations for multiple pairs of distributions by attempting to create rules to generalize patterns of histogram bars to make comparisons (delMas & Liu, 2005). delMas and Liu noted that very few students employed a conceptual approach to coordinate the location of the mean, estimated by using characteristics of the distribution, with deviations from the mean. The body of this research with introductory-level statistics students reveals that even though many students are able to reason informally about variation, they may not be coordinating their intuitions about variation with their knowledge of formal measures of variation to reason about distributions of data and to make comparisons between distributions. It seems that even if individuals study statistics formally, they exhibit a tendency to employ rule-based approaches to reason about variability.

Although the tertiary students whose reasoning and understandings were described in this section arguably may have less sophisticated mathematical understandings than preservice secondary mathematics teachers, there is no reason to believe that teachers do not experience the same difficulties. For example, few of the prospective science and mathematics teachers participating in Makar and Confrey's (2005) study compared data sets by using standard deviation. The researchers observe that "it would appear that the notion of standard deviation as a measure of variation did not hold much meaning" (p. 38). Studies with preservice teachers provide little evidence to suggest that many preservice secondary mathematics teachers understand the formal measures of variation as anything more than numerical values or as computations (e.g., Makar & Confrey, 2005; Sorto, 2004).

Reasoning About Measures of Center

For many teachers, even if they are able to calculate a value for standard deviation and discuss standard deviation as a measure of variation, they may be unable to reason about standard deviation in conjunction with the mean (Clark, Kraut, Mathews, & Wimbish, 2007; Silva & Coutinho, 2006). Part of this difficulty might stem from an impoverished understanding of mean. Research investigating elementary-aged through college-aged students' conceptions of average and mean reveals some of the same struggles that students exhibit in their conceptions of variation—many are able to calculate numerical summary values without understanding the meaning of their results (e.g., Clark, Kraut, Mathews, & Wimbish, 2007; Mokros & Russell, 1995). Research with experts suggests that a deep understanding of the mean includes understandings of both the algorithm for the arithmetic mean and the arithmetic mean as a mathematical point of balance (MacCullough, 2007). Experts not only use the algorithm to calculate a value for the mean but also understand the meaning of the operations within the algorithm and the nature of the results. Captured within understandings of the algorithm is the notion of the average as a representative value for a set of data.

Inservice and preservice secondary teachers' conceptions of the mean and of average are similar to those seen from students. Although preservice and inservice secondary mathematics and science teachers may use the computational algorithm to calculate values for means, teachers struggle to conceive of the mean in multiple ways (Gfeller, Niess, & Lederman, 1999), to apply the mean to higher-level problems (Gfeller, Niess, & Lederman, 1999), and to estimate values for the mean from graphical representations of data (Callingham, 1997; Sorto, 2004). In short, research suggests that many teachers have little conceptual understanding of the mean, which has implications for their understanding of variation. If an understanding of standard deviation requires a dynamic conception of distribution that coordinates changes to the relative density of values about the mean with their deviation from the mean (delMas & Liu, 2005), then it would appear that an understanding of mean as a mathematical point of balance is needed to reason about estimating a value for the mean of a set of data, particularly when displayed graphically by dotplots or histograms, and to reason about standard deviation. In particular, reasoning about how individual data values affect values for the mean and standard deviation is useful for detecting data entry errors that may affect both the mean and standard deviation.

Reasoning About Distribution

Students who understand the algorithm for the mean and who are able to view the average as a representative value for a set of data seem to have a view of the average as a value that represents the data distribution as an entity. Researchers have identified the importance of students' developing an aggregate view of data—being able to view data in terms of the whole distribution—for reasoning about data and variability in data (e.g., Ben-Zvi & Arcavi, 2001; Hancock, Kaput, & Goldsmith, 1992; Konold, Harradine, & Kazak, 2007).

The ability to view data as a single aggregate collection of values rather than as a collection of individual values seems to be needed for understanding distribution, and an understanding of distribution seems to be needed to reason about variation. Wild (2005) describes distribution as the "pattern of variability in a variable" that "underlies virtually all statistical ways of reasoning about variation" (p. 4). Viewing data as an aggregate focuses on patterns of variability, which includes notions of shape, center, and spread, whereas viewing data pointwise allows for calculation of summary values such as the mean, median, range, interquartile range, and standard deviation and consideration of individual deviations from the pattern (Bakker & Gravemeijer, 2004). Statistical "experts" are capable of moving flexibly between pointwise and aggregate views of data, and understanding distribution from this dual perspective seems to lay foundations for reasoning about variation within and between groups to compare data collected from two or more groups. A dual perspective of distribution seems to align with views that understandings of distribution can be enhanced by viewing data as a "'distribution around' a signal" (Konold & Pollatsek, 2004, p. 171), whereby the notion of central tendency embodies the idea of signal and variation embodies the idea of noise. Konold and Pollatsek conjecture that interpreting average as "signal in noise" (2004, p. 177) is a useful interpretation for making group comparisons.

Describing Variability

Investigations to explore students' reasoning about variation from the data-centric perspective largely utilize tasks that focus on group comparisons. Although variation can be described by summary measures, many students and teachers do not seem to recognize the utility in using summary measures to compare or describe groups. For example, preservice secondary mathematics and science teachers appear to struggle in applying their knowledge of the mean to make comparisons between two groups of data (Makar & Confrey, 2003)—they struggle with viewing the mean as a representative value for a set of data. The preservice secondary mathematics and science teachers in Makar's (2004) study seem to prefer describing variation with non-standard language to make comparisons. Makar claims that these teachers learned statistics concepts, including variation, but chose to express their understanding in informal terms. Although their nonstandard language at times revealed sophisticated reasoning, Makar notes that the informal nature of their conceptions may prove to be insufficient for applying the concepts in future statistical study, suggesting that the teachers may face difficulties in applying their understandings of the concepts to second-order concepts such as sampling distribution. Even if teachers focus on the center and variation within groups—with or without formal measures—

they may still struggle to use this information to make comparisons between groups. The secondary mathematics teachers investigated by Makar and Confrey (2002, 2004) readily described center and variation within groups but did not apply their knowledge of variation to describe the variation between groups in order to reason about the existence or nonexistence of a difference in groups.

The totality of research that investigates secondary teachers' descriptions of variation suggests that even after studying formal measures to describe variation and formal inferential techniques for comparing distributions, teachers prefer using informal reasoning about variation to make comparisons. The teachers who participated in these studies either chose not to reason or could not reason about variation using formal measures and techniques, suggesting that impoverished understandings of variation may be at the center of their difficulties. Garfield and Ben-Zvi (2005) suggest that deep understandings of variation are partially exhibited when individuals use global summary measures of variation to compare groups and include examinations of and distinctions between within-group and between-group variation in their comparisons. In their work to examine tertiary students' consideration of variation, Reid and Reading (2008) considered linking within-group variation to between-group variation to make inferences from data as the difference between students exhibiting strong considerations of variation and those exhibiting developing considerations of variation.

Students' and Teachers' Reasoning From a Modeling Perspective

Wild and Pfannkuch (1999) tell us that an important consideration of variation involves modeling the variation in data "for the purposes of prediction, explanation, or control" (p. 226)—ideas that entail reasoning about variation from the modeling perspective. The modeling perspective forms the basis for inferential statistics in that it involves viewing data in comparison with some theoretical model, including binomial, normal, and linear models.

Variation and Binomial Models

Although many individuals would think about modeling a sampling distribution with a normal distribution for statistical inference, there are other models that seem to be more approachable for students. In the context of tossing a die, for example, third grade students were able to suggest that results of a specified number of die tosses can vary and are likely to vary (Watson, 2005). They were able to informally hypothesize binomial models for toss outcomes by exhibiting appropriate deviation from the expected value for outcomes. These third graders displayed informal reasoning about data that could be modeled by a binomial distribution.

Watson and colleagues investigated students' reasoning about data and chance as students advanced through grade levels in the data and chance curriculum (Kelly & Watson, 2002; Watson & Kelly, 2002a, 2002b, 2003a, 2003b, 2004a, 2004b, 2005). They observed that students are increasingly able to intuit characteristics of binomial distributions to make conjectures about die toss results with appropriate deviation from expectation for a specified number of tosses. Students at higher grade levels are more likely to respond based on probabilistic expectation or with too little variability in tosses, and some students express being torn between expected values based on probability and their expectations for varying results (Reading & Shaughnessy, 2000; Shaughnessy, Ciancetta, & Canada, 2004). Fischbein and Schnarch (1997) also employed the use of a task with a binomial setting to investigate "the evolution of probabilistic misconceptions as an effect of age" (p. 101). Their task requires students to consider the likelihood of results as opposed to hypothesizing results and thus requires greater sophistication in reasoning than the die toss problems. Fischbein and Schnarch's problem can be answered by using a binomial model to calculate probabilities or by using a normal model to approximate a binomial distribution. What the researchers found is that the *mis*conception that sample size is irrelevant occurred more frequently with increased age. The authors posit that individuals tend to believe that ratios should be used to solve binomial probability problems and thus fail to use the law of large numbers as appropriate for the situation. The preservice mathematics teachers who participated in the study apparently had no background in statistics, and their mathematical experiences and intuitions did not seem to help them in this setting. At the very least, this study highlights the importance of being able to reason about the effects that sample size can have on results.

More recently, a version of the same problem was given to preservice secondary mathematics teachers (Watson, 2000). While slightly more than half of the preservice teachers responded correctly to the problem, few combined intuition with mathematical justification. Teachers who correctly set up a calculation to solve the problem but failed to reach a correct solution due to an arithmetic error did not seem to notice any problem with their solutions, whereas other teachers entirely relied on intuition rather than calling on their formal background in mathematics to reason towards a solution. Many of the teachers did not seem to be aware that, in general, as sample size increases, empirical relative frequencies approach theoretical probability. It would seem that part of reasoning about variation from a modeling perspective entails being able to reason about the effects of sample size on variability in novel problem contexts. These studies suggest one reason why students seem to have difficulty in reasoning about sampling distribution, in that the reasoning needed for sampling distribution appears to be counterintuitive.

Variation, Models, and Statistical Inference

As a concept that underlies most areas of statistics, variation plays a role in developing students' understandings for informal and formal inference. Formal inferential reasoning requires reasoning with the concepts discussed in previous sections, including reasoning with and about center and measures of center, variation and measures of variation, distribution, sampling, and probability to meaningfully draw conclusions about a population from a sample selected from that population (Pfannkuch, 2005). Success in understanding formal statistical inference beyond scripted steps for calculating a *p*-value or finding a confidence interval requires reasoning about the concept of sampling distribution—a concept for which understanding seems elusive for many students of introductory statistics.

In studies designed to investigate students' ability to interrelate the ideas of variability, sampling, and sampling distribution, researchers note students' tendency to confuse a distribution of a sample with a distribution of sample means (e.g., Saldanha & Thompson, 2002). Students struggle to reason about the variation of individual observations in a sample and the variation of sample means in a sampling distribution (e.g., Garfield, delMas, & Chance, 2007; Meletiou-Mavrotheris & Lee, 2003). To understand sampling distribution, individuals need to juxtapose "the individual sample result against an aggregate of similar sample results to compare the one against the many" (Saldanha & Thompson, 2002, p. 267)—the multiplicative conception of sample and sampling noted earlier. Students with multiplicative conceptions are able to reason proportionally about the likelihood of sample results (Saldanha & Thompson, 2001) by examining a distribution of a simulated collection of sample statistics, for example. Modeling-based activities that include simulation are being investigated to determine their viability in aiding students' constructions of foundational knowledge from which they can build more formal understandings of statistical inference (Konold, Harradine, & Kazak, 2007). Research suggests that a critical juncture for students occurs when they attempt to link a simulated sampling distribution to a theoretical sampling distribution (Lipson, 2002). Lipson found that the students in her study had difficulty transitioning from a computer simulation to formal inference. Being able to complete this link to successfully reason about sampling distribution lays the foundation for reasoning formally about inferential methods.

Research on Variation: Concluding Remarks

As the body of literature discussed in the preceding sections suggests, research that examines students' reasoning about the concept of variation and related concepts reveals that students have many intuitions about variation and concepts related to variation. As a whole, however, these studies reveal that despite students' improved reasoning about variation as they progress through grade levels with appropriate instruction, most students continue to express only intuitive understandings of variation. The limited body of work to investigate teachers' reasoning and understanding of statistics concepts suggests that teachers have difficulties similar to those identified for students and struggle to construct both procedural and conceptual understanding of statistical concepts as well as to identify connections between and among concepts. Developing understandings of variation and applying knowledge of variation to problem solving is problematic for both students and teachers. What is abundantly clear, however, is that variation connects to and interrelates with many concepts in statistical study, which suggests that understanding variation is critical for understanding statistics and for recognizing the utility of statistics.

Research About Teacher Development

One focus of current mathematics education research is knowledge required for teaching. Whereas teacher knowledge has been a subject of mathematics education research for a number of years, research to investigate the impact of teacher knowledge on student achievement in ways that go beyond using proxy measures for teacher knowledge, such as the number of undergraduate mathematics courses completed (e.g., Monk, 1994), is a relatively new phenomenon. In one of the few studies to examine the connection, Hill, Rowan, and Ball (2005) "found that teachers' mathematical knowledge for teaching positively predicted student gains in mathematics achievement" (p. 399) for the first and third graders included in their study, providing support for a prevalent belief that teacher knowledge affects student achievement in mathematics.

In addition to content knowledge, teachers should display pedagogical knowledge as well as other types of more delineated knowledge to be successful in teaching. Ma (1999) references a need for teachers to have "profound understanding of fundamental mathematics." Shulman (1986) makes distinctions among content knowledge, pedagogical knowledge, and pedagogical content knowledge, the knowledge of pedagogy unique to a content area, as necessary for teachers. There also is work to suggest that teachers' pedagogical content knowledge supports student learning (e.g., Krauss, Baumert, & Blum, 2008). Somewhat overlapping with other knowledge types, Hill and Ball (2004) add mathematical knowledge for teaching, which includes "common content knowledge but also [the] specialized knowledge for teaching mathematics" (p. 335). This specialized knowledge includes understanding how and why procedures work in addition to being able to apply procedures as well as the deep understanding of mathematics needed to understand and react to students' sometimes unconventional mathematical understandings and processes.

Groth (2007) argues that the knowledge needed to teach statistics differs from the knowledge needed to teach mathematics based in the differences between mathematics and statistics. The differences he hypothesizes stem from differences between deterministic and stochastic reasoning, designing studies, considering context, and distinguishing between practical significance and statistical significance. These differences seem to align with differences between the art and science of statistics (Peters, in press). Although much of the research cited here examines the knowledge required to teach mathematics, at a minimum, there is very little reason to believe that the types of knowledge required to teach scientific aspects of statistics differ significantly from that required to teach mathematics, particularly since a considerable amount of research investigates knowledge requirements for teachers in multiple subject areas, including pedagogical content knowledge for teaching science (e.g., van Driel, Verloop, & de Vos, 1998), practical knowledge for teaching language (e.g., Meijer, Verloop, & Beijaard, 1999), and teacher knowledge in action for teaching social sciences (e.g., Department of Education Training and Youth Affairs, 2000).

Professional Development

Teachers attempting to reform their teaching practices in mathematics express how their limitations in knowledge impact their ability to enact national, state, and local educational recommendations in the area of mathematics (Peterson, 1990; Spillane, 2000b; Wilson, 1990). Researchers also note how science and mathematics teachers’ knowledge affects their ability to align their practices with the practices called for by reform efforts (Borman & Kersaint, 2002; Firestone, Mayrowetz, & Fairman, 1998; Spillane, 2000a, 2000b). Given the perception that teachers have few experiences with statistics content (e.g., Shaughnessy, 2007) and receive little preparation for teaching statistics (e.g., Garfield & Ben-Zvi, 2008), it seems reasonable to believe that teachers' knowledge of statistics can impact their abilities to teach the statistics content they are being asked to teach.

In their work to examine Australian teachers' knowledge of data and chance, Watson and colleagues (Callingham, Watson, Collis, & Moritz, 1995; Watson, 1998) concluded that many issues related to teacher training need to be addressed to produce a statistically literate society. Even when prospective secondary mathematics teachers participate in training focused on integrating statistics with undergraduate methods courses, they "struggle with the 'spirit' of statistics" (Burrill, 2008, p. 3). As an example of struggle, consider the results of a study conducted by Coutinho (2008). Although the teachers in the study espoused using exploratory approaches with data, they enacted what Coutinho termed a "technicist approach" in their classrooms—they focused attention on algorithms to calculate summary values from data rather than true explorations of data. Given teachers' difficulties in teaching mathematics consistent with recommendations, this collection of work suggests that teachers may need extensive professional development to develop multifaceted understandings of statistical concepts and procedures in order to teach statistics in ways consistent with recent recommendations.

Researchers have conducted studies to examine the characteristics of "high quality" professional development, with a number of characteristics identified in this collective body of work. Of those who have studied the impact of professional development on mathematics teachers' practices, there is agreement about the necessity for professional development to be content-focused and sustained (e.g., Cohen & Hill, 1998, 2000, 2001; Darling-Hammond & Ball, 1998; Goos, Dole, & Makar, 2007; Smith, Desimone, & Ueno, 2005). There is also some suggestion that professional development should focus on curriculum (Cohen & Hill, 1998), students' work on tasks from that curriculum (Darling-Hammond & Ball, 1998), and other artifacts from practice, including instructional tasks (Ball & Cohen, 1999). Recommendations for professional development specific to statistics include providing teachers with opportunities to experience as learners the statistical content they are expected to teach enacted with the pedagogical strategies they are expected to use (e.g., Heaton & Mickelson, 2002; Lee & Hollebrands, 2008; Peck, Kader, & Franklin, 2008). Mathematics educators make similar recommendations for the learning of mathematical content (e.g., Artzt, Curcio, & Sultan, 2004), and teachers see opportunities in which they can actively learn in ways similar to their students as characteristic of effective professional development (Rogers et al., 2007). Teachers also view opportunities to interact with teachers from other schools as beneficial in ways beyond what engagement with formal activities can offer (Rogers et al., 2007)

Although the preceding description largely focuses on qualities of professional development that occur in formal educational settings, teachers in all areas experience professional development through self-initiated and self-directed efforts, such as initiating the use of innovative curriculum materials (e.g., Lohman & Woolf, 2001). Studies that investigate the collective impact of teachers' learning could benefit by focusing not only on formal learning opportunities experienced by teachers but also on types of self-directed learning opportunities. Gaining knowledge about the characteristics of both informal and formal activities that promote meaningful teacher learning seems important for informing future professional development and teacher education initiatives in statistics.

Teacher Beliefs

In addition to knowledge, teacher beliefs may affect the pedagogical strategies employed in classrooms and affect students' subsequent learning. Because research to investigate teacher beliefs related to statistics is almost nonexistent (Watson, 2001), particularly for secondary-level teachers, research related to teacher beliefs about mathematics is examined and followed by speculation for how teacher beliefs about mathematics might compare with teacher beliefs about statistics. Research suggests that teachers' beliefs about mathematics and beliefs about mathematics teaching and learning may impact their subsequent decisions about what content is taught and how that content is taught (Thompson, 1992). Teachers' beliefs about statistics and about statistics teaching and learning presumably could have a similar effect on their decisions regarding statistical content.

Researchers studying the impact of national and state efforts to reform teaching practices note how a teacher's beliefs can affect his or her classroom practice. For example, a teacher who views mathematics as a collection of fixed procedures used to arrive at a single correct answer may feel that his or her practice is consistent with the calls for reform and yet enact classroom discourse that ignores mathematical explanation, justification, and argumentation (Cohen, 1990). Given speculation that teachers are likely to apply their beliefs about the nature of mathematics to statistics, that same teacher may enact classroom discourse that ignores explanation, justification, and argumentation situated within the context of data, particularly if that teacher as a learner experienced statistics as a mathematical topic that focuses on computations and procedures (Cobb, 1999; Gal & Garfield, 1997). That teacher may neglect the issues of uncertainty and variability inherent to statistics (Meletiou-Mavrotheris & Stylianou, 2003). Unless teachers make distinctions between statistics and mathematics, even if a reformed view of mathematics teaching is adopted, the need to reason within a context and in consideration of variation may prevent the teacher from viewing statistics as the "science of uncertainty" (e.g., Tabak, 2005). Teachers may still teach statistics deterministically.

Another teacher may exhibit a strong belief system that he or she readily admits may affect his or her willingness to make changes in practice. For example, a teacher who believes that an understanding of mathematics consists of the mastery of symbols and procedures may use reform recommendations to guide the mathematical topics addressed in class, but refuse to incorporate new ideas for how mathematics should be taught (Wiemers, 1990). Similarly, a teacher who views statistics in procedural terms might ignore new ideas for the teaching of statistics. It is possible for a teacher to profess a desire to teach for understanding, believing that students learn through engaging actively using concrete or physical manipulatives, using technological applications, and comparing multiple representations for learning mathematics concepts and procedures and yet not teach in ways consistent with reform. If the teacher presents solutions by using models or representations but makes the connections between the model and symbols for students, envisions active engagement in terms of students' physical activity rather than mental activity, or believes that only one right way exists to arrive at any answer, this teacher has not made substantial progress towards enacting a practice consistent with reform recommendations (Ball, 1990). Similarly, a teacher could have students collect data in the classroom, such as having students time how long they can hold their breaths, follow a step-by-step procedure to enter data into a calculator or software package, and follow a step-by-step procedure to produce summary statistics and various graphical representations. The teacher could then describe the connections among the summary measures and the various graphical displays, suggesting a parallel possibility for teaching statistics. These examples illustrate that even though there are some fundamental differences between mathematics and statistics, the effects of teachers' beliefs about mathematics and the practice of teaching mathematics may be similar to the effects of their beliefs about statistics and the practice of teaching statistics.

Professional Development, Teacher Beliefs, and Teacher Change

In general, professional development in education typically is designed to effect change in teachers' beliefs, practices, or knowledge (Guskey, 2002). Historically, professional development efforts in education focused on a transmission model of teaching and learning that assumed developers could present knowledge and pedagogical strategies to teachers, who would then replicate the techniques in their classrooms (Richardson, 1998). Activities for this type of professional development consisted of conference or workshop attendance that ultimately served to pique a majority of teachers' curiosities at best (e.g., Cohen & Hill, 2000; Desimone, Porter, Garet, Yoon, & Birman, 2002). Research literature for teachers of mathematics, science, and English suggests that these approaches are largely ineffective for changing what teachers teach or how they teach (Boyle, White, & Boyle, 2004). This traditional view of professional development suggests a view that assumes that teachers' beliefs and attitudes will change in response to professional development and that teachers subsequently will transform their classroom practices for the result of improved student learning (Guskey, 2002). Changes in beliefs are a significant predictor for changes in practice, and even though teachers readily make superficial changes (e.g., changes in classroom organization), research suggests that deeply embedded implicit beliefs are much more difficult to change (Richardson & Placier, 2001).

A related perspective on teacher change posits that changes in classroom practices result in increased student learning, which then prompts changes in attitudes and beliefs (Guskey, 2002; Nathan & Knuth, 2003). Proponents of this view suggest that beliefs significantly change only after evidence of student improvement exists and subsequent to changes in practice. The proponents of this view acknowledge, however, that some change in attitude or beliefs necessarily precipitates a change in practice (Guskey, 2002), suggesting that the process of change may not be linear.

Another view of teacher change suggests the process is much more complicated than the described views might suggest. For example, researchers from the Cognitively Guided Instruction (CGI) teacher development program found little consistency between whether a teacher's change in practice precipitated a change in beliefs or vice versa (Fennema et al., 1996). In one of the studies conducted by this group, they found that 6 of the 21 elementary teachers enrolled in their program changed their beliefs before their practices. For five of the teachers, their practices seemed to prompt a change in beliefs, and for six teachers, changes to beliefs and practice occurred simultaneously (Fennema et al., 1996). Other studies suggest that teachers may enact new practices in their classrooms without a corresponding change in beliefs, particularly when the new practices are seen as consistent with already existing beliefs (e.g., Nathan & Knuth, 2003). In keeping with Fennema and colleagues' (1996) comments from more than a decade ago, the current body of work suggests that research has not yet provided the key for ascertaining why some teachers change their beliefs or practices whereas others do not and why some teachers are able to change their beliefs or practices more than other teachers.

Research suggests that national, state, and local policies that attempt to effect a change in teachers' practices require many teachers to alter their beliefs about teaching and learning (e.g., Spielman & Lloyd, 2004). Many current teachers experienced teacher-directed instruction as students (Stigler & Hiebert, 1999), which created what can be called a "culture of schooling" in the United States (Stigler & Hiebert, 1999; Weissglass, 1992). Historically, many policy attempts to alter this culture failed to provide teachers with the direction or tools necessary for them to make the recommended changes in their practices (Cohen & Ball, 1990; Spillane, 2002). How, then, can this culture be changed? One hypothesis is that changing the "'culture of schooling' will occur, if it occurs at all, in the context of identifying and discussing values and beliefs about all school practices, listening to and grappling with views that are different from our own, and working through feelings and attitudes that inhibit change" (Weissglass, 1992, p. 198). It seems that changing the predominant "culture of schooling" requires teachers to reflect critically on values and beliefs and to engage with others in rational discourse about their beliefs. Such reflection and discourse would need to occur over time and may eventually result in classroom instruction that is reflective of transformed beliefs and values.

Weissglass (1994) proposes a model that addresses teachers' feelings and beliefs during the process of change just described, as shown in Figure 2-1. This model seems to capture some of the complexity and nonlinearity of the teacher change process and suggests that the element of emotional support ignored in other models may be important for teachers to accomplish change. Four essential components form the model to suggest that teachers continually need to learn about content and pedagogy, reflect on their beliefs and practice, and obtain emotional support from colleagues in planning for and enacting change. These processes culminate with the teacher taking action to make changes in practice, which precipitates the need for further reflection and emotional support, and so on.

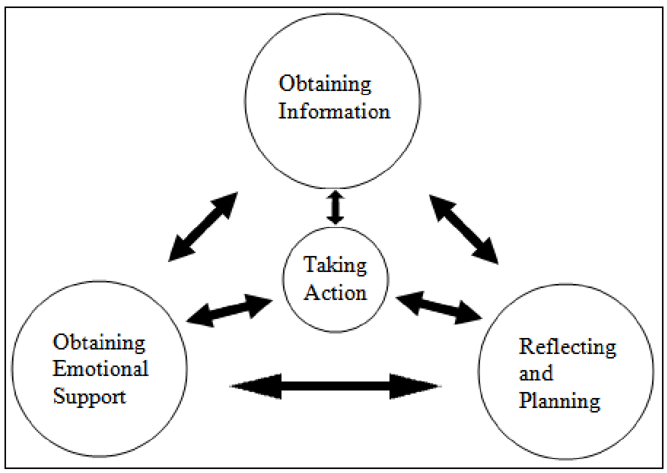

Figure **2-1**: Model for Addressing Teachers' Feelings and Beliefs (Weissglass, 1994, p. 70).

Whereas the model posited by Weissglass is important because it captures the nonlinear pattern of teacher change and includes important components of change such as reflection, the need for emotional support, and action, the model is somewhat limited in that it does not suggest how the change process begins. Clarke and Hollingsworth (2002) offer a model of professional growth that accounts for both internal and external stimuli that may precipitate the process of change, as shown in Figure 2-2. Their empirically derived model consists of four domains of change, which when viewed holistically portray a teacher's professional world of practice. The external domain refers to information or stimuli the teacher encounters from external sources, while the remaining domains are part of the teacher's individual world. A teacher's knowledge, beliefs, and attitudes form the teacher's personal domain, whereas the domain of consequence consists of the teacher's inferences about the salient outcomes of practice. Professional experimentation, which includes but is not limited to classroom experimentation, forms the domain of practice. The model also depicts two mechanisms for change: enactment and reflection. A teacher's changes take place within a school environment, which can help or hinder the change process at any point. The change environment contributes to a teacher's affective disposition.

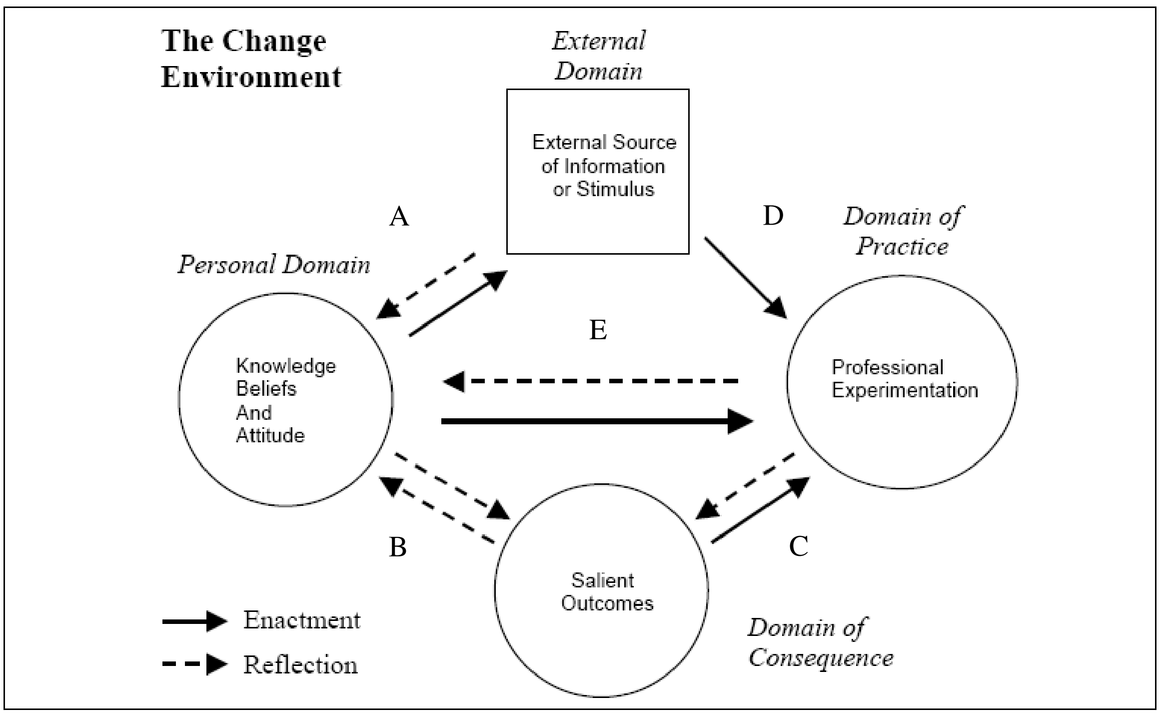

Figure **2-2**: Adaptation of a Model of Professional Growth (Clarke & Hollingsworth, 2002, p. 951).

Clark and Hollingsworth's model of professional growth embodies each of the previously discussed views and models. In the traditional view mentioned at the beginning of this section, professional development presents an external stimulus that upon reflection brings about new knowledge, beliefs, or attitudes (A). The teacher then enacts changes in his or her practice (E) that the teacher infers will result in salient outcomes (C). In contrast with Weissglass' model, the external domain aligns with information gathered from others, and the domain of practice aligns with the process of taking action. Rather than considering reflection as a separate component of change, reflection appears throughout the model of Clark and Hollingsworth, and emotional support can be part of the change environment or action on the external domain if the teacher deliberately seeks emotional support from others. The professional growth model is also consistent with suggestions for change from the professional development and research literature. Furthermore, the model is consistent with the adult learning theory known as transformation theory (Mezirow, 1991, 2000) that forms the theoretical frame for this study and is discussed in detail in Chapter 3.

Literature in both professional development and research contains various suggestions for creating an environment that is conducive to teacher change. These suggestions include providing opportunities for teachers to interact during inservice programs and purposefully planning teachers' schedules to allow time for interaction during the school day (Weissglass, 1994). Interaction gives teachers the opportunity to form support networks and provides an opportunity for teachers to engage in rational discourse with colleagues (Saavedra, 1996; Weissglass, 1994). Opportunity and time for teachers to reflect critically on their practice (Weissglass, 1994) and to develop the skills and knowledge needed for change seem crucial for change to occur and supports calls for extended professional development opportunities (e.g., Senger, 1998-1999). Along with the aforementioned forms of administrative and systemic support, teachers also need access to resources for successful change to occur (Lohman & Woolf, 2001).

In response to a need for change in teacher education related to teaching statistics, guidance exists to suggest needed prerequisite knowledge (CBMS, 2001). Additionally, various organizations provide suggestions for appropriate pedagogical strategies and techniques for teaching statistics, along with a logical progression of topics for curricula (e.g., Franklin et al., 2007). These recommendations stem largely from their authors' beliefs about effective statistics instruction, rather than from the results of research on teaching and learning. Researchers paid little attention to teachers' understanding of statistical concepts or teachers' professional development needs for teaching statistics prior to the release of the 1989 NCTM Standards (Konold & Higgins, 2003). Since that time, calls for research on teachers' conceptions of statistical concepts have been made (Shaughnessy, 1992; Shaughnessy, Garfield, & Greer, 1996) but such research is not sufficiently reported (Batanero, Garfield, Ottaviani, & Truran, 2000). Calls have also been made to "establish effective ways of training current and future teachers of statistics" (Batanero, Garfield, Ottaviani, & Truran, 2000, p. 5), yet little research conducted with teachers of statistics has been reported (Batanero, Burrill, Reading, & Rossman, 2008). At the current time, it makes sense to examine the learning opportunities and support experienced by current statistics teachers who exhibit robust understandings of fundamental statistical concepts in order to provide timely information for the design of professional development programs to train current and future teachers of statistics. It also makes sense to examine these opportunities through a theoretical frame that is consistent with teacher learning as described in the professional development literature and adult learning theory to allow for differences in how adults learn from how children learn.

Concluding Remarks

Much research exists to illuminate how students from the elementary grades through the undergraduate level learn about statistical concepts, and in particular, recent research has greatly expanded what we know about students' conceptions of and learning of variation. Most of this research, however, investigates students reasoning informally about variation and statistical content more generally. Few reports of research describe how experts may think about or learn statistical content. When the results of research are combined with statisticians' and statistics educators' expositions about the multifaceted concept of variation, however, a sense of what it means to deeply understand statistical variation emerges. Discussion of the initial framework for robust understanding of variation that arose from the synthesis of literature presented here is discussed in Chapter 4.

As the absence of research devoted to how advanced knowers think about or learn statistical content might indicate, there is little work that directly informs how advanced knowers learn about complicated concepts such as statistical variation. The wealth of literature surrounding teacher learning and teacher change, however, provides insights into personal and environmental factors that may affect teacher learning in general and thus affect teacher learning about statistical variation in particular. The literature on teacher change in particular provides insights into important components of change in addition to suggesting events, activities, or conditions that may trigger the process of change and the types of nonlinear paths among components that teachers may take during their process of change. A learning theory appropriate for describing teacher learning should be consistent with these literature-based observations. Transformation theory, discussed next in Chapter 3, is consistent with the models presented here and provides explanatory power for teachers' learning about statistical variation.

Chapter 3

Conceptual and Theoretical Grounding

Transformation Theory

Transformation theory is a theory of adult learning that may have explanatory power for learning that results in teacher change. Primarily credited to Jack Mezirow (Merriam & Caffarella, 1999), the theory is based on constructivist assumptions, including the assumptions that meaning resides within each person through personal constructions and that personal meanings are acquired and confirmed through social interaction and experiences (Merriam & Caffarella, 1999; Mezirow, 1991). Mezirow's theory of transformative learning extends the work of Malcolm Knowles (e.g., Knowles, 1984), who provided the foundation for most current studies in adult education (Cranton, 2006). Knowles acknowledged that adults may learn in ways different from school-aged children. Although some scholars describe Knowles' work as a theory of adult education, Knowles referred to his work as a conceptual framework that can serve as a basis for theory (Knowles, Holton III, & Swanson, 2005). Building from Habermas' (1971, 1984) distinctions between two learning domains and the transformative nature of learning for the development of transformation theory, Mezirow describes adult learners in a manner consistent with the characterization of adults as self-directed individuals who learn from experience (Knowles, Holton III, & Swanson, 2005; Merriam, 2001; Mezirow, 1985). The transformation theory of adult learning resulted from Mezirow's study of learning of adult women who enrolled in a community college program after a significant period of time away from formal education (Taylor, 1997). As with Mezirow's work, this study includes some teachers who had to learn statistics years after they completed their last formal undergraduate or graduate course.

Forms of Learning

Although Mezirow and others who are conducting research to validate or refine transformation theory may view the theory as a “theory in progress” (Jack Mezirow and Associates, 2000), the main elements of the theory remain constant throughout Mezirow’s discussions, including Mezirow’s characterization of learning. He articulates four forms of learning and describes learning as “the process of using a prior interpretation to construe a new or revised interpretation of the meaning of one’s experience as a guide to future action” (Mezirow, 2000, p. 5).

Learning Through Meaning Schemes

Mezirow’s (1991) first form of learning is “learning through meaning schemes” (p. 93). *Meaning schemes* consist of specific expectations, knowledge, beliefs, attitudes, and feelings (Mezirow, 1991) that are the habitual, implicit rules that we use to interpret our everyday experiences (Cranton, 2006). Because meaning schemes are based on common expectations, people are often unaware of their meaning schemes as they interpret their experiences (Mezirow, 2000). Learning through meaning schemes involves differentiating among or elaborating upon preexisting meaning schemes.

To examine an example of learning through a meaning scheme, consider an individual’s meaning scheme for the statistical concept of standard deviation. In general, researchers have found that many students and adults understand statistical concepts such as the mean purely procedurally (e.g., Mokros & Russell, 1995; Pollatsek, Lima, & Well, 1981). If an individual understands the standard deviation as a computation, then without calculating values, that individual would struggle to describe how adding an outlier to a set of univariate data might affect the standard deviation. What that individual needs is an elaboration of his or her existing meaning scheme. By using dynamic software to explore how changing the value of an outlier affects the values of the mean and standard deviation, that individual may develop a dynamic conception that coordinates changes to the relative density of values about the mean with their deviation from the mean, a conception suggested by some researchers as necessary for understanding standard deviation (delMas & Liu, 2005, p. 56). With the development of this dynamic conception, the individual's meaning scheme expands to include both procedural as well as conceptual aspects of standard deviation. Although this example illustrates an individual's possible meaning scheme for standard deviation, there are other measures of variation in statistics, each of which would also be associated with different meaning schemes.

Meaning schemes are contained within meaning perspectives, which consist of the web of interwoven assumptions and expectations through which the world is viewed (Cranton, 2006). A *meaning perspective* consists of broad predispositions formed from culture, personality, and prior experiences and is used to interpret current experiences (Mezirow, 2000). Perspectives are expressed through a *point of view*. For example, a mathematics teacher may have a personal theory of learning that assumes meaning is transmitted to learners by the mathematical authority in the classroom—a learning theory that is consistent with the transmission model of teaching. This personal theory, of which the individual may not be aware on a conscious level, most likely formed subconsciously from years spent in the "apprenticeship of observation" (Lortie, 1975) as a student. This individual is likely to exhibit his or her point of view about learning through his or her teaching, which most likely is teacher centered and lecture driven.

Learning New Meaning Schemes

Although Mezirow's first form of learning involves making changes to existing meaning schemes, a second form of learning involves the learning of new meaning schemes to complement and expand upon an existing meaning perspective or to lead to a new meaning perspective. Consider an individual's meaning perspective for variation that may consist almost entirely of knowledge about summary statistics values to describe variation in a data set. When learning about sampling and experimental design, the individual might form a meaning scheme for different types of variability, such as sampling variability or measurement variability. These new meaning schemes then become associated with the individual's existing meaning perspective for variation, with learning resulting in new knowledge about variation. Although an individual learns from both the elaboration of and the creation of meaning schemes, these forms of learning result in changes to *what* the individual knows and not *why* the individual knows (Kegan, 2000); neither of these two forms of learning results in transformational learning.

Learning by Transforming Meaning Schemes

Mezirow's third form of learning, *learning by transforming a meaning scheme*, occurs from reflecting on assumptions and results related to a particular meaning scheme when existing values or beliefs appear to be inadequate for current circumstances (Mezirow, 1991). Think about an individual's meaning perspective for statistics and his or her meaning scheme for variation—specifically that individual's beliefs about statistics and variation. That person may have general assumptions and beliefs about statistics consistent with the somewhat traditional view of statistics as the "churning out [of] dry statistical techniques" (Karpadia, 1980, p. 415). Given the complexity and multifaceted nature of variation, however, that individual may encounter circumstances that precipitate reflection on those existing beliefs about variation. That individual may reject the belief that variation consists merely of dry techniques and transform his or her meaning scheme for variation into a more encompassing view of variation. At the same time, that individual's beliefs about statistics in general, namely, his or her meaning perspective for statistics, may not change. Mezirow (1994) acknowledges that transformations to a small percentage of an individual's meaning schemes occur regularly during the course of everyday events; a less common experience is the learning that results from the transformation of a meaning perspective, or a perspective transformation.

Learning by Perspective Transformation

Transformational learning results from transforming a meaning scheme or from transforming a meaning perspective, a fourth form of learning often referred to as a *perspective transformation* (e.g., Cranton, 2006). Often an individual's transformation of a particular meaning scheme within a meaning perspective precipitates the transformation of other meaning schemes within the same meaning perspective. A sequence of transformed meaning schemes within a particular meaning perspective can provoke transformation of the meaning perspective that encompasses the meaning schemes (Mezirow, 1991, 2000). Transformation of a meaning perspective is the fourth and most powerful form of learning identified by Mezirow (2000). A perspective transformation occurs when an individual reflects on the specific presuppositions upon which a current meaning perspective is based and for which these assumptions and beliefs are now seen by the individual as incomplete or invalid (Mezirow, 1991). Mezirow (1991) refers to these incomplete or invalid assumptions as *distorted assumptions*, although the negative connotation associated with the term, "distorted," has caused others to propose labeling these assumptions as “unquestioned” or “unexamined” (Cranton, 2006). No matter how the assumptions are labeled, perspective transformation is the process by which meaning perspectives, consisting of “uncritically assimilated assumptions, beliefs, values, and perspectives” (Cranton, 2006, p. 2), become transformed to give new meaning to an old experience. The result is that the meaning perspective becomes “more open, permeable, and better validated” (Cranton, 2006, p. 2). These transformed meaning perspectives “are more likely to generate beliefs and opinions that will prove more true or justified to guide action” (Mezirow, 2003, p. 59), making the transformed meaning perspective better than the original perspective. As mentioned earlier, an individual’s perspective transformation can be triggered internally by the accumulation of transformed meaning schemes within the same meaning perspective, or that individual’s perspective transformation can be triggered externally by an event that Mezirow refers to as a *disorienting dilemma* (Mezirow, 1990). A disorienting dilemma seems to equate with what is referred to as “perturbation” or “cognitive conflict” in other research literature (e.g., Leikin & Zazkis, 2007; Polettini, 2000).

To illustrate transformations of a meaning perspective, consider an individual’s meaning perspective for statistics. That individual may believe that statistics is merely a subject of data manipulation, display, and calculation. Events may occur that prompt the individual to question assumptions about various statistical concepts and subsequently to transform the meaning schemes for those concepts, as in the example related to variation. This accrual of transformed meaning schemes may trigger an eventual perspective transformation for statistics in which statistics is viewed as a problem-solving process that allows decisions to be made from data. Alternatively, that same individual may attend a professional development workshop in which the individual experiences statistics actively as the study of the collection, organization, and analysis of data within a particular context. Attending this workshop may trigger a disorienting dilemma, such as being confronted with analysis that requires considering artistic aspects of statistics in addition to scientific aspects to draw conclusions, which may eventually lead to a perspective transformation.

Much of the literature on transformation theory focuses on the types of events that lead to a perspective transformation or on designing adult education to prompt an individual's inspection of distorted assumptions. As the examples used throughout the preceding sections might suggest, indications of transformed meaning schemes with respect to variation or transformed meaning perspectives with respect to statistics are important to identify because they are associated with significant learning. Of particular interest are the events that trigger the transformations of meaning schemes and meaning perspectives. The importance of perspective transformations warrants a more careful examination of perspective transformations, and in particular, the different types of perspectives that may be transformed.

Types of Meaning Perspectives

In his writings about transformation theory, Mezirow (1991) concentrated on perspective transformations with regard to three types of meaning perspectives: epistemic, sociolinguistic, and psychological. Although these meaning perspectives will be discussed separately in the next few sections, the perspectives are not clearly demarcated. An individual's interrelated perspectives comprise the individual's worldview.

Epistemic Perspectives

Epistemic meaning perspectives pertain to knowledge, including what an individual knows, how the individual gains or gained that knowledge, and the way the individual uses or acts upon that knowledge (Cranton, 1996). One way an individual's epistemic assumptions and beliefs can be distorted is if the individual assumes that all knowing can be verified empirically and that there exists a correct solution for every problem (Kitchener & King, 1990; Mezirow, 1991). Within education, a teacher may have one particular view for the defining features of effective teaching. Unless that teacher's assumptions about effective teaching are transformed, that teacher would not be open to alternative pedagogical techniques that are not characteristic of those in his or her meaning perspective (Cranton, 1996). Teachers can also have distorted epistemic assumptions if they have a narrow and limited view of teaching and learning. One example of a teacher with a limited perspective would be someone who has a particular preferred learning style, believes that this learning style should be the learning style of everyone, and teaches towards only this learning style (Cranton, 1996). A third example of a distorted epistemic assumption can result from an individual who views socially derived phenomena as beyond one person's control (Mezirow, 1991). In response to this assumption, for instance, a teacher might adopt district administrators' beliefs about teaching and learning unquestioningly (Cranton, 1996).

General examples of Mezirow's (1991) four forms of learning (learning through existing meaning schemes, learning new meaning schemes, learning by transforming meaning schemes, and learning by transforming a meaning perspective) that can change an individual's epistemic meaning perspectives are displayed in Table 3-1. Table 3-1 also includes descriptions of the four forms of learning for sociolinguistic and psychological perspectives and labels examples related to each of the meaning perspectives for the two domains of learning identified in transformation theory, the instrumental and communicational domains (Mezirow, 1991). These other types of perspectives and domains of learning are discussed in subsequent sections.

Table **3-1**: Learning Related to Meaning Perspectives.

Type of Learning	Meaning Perspective	Instrumental and Communicational Domains of Learning
Learning Through Meaning Schemes		Learning through existing meaning schemes—differentiating among or elaborating upon current expectations, feelings, attitudes, or judgments pertaining to …
	Epistemic	empirically testable problems within a particular epistemic meaning perspective (instrumental) or based upon what others say about an issue related to a particular epistemic meaning perspective (communicational)
	Sociolinguistic	what is factually related to social norms and empirically testable within a particular sociolinguistic meaning perspective (instrumental) or subsequent to hearing what others say about social norms within a particular sociolinguistic meaning perspective (communicational)
	Psychological	empirically testable knowledge of oneself and how one came to form that image of self within a particular psychological meaning perspective (instrumental) or subsequent to hearing what others say about oneself within a particular psychological meaning perspective (communicational)
Learning New Meaning Schemes		Learning of new meaning schemes …
	Epistemic	related to factual knowledge (instrumental) or related to the perspective of others (communicational)
	Sociolinguistic	pertaining to factual information about social norms (instrumental) or related to the perspective of others (communicational)
	Psychological	pertaining to knowledge of oneself or how that knowledge complements or expands upon an existing psychological meaning perspective or that leads to a new psychological meaning perspective (instrumental) or the learning of new meaning schemes of oneself related to the perspective of others (communicational)
		that complements or expands upon an existing (epistemic, sociolinguistic, psychological) meaning perspective or that leads to a new (epistemic, sociolinguistic, psychological) meaning perspective

Learning by Transforming Meaning Schemes		Learning by transforming meaning schemes, such as those that result from transforming and testing what is viewed as factual …
	Epistemic	along with examining the validation for this view (instrumental) or those that result from transforming what one views as factual by examining the views of others as well as integrating the views of others (communicational)
	Sociolinguistic	and related to social norms along with validation for this view (instrumental) or those that result from transforming what one views as a social or cultural norm by examining the views of others as well as transforming one's views based upon one's analysis of the views of others (communicational)
	Psychological	and related to one's image of oneself along with validation for this view (instrumental) or those that result from transforming a particular view of oneself by examining the views of others as well as transforming one's views based upon one's analysis of the views of others (communicational)
Transformational Learning		Learning by transforming an (epistemic, sociolinguistic, psychological) perspective, such as transforming perspectives …
	Epistemic	about why particular knowledge is (or is not) important (instrumental) or transforming perspectives about the validity and utility of a perspective based upon the conclusions drawn by using another's point of view (communicational)
	Sociolinguistic	about why knowledge related to social norms is (or is not) important (instrumental) or transforming a sociolinguistic meaning perspective about the validity and utility of a social norm based upon consideration of the conclusions drawn by using another's point of view (communicational)
	Psychological	about why knowledge related to one's image or perception of self is (or is not) important (instrumental) or by transforming perspectives about the validity and utility of a person's image based upon consideration of the conclusions drawn using another's point of view (communicational)

Sociolinguistic Perspectives

An individual's sociolinguistic meaning perspectives form initially from the social and historical context in which the person lives and participates. His or her meaning perspectives include social norms and cultural expectations of which the individual may be consciously unaware (Cranton, 2006). Distortions in an individual's sociolinguistic assumptions result from the way that individual's society, culture, and language may limit understanding (Mezirow, 1991). Related to education, a teacher may not question some of the negative ways educators are portrayed in the media, or the teacher may feel constrained by the educational system in which he or she teaches but never question or closely examine the system (Cranton, 1996). The underlying distorted beliefs could be that the media accurately portrays teachers, and the educational system in which the teacher participates is the best system for educating youth. Sociolinguistic perspectives often include metaphors that do not register within conscious thought processes, but which may affect behavior directly. One such perspective is the teacher who has an image of students as "blank slates." This teacher's view is consistent with a transmission theory of learning, which may translate to a practice that is teacher centered and lecture driven. This metaphor can provide a glimmer of the teacher's assumptions about and perspective of learning (Cranton, 1996). In many cases, metaphors for sociolinguistic perspectives reveal an image of distorted assumptions. General examples of learning that can change an individual's sociolinguistic perspectives are displayed in Table 3-1. In addition to assumptions and beliefs related to knowledge and society, individuals also form meaning perspectives about their inner beings, which relates to the psychological meaning perspectives.

Psychological Perspectives

An individual's psychological meaning perspectives are formed by his or her self-concept, feelings, and personality traits to form that person's self-image (Cranton, 2006). Distortions in psychological assumptions, when recognized, can cause individuals intense pain if there are inconsistencies with the way the individual views himself or herself (Mezirow, 1991). A teacher may have a distorted psychological belief or assumption if that teacher's behavior is influenced by personal negative childhood school experiences. An example of a teacher with this distortion is a teacher who subconsciously avoids discussing a child's potentially problematic behavior with the child, based on a subconscious reaction to another teacher's treatment of the teacher as a student (Cranton, 1996). Additionally, a mathematics teacher may see himself or herself as the mathematical authority in the classroom and thus have a self image of being the most knowledgeable person in the classroom. If continually confronted by a precocious student's questions, however, that teacher's self-image may be questioned and subsequently changed or transformed. General examples of the four forms of learning that can change an individual's psychological perspectives are displayed in Table 3-1.

Perspectives and Learning

The three types of perspectives interrelate to form a person's knowledge base. For that reason, when a meaning scheme or meaning perspective of any type is transformed, it can affect meaning schemes and perspectives of other types. Although this study focuses on teachers' epistemic meaning schemes and perspectives for variation and statistics, dilemmas triggered and resolved with respect to sociolinguistic or psychological perspectives can cause a teacher to question assumptions, beliefs, and perspectives related to their epistemic schemes and perspectives. As a result, an awareness of different types of perspectives can inform learning related to the specific epistemic perspectives of interest. As an example, reconsider a mathematics teacher who sees himself as the mathematical authority whose self-image is questioned and transformed subsequent to critically reflecting on a precocious student's questions. Although that teacher transformed a psychological meaning perspective, the student's questions are likely to have impacted epistemic meaning schemes and perspectives for the teacher based on factual information related to the questions. That teacher may accumulate factual information in answer to the student's questions or may learn by inquiring about and considering the student's perspectives behind what prompted the questions. The teacher's subsequent learning may occur in the instrumental domain of learning, the communicative domain, or both.

Domains of Learning

Table 3-1 displays characteristics typical of learning for the four forms of learning identified by Mezirow and displays learning within two distinct domains: the instrumental domain of learning and the communicational domain of learning. Both instrumental learning and communicational learning can result in changes to an individual's epistemic, sociolinguistic, or psychological perspectives. Learning rarely occurs in a distinctly single domain of learning and can be triggered from dilemmas for any of the three types of meaning perspectives.

When an individual experiences a disorienting dilemma or undergoes a succession of meaning scheme transformations within a meaning perspective, that individual may not possess the knowledge needed to resolve the dilemma or to transform the meaning perspective. Additional learning is needed for transformation to be achieved. Based on Habermas' (1971) classifications for knowledge generation, Mezirow (1991) identified two broad domains of learning in which adults may acquire the necessary knowledge for transformation. The instrumental domain of learning pertains to technical learning, or learning that concerns “the way we control and manipulate our environment, including other people” (Mezirow, 1991, p. 73), and the communicative domain of learning pertains to practical learning, or learning related to understanding others, both through attempting to understand others’ meanings and through communicating one’s own meanings. Most adult learning occurs in both domains, with instrumental learning occurring subsequent to communicative learning, for example. Although learning typically involves aspects characteristic of both domains, these two domains of learning are discussed separately.

Instrumental Domain

Learning in the instrumental domain results in technical knowledge, or knowledge related to empirical, task-based problem solving. It is learning predominantly related to learning “how to do something” that can be validated empirically. Instrumental learning results from examining factual information related to a problem, examining alternative strategies for solving the problem in the most efficient and effective manner, and testing the envisioned course of action empirically to solve the problem (Cranton, 2006; Mezirow, 1990). This learning extends an individual’s current technical knowledge and deepens the knowledge in existing meaning perspectives, as shown in the descriptions of learning depicted in Table 3-1 (Kegan, 2000). Action within this domain is based upon empirical knowledge and “centrally involves determining cause–effect relationships” (Mezirow, 1991, p. 73). Learning occurs when an individual reflects on the contextual or procedural assumptions used to guide problem solving and when the assumptions lead to strategies and tactics that are more efficient in producing the cause–effect relationship (Mezirow, 1990, 1991). Learning can also occur upon examining and critically reflecting upon the importance of instrumental knowledge. Most adult learning from experience falls within the instrumental domain (Marsick, 1990). Consider someone who may learn a new technique for modeling data, such as the use of time-series models rather than least squares regression for modeling data collected over time. This method then becomes a part of that person's meaning perspective for data analysis. Because most learning falls within the instrumental domain, most learning is associated with the instrumental domain. Mezirow (1991) considers the learning that occurs within the communicative domain to be more significant than instrumental learning, since it involves "*understanding the meaning* [italics in original] of what others communicate" (Mezirow, 1990, p. 8).

Communicative Domain

Learning in the communicative domain is learning that occurs through discourse and is validated through consensus (Mezirow, 1985). Learning in the communicative domain may occur when an individual attempts to understand the meaning of what others communicate either verbally or in written form with respect to values, feelings, and beliefs. It also can occur when an individual attempts to make his or her own values, feelings, and beliefs be understood or when an individual attempts to integrate the points of view of others into his or her own perspectives (Mezirow, 1990, 1991). In this domain of learning, the focus is on the justification of beliefs through understanding, describing, or explaining values, feelings, and beliefs and reaching consensus on the validity of the beliefs (Cuddapah, 2005; Mezirow, 1991). Learning in the communicative domain potentially has the greatest effect on an individual's sociolinguistic and psychological meaning perspectives (Marsick, 1990), as depicted in Table 3-1. By interacting and communicating with others, an individual may need to view an experience in terms of a different meaning scheme, necessitating one of the first three forms of learning (Mezirow, 1990). The individual also may reflect critically on the assumptions of others in relation to his or her own assumptions to transform a meaning perspective. Continuing with the previous example of a someone whose meaning perspective for data analysis expanded to include time-series models, consider a difference in beliefs between individuals about the most appropriate technique, using linear regression models or time-series models, to analyze a particular set of data. In such a case, communication between the individuals could ensue, with each person attempting to understand the analysis beliefs of the other. In this manner, a change in meaning schemes may occur for one or both individuals for those particular statistical techniques.

Elements of Transformational Learning

Learning in the instrumental domain results from reflection on the content of problem solving through questioning the problem, context, and premises about the importance of the problem (Cranton, 2006; Mezirow, 1985). Learning in the communicative domain results from interacting with and sharing values and beliefs with others, often in the context of problem solving, and while questioning the validity of existing assumptions and beliefs (Cranton, 2006; Mezirow, 1985). As mentioned previously, learning in both the instrumental and communicative domains is interdependent and interrelated in an individual’s learning and results in that individual’s technical and practical knowledge. An individual’s reflection on the *content* and *process* of problem solving, often while engaging in dialogue with others about the problem, typically results in meaning schemes that are changed, created, or transformed (Mezirow, 1991). An even more powerful form of learning may result from an individual critically reflecting on the *premises* of problem solving, that is, questioning the importance of or the validity and utility of the problem-solving content and process, often while engaging in dialogue with others (Cranton, 2006; Mezirow, 1985). Critical reflection can result in a perspective transformation and begins by critiquing the presuppositions upon which beliefs from prior learning are built (Mezirow, 1990, 1991).

Most current professional development efforts focused on educational reform are undertaken to bring about a change, or transformation, in teachers' beliefs and ideals. Although few individuals have linked transformation theory directly to their descriptions of teacher learning and change (Cuddapah, 2005), the applicability of transformation theory can be seen in researchers' descriptions about reform. For example, Thompson and Zeuli (1999) state that reforms "call for very deep changes—even a transformation—in teachers' ideas about and understandings of subject matter, teaching, and learning" (p. 350). In particular, they note that "transformative learning [would] be required to realize science and mathematics education reformers' visions of curriculum and teaching" (p. 350). For reform to occur, Thompson and Zeuli posit that many mathematics teachers would need to examine and change their beliefs about mathematics, teaching, and learning and seek to enhance their understandings related to each. Transformation theory provides the mechanisms behind transformed learning. To examine the processes associated with perspective transformations in teaching, it makes sense to look at the processes in conjunction with the processes contained in a model for teacher change. In particular, the model of professional growth described by Clarke and Hollingsworth (2002) aligns well with elements of perspective transformation.

In his initial research, Mezirow identified ten phases for perspective transformation, but recent research suggests that not every one of these ten elements is necessary for a perspective transformation to occur (Taylor, 1997, 1998, 2000). This same body of research also suggests that the transformation process is less linear and more recursive than originally envisioned by Mezirow (Merriam, 2001; Taylor, 1997, 1998, 2000). Agreement does appear to exist on the three main phases of transformative learning that result in a perspective transformation: *critical reflection*, *rational discourse*, and *action* (Cuddapah, 2005; Merriam & Caffarella, 1999). Table

3-2 displays those three elements for perspective transformation, along with the corresponding elements originally posited by Mezirow. The table also shows how the elements of Clarke and Hollingsworth's model for professional growth correspond with Mezirow's elements of transformation. Whereas the table gives the impression of a linear process or sequential steps for perspective transformation and teacher change, individuals' transformational learning and change may not occur in this order. Elements related to the three main phases of transformative learning in combination with elements from Clarke and Hollingsworth's Model of Professional Growth are examined in the next section.

Table **3-2**: Comparison of Perspective Transformation With Professional Growth.

Main Elements of a Perspective Transformation	Mezirow's Elements of Perspective Transformation	Corresponding Elements of Clarke and Hollingsworth's Model of Professional Growth
Critical Reflection	Disorienting dilemma or sequence of transformed meaning schemes	External stimuli or culmination of internal reflections
	Self-examination, accompanied by emotions	Reflecting on the personal domain (knowledge, beliefs, and attitudes) and/or the domain of practice
	Critical assessment of assumptions related to epistemic, sociolinguistic, or psychological perspectives	Reflecting critically on the personal domain and reflecting critically on the domain of consequence (salient outcomes)
Rational Discourse	Recognition that others have experienced similar discontent with their perspectives	Acting on and reflecting on the external domain (sources of information)
	Exploring new roles, relationships and actions through engaging in rational discourse with others – learning in the communicative domain	Acting on the domain of practice (professional experimentation) and acting and reflecting on the external domain (sources of information)
Action	Planning a course of action	Reflecting on the personal domain, the domain of practice and/or the domain of consequence
	Constructing the knowledge and skills needed to enact the plan – learning in the instrumental domain and possibly in the communicative domain	Acting on and reflecting on the external domain and reflecting on the personal domain
	Experimenting with new roles	Acting on the domain of practice
	Building a sense of competence and self-confidence for new roles and relationships	Reflecting on the domain of practice and the domain of consequence
	Reintegration into life based on the transformed perspective	Acting on the domain of practice

Critical Reflection

Perspective transformations typically begin with one or more events that precipitate a disorienting dilemma or an examination of presuppositions for which current problem-solving processes do not provide resolution to the problem at hand (Merriam & Caffarella, 1999). Although Clarke and Hollingsworth do not specify the need for a disorienting dilemma to stimulate professional growth, teachers typically have some motivation for pursuing professional development that results in change. That motivation may come internally from an individual's experience of a disorienting dilemma or transformed meaning schemes, or the motivation may come externally from reform efforts within the individual's school or professional community. Internal motivation proceeds in a manner similar to the process described by transformation theory; external motivation, however, requires some action to precipitate the teacher's need to reflect on practice and on assumptions and beliefs about teaching and learning.

A second characteristic of critical reflection is self examination, which is often accompanied by strong emotions (Mezirow, 1991). Mezirow (2000) acknowledges that becoming aware of previously implicit presuppositions and recognizing a need for change can present a threatening, emotional experience for many adults. As a result, some individuals resort to accepting the status quo or succumbing to the perspective of an authority, whereas others proceed to critically assess their epistemic, sociolinguistic, or psychological assumptions through critical reflection (King, 2002; Mezirow, 1994, 2000). Clarke and Hollingsworth incorporate the need for teachers to reflect on practice and for teachers to reflect on assumptions and beliefs about teaching and learning for professional growth throughout their model.

Rational Discourse

Upon engaging in critical reflection, an individual may take consolation in the fact that others have experienced similar discomfort when examining their assumptions and beliefs (Mezirow, 1991). In advance of preparing a plan of action to resolve a dilemma, an individual may explore options for new roles, relationships, and actions by either engaging in rational discourse with others or internally engaging in rational discourse. Through discourse, the individual can examine the experiences of others and gain insights into both his or her own assumptions and the assumptions of others. The resultant learning falls mainly within the communicative domain (Mezirow, 1991, 2000). It is in this phase that the establishment of trusting relationships with others may play a role, since open and frank discussions can occur (Mezirow, 2000). Obtaining emotional support can take the form of support from colleagues who simultaneously engage in a change process or through rational discourse with colleagues pertaining to assumptions and beliefs about teaching and learning. Although Clarke and Hollingsworth do not explicitly address the role of others in rational discourse, their model of professional growth takes into account information gained from external sources, which does not preclude information gained through rational discourse with colleagues. Additionally, the support of others could be one part of the change environment in which the teacher participates.

Action

To bring about resolution to a dilemma, an individual may plan a course of action and then go about constructing the knowledge and skills for enacting the plan. The resultant learning falls mainly within the instrumental domain (Mezirow, 1991). Acting on plans for change is a major element of Clarke and Hollingsworth's model for professional growth. After a teacher develops a sufficient knowledge base, the teacher may experiment by taking on new roles to build competence and self-confidence with the new roles and relationships (Mezirow, 2000). The final result is that the individual lives life based on a newly transformed perspective—a perspective that is more inclusive, open, and discriminating (Mezirow, 2000). Taking action to make changes in practice may precipitate the need for further reflection and emotional support, which may suggest the need for increased knowledge in preparation for further action.

Conditions for Perspective Transformation

Much of the research on transformation theory has been focused on validating and expanding the theory; less research has been conducted to investigate under what conditions a perspective transformation might be likely to occur (Taylor, 1997, 1998, 2000). Several adult educators have speculated about some of the conditions they believe are conducive to enabling perspective transformations. These conditions include some or all of the following for teachers: dissatisfaction with current practice, occurrence of a disorienting dilemma, critical examination of beliefs, support and freedom to pursue alternatives, support and opportunity to engage in rational discourse, readiness for change, and openness to alternative perspectives (Cranton, 2006; Cuddapah, 2005; Merriam, 2004a). A teacher's experience with a disorienting dilemma related to teaching or learning or a teacher's dissatisfaction with teaching practices are conditions that may precipitate the teacher's critical examination of beliefs (Cranton, 1994). A teacher's readiness for change might also increase the likelihood of engaging in the difficult task of critically examining beliefs. Critically questioning previously unexamined assumptions and beliefs about teaching and learning, particularly through discourse with others, may motivate a teacher to listen to and consider alternative perspectives. Administrative support and emotional support from peers may create "safe" conditions under which the teacher feels free to experiment with new roles. Others have speculated that long-term work with a mentor, faculty developer, group of faculty members, or network of teachers from multiple districts can sustain transformative learning in teachers by providing the emotional support needed for perspective transformation.

Professional development workshops in many districts are of short duration and focus strictly on technical knowledge; they result in little meaningful teacher change (Desimone, Porter, Garet, Yoon, & Birman, 2002). In contrast, researchers who examined programs designed to effect change found that the most successful programs were those that provided long-term support and assistance to teachers (Shields, Marsh, & Adelman, 1998). To facilitate the change process, teachers need support to battle through the emotional strain that accompanies the examination of underlying beliefs. Teachers also need to engage in rational dialogue with colleagues and take action to learn the new content and strategies required for transformed practices. Current prevalent professional development practices for teachers of grades K-12 often appear to be unfocused and disconnected (Cohen & Hill, 2001). These programs do not seem to provide a focus on creating conditions that are indicative of transformation-provoking and transformation-sustaining activities, and most importantly, activity to support teachers' examination of beliefs and engagement in rational discourse.

In contrast to traditional professional development endeavors, Whitelaw, Sears, and Campbell (2004) provide an example of a program that was supportive for provoking teachers' transformations. They studied instructors' learning within an initiative designed to promote faculty members' technology use. The program invited teachers to examine their teaching practices under the umbrella of incorporating technology into practice. The program itself focused on technical knowledge. Teachers worked in pairs to explore the technology and to dialogue about how the technology could be used. For some pairs, this dialogue resulted in practical knowledge. The researchers found that those pairs who self-reported significant learning recognized the value of examining their beliefs about teaching, engaging in dialogue with their partners, and exploring how technology use could support their beliefs. Additionally, several participants experienced a misalignment between their expectations for the initiative and their experiences while participating in the initiative. The authors contend that this misalignment provided an opportunity for participants to engage in critical reflection on their practice through the triggering of a disorienting dilemma.

There are also other professional development opportunities that are available to teachers and that may precipitate a change in practice. Within mathematics and statistics education, there exist several independent educational organizations, such as the National Council of Teachers of Mathematics. These organizations provide both formal and informal learning opportunities for self-directed learners, including workshops, conferences and publications (Merriam & Caffarella, 1999). Such workshops and conferences provide teachers with opportunities to exchange information with others and to examine educational publications and materials— activities that may prompt critical reflection and discourse (Cranton, 1994). Similar venues for prompting critical reflection and discourse are available in formal educational settings, such as graduate classes or degree programs offered by colleges and universities (Merriam & Caffarella, 1999).

Lastly, proponents of adult education suggest that emotional maturity is required for a perspective transformation to occur (Mezirow, 2000). Although there is some question about whether precocious teens may experience perspective transformations, transformation theory is considered by many to be strictly an adult theory of learning (Taylor, 2000). Mezirow (1985) defines an adult as "one who fulfills adult social roles and who sees himself or herself as a self-directed person" (p. 17). Associated with emotional maturity are the abilities to be critically self-reflective as well as to engage in rational dialogue (Cuddapah, 2005). Others (e.g., Knowles, Holton, & Swanson, 2005) suggest that only adults have the types of experiences and resources necessary to be able to critically examine previously uncritically assimilated assumptions and beliefs, and only adults have the desire to resolve any contradictions between beliefs and experiences (Mezirow, 1991). The main factors that seem to distinguish transformation theory from learning theories used to investigate students' learning are recognition of the need for emotional maturity and acknowledgement of the role of emotion in transformational learning coupled with the realm of experiences needed for critical reflection. In addition, others (e.g., Merriam, 2004b) suggest that a "mature level of cognitive functioning" (p. 60) is required for transformational learning—cognitive development that may extend beyond Piaget's formal stage of cognitive development. As the preceding descriptions suggest, viewing teacher learning through the lens of transformation theory is consistent with current explanations for teacher change.

The SOLO Model

Transformation theory provides explanatory power for what prompts adults to construct the knowledge and beliefs necessary to function at advanced cognitive levels as well as for how adults construct that knowledge. Whereas transformation theory does provide explanatory power for adult learning, it offers only a global perspective of learning and gives little insight into the intricacies of an individual's meaning perspectives and the complex, interrelated web of knowledge, assumptions, and beliefs associated with the meaning schemes that combine to form these meaning perspectives. In relation to this study, transformation theory may provide insight into how professional development can provoke individuals' construction of robust understandings of variation by examining closely teachers' activities and actions that may contribute to their development of that understanding. Ensuring that teachers have robust understandings of variation to answer the second research question of this study, however, can be accomplished best by examining teachers' understanding through a lens different from transformation theory. The Structure of the Observed Learning Outcome (SOLO) Model (Biggs & Collis, 1982, 1991) provides such a lens.

Establishing teachers' understandings of variation can be accomplished by examining their responses to interview tasks using the SOLO Model (Biggs & Collis, 1982, 1991). The SOLO Model is an empirically derived, neo-Piagetian model of cognitive development that assumes individuals actively construct knowledge through interactions with the world around them (Pegg & Tall, 2001, 2004), thus sharing similar constructivist assumptions with transformation theory. Unlike transformation theory, however, SOLO allows for both a global analysis of long-term growth and a local analysis of conceptual growth (Pegg & Tall, 2001), thus enabling description of both cognitive development and the complexity and cogency of the knowledge that results from learning (Cantwell & Scevak, 2004). It is in the latter sense that SOLO is considered for this study.

Collis, Romberg, and Jurdak (1986) describe the dual phenomena that the SOLO Model addresses and mention two descriptors initially used by Biggs and Collis for the phenomena: Hypothetical Cognitive Structure and the Structure of the Learned Outcomes or Responses. The Hypothetical Cognitive Structure is the tool for describing cognitive development, and the Structure of the Learned Outcomes or Responses is the tool for describing an individual's structure of response to a task, with the response not necessarily indicative of development. As Biggs and Collis (1982) observe, the SOLO levels "describe a particular performance at a particular time" (p. 23). An individual may respond to a task using a lower mode of reasoning even though the individual is capable of reasoning at a higher mode. Much of the research in mathematics education and in statistics education alludes to the developmental nature of the SOLO Model while using the model to describe the structure of students' responses to mathematical or statistical tasks. For this study, SOLO is used to examine the structure of teachers' responses to statistical tasks and provides a useful lens through which to design tasks to elicit understanding and to examine an individual's knowledge subsequent to learning. Because it can be used to describe developmental levels, SOLO offers a way to describe varying depths of understanding. In this study, the varying depth is not that of one individual across time, as it would be in developmental research, but rather the varying depths of understanding at one time across individuals.

The SOLO Model (Biggs & Collis, 1982, 1991) consists of five modes of functioning that correspond closely to Piaget's developmental stages: the sensorimotor, ikonic, concrete symbolic, formal, and postformal modes. Each mode of learning reveals the "level of abstraction that a learner uses when handling the elements of a task" (Biggs & Collis, 1982, p. 152), with each successive mode increasing in the degree of abstraction needed to reason within that mode. Thought processes in each successive mode are qualitatively more complex than thought processes in earlier modes.

This study focuses strictly on the formal mode of reasoning. Reasoning within the formal mode can be used to generate speculations that both incorporate and transcend particular situations; characteristic of this mode is reasoning that does not require reference to a particular concrete setting. Although some adults never develop sufficient understanding of concepts within particular areas of study to reason in the formal mode for those areas, thinking in the formal mode is seen as representative for the type of thinking exhibited by undergraduates and professionals (Biggs & Collis, 1982; Groth & Bergner, 2006). Secondary statistics teachers in particular should be able to think and reason about variation in the formal mode. Not only does the SOLO Model provide a means to examine individual responses on a global level, SOLO also affords the use of a finer grain of analysis by examining learning cycles within each mode (Pegg & Tall, 2005).

Levels of Response

There exists essentially a cycle of three levels of response, or levels of thinking, to describe cognitive growth in the formal mode: unistructural, multistructural, and relational (Biggs & Collis, 1991). The underlying cognitive structure of a response may reveal different foci in an individual's response. At the unistructural level, an individual focuses on one relevant aspect that lies in the formal mode, whereas at the multistructural level the individual focuses on more than one relevant aspect without integrating the aspects. At the relational level, the individual integrates all relevant aspects to reveal coherent structure and meaning. The three levels within a mode form a cycle of cognitive growth (Pegg, 2003), with the cycle being hierarchical in nature and generally accepted as developmental in nature (Pegg, 2003).

Recent empirical studies identify more than one cycle of levels of response that a learner exhibits within a mode before possibly reasoning beyond that mode (Callingham, 1997; Pegg, 2003; Watson, Collis, Callingham, & Moritz, 1995). Pegg and Tall (2001) provide one of the few descriptions of the SOLO Model that acknowledges the possibility of multiple cycles of levels within a mode. They suggest that multiple cycles within a mode occur after

> the individual meets new stimuli and begins to react first to one aspect, then another, to give multiple responses, which begin to be related together, then the whole structure is conceptualized as a new single structure. This structure can retain characteristics of the initial cycle. (Pegg & Tall, 2001, p. 2)

This single new structure characterizes the unistructural level in a second cycle of levels within a mode. Given the complex nature of variation, it seems possible that individuals may be able to reason relationally about aspects of variation in the formal mode while not integrated reasoning among aspects, suggesting that relational reasoning about variation in the formal mode might encompass more than one level of reasoning.

The SOLO Model as a Tool for Eliciting Understanding of Variation

To date, researchers have used the SOLO Model to explore issues of individuals' understanding and reasoning about variability in the concrete-symbolic and ikonic modes (Reading, 2004; Watson, Kelly, Callingham, & Shaughnessy, 2003). A few statistics education researchers have used the SOLO Model to examine and describe reasoning in the formal mode (e.g., Groth & Bergner, 2006), and mathematics education researchers have described students' reasoning in the formal mode for mathematical topics in several areas of mathematics (e.g., Pegg & Tall, 2005; Serow, 2006). The body of statistics education research work consists of analyzing students' responses to form descriptions of understanding and reasoning in relation to the levels of reasoning within a mode or to distinguish reasoning between modes. In these studies, either the participants were unable to reason in the formal mode, or the tasks did not elicit reasoning indicative of the formal mode. Mathematics education researchers (e.g., Collis, Romberg, & Jurdak, 1986) have used an alternative process with the SOLO Model to reverse the usual order from response to level classification to address the latter situation. The alternative process is used to write tasks that elicit responses reflective of SOLO levels within the mode intended to be studied or that distinguish between modes (Collis, Romberg, & Jurdak, 1986; Rider, 2004). This alternative process was used in this study to design tasks to elicit reasoning in the formal mode along with empirical verification of responses to confirm the tasks elicited formal reasoning. The process also was used to analyze responses to describe reasoning and understanding about variation in the formal mode.

Figure 3-1 provides a simplified graphical representation for the first cycle of levels of understanding variation using the SOLO Model. This figure shows that an individual can have an integrated understanding for each perspective of variation in the formal mode. The first cycle of unistructural, multistructural, and relational levels within the formal mode is a cycle of levels within each perspective of variation, with the levels subscripted with a "1" depicting the first cycle for each perspective and the arrows representing the hypothetical development of an understanding of variation and increased sophistication in reasoning from each perspective. For example, an individual who provides evidence of relational reasoning (R_1) within the design perspective displays integrated reasoning with statistical problem-solving elements related to design. In contrast, unistructural reasoning (U_1) from the design perspective is evidenced by reasoning about a single problem solving element related to design, and multistructural reasoning (M_1) is evidenced by reasoning about multiple elements without integration of the elements.

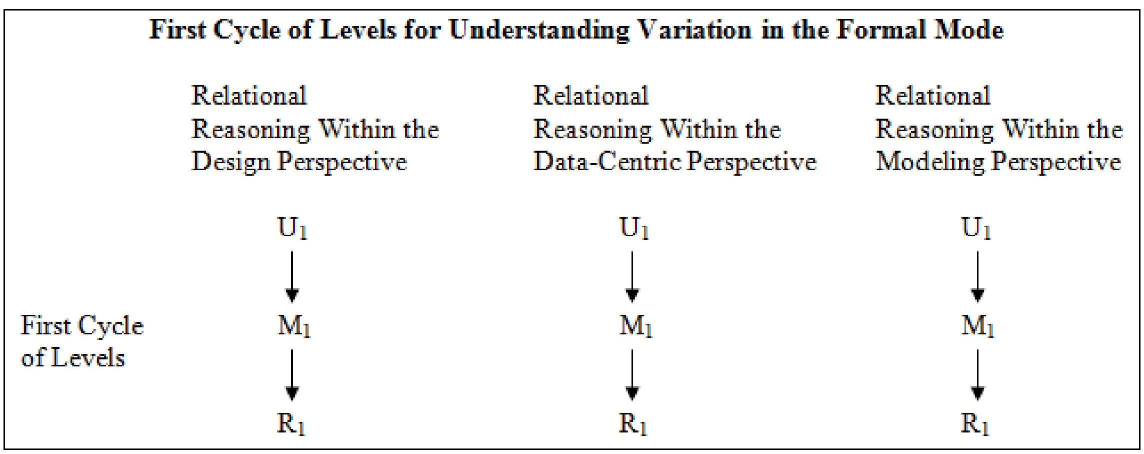

Figure **3-1**: SOLO and the Cycle of Levels for Each Perspective.

The second cycle of levels in the formal mode (the cycle in the bottom half of Figure 3-2) requires integrating the three perspectives to reveal a holistic understanding of variation. The levels subscripted with a "2" depict this second cycle and the arrows represent hypothetical development and increased sophistication towards relational reasoning across perspectives. An individual who reasons relationally from each perspective and integrates reasoning from the three perspectives is reasoning at a relational level (R_2) in the second cycle of levels in the formal mode. Relational reasoning across perspectives is indicative of robust understanding of variation. Relational reasoning within only one perspective is indicative of unistructural reasoning (U_2) in the second cycle, and relational reasoning within more than one perspective without integration between perspectives is indicative of multistructural reasoning (M_2) in the second cycle. Although individuals who reason at the unistructural and multistructural levels in this second cycle exhibit relational reasoning from one or more perspectives, they do not exhibit reasoning indicative of an overall robust understanding of variation. The interview tasks and the lines of questioning related to each task in this study were designed to elicit reasoning about variation in the formal mode and to allow determination of levels of reasoning in both cycles of levels, thus providing a view of robust understanding of variation as relational reasoning across the three perspectives.

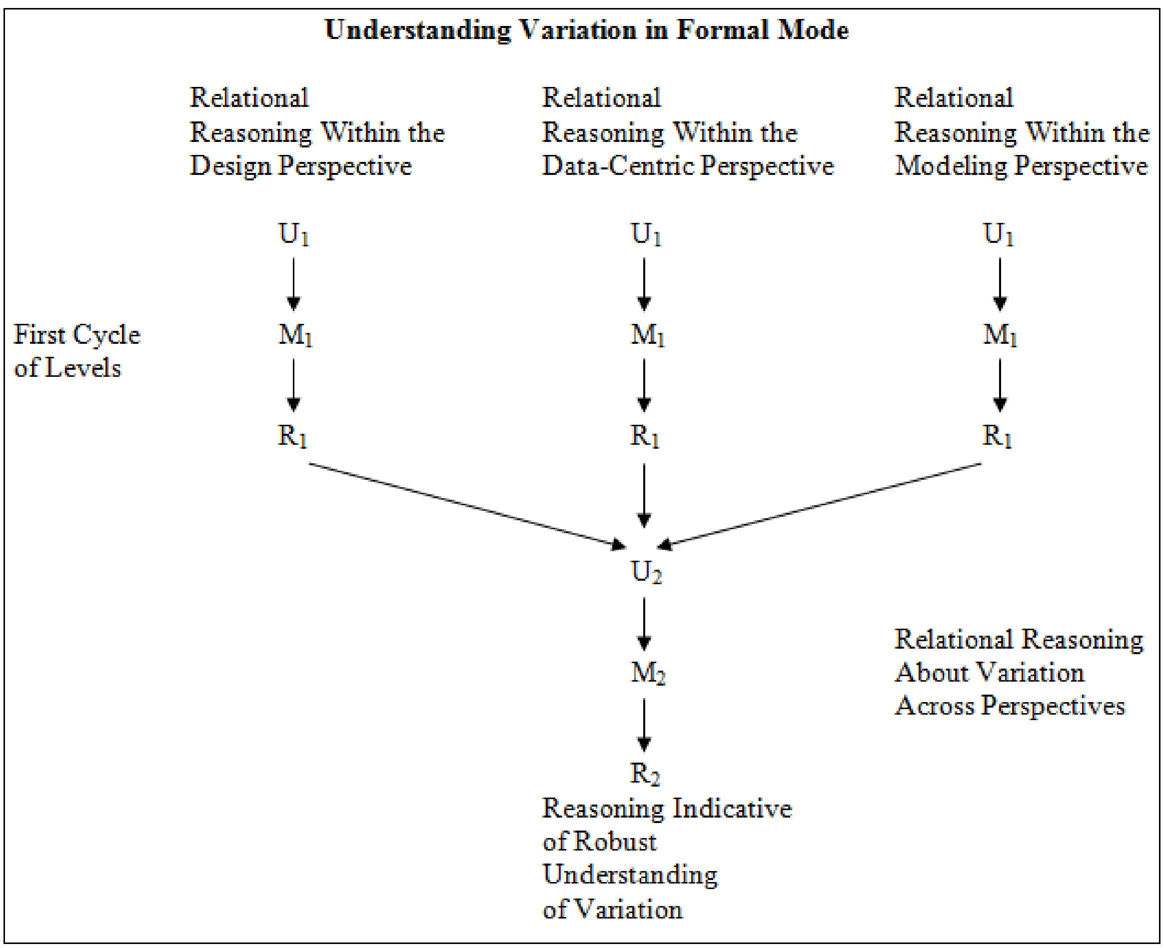

Figure **3-2**: The SOLO Model and Robust Understanding of Variation.

Chapter 4

Research Methods

In addition to investigating teachers' conceptions of variation, a major goal of this study is to understand the nature of experiences that secondary mathematics teachers believe contributed to their development of robust understandings of statistical variation. To examine the fundamental nature of these experiences and to obtain a holistic description of the nature of the experiences, I use phenomenological methods (Moustakas, 1994), for which the phenomenon under study is secondary mathematics teachers' development of robust understandings of statistical variation. A major requirement for phenomenological study is that the topic and research questions be of social significance (Moustakas, 1994); the perceived need for students to become statistically literate and to have statistically literate teachers suggests the significance of this study. This study can contribute to the future design of programs that advance the development of a statistically literate society. In this chapter, I describe in detail the processes that I used to conduct this study. Specifically, I describe the processes used to select participants, to collect data, and to analyze data in order to examine teachers' conceptions of variation and to establish that teachers experienced the phenomenon under study as well as to identify characteristics of experiences identified by teachers as important for their learning during the course of experiencing the phenomenon.

Participant Selection

In order to study the phenomenon of coming to understand variation, participants must have experienced the phenomenon (Moustakas, 1994). To ensure the highest probability of finding teachers who have robust understandings of variation, I established explicit criteria for individuals' addition to the participant pool (Merriam, 1998). The resulting purposeful sample consisted of 16 high school statistics teachers from across the country; a different group of three teachers participated in a pre-pilot study and another three teachers participated in a pilot study. To establish the selection pool, I considered teachers who have been active in The College Board's Advanced Placement® (AP) Statistics Program.

Background Information

The AP program allows secondary students to receive college credit for courses taken during their high school years (Collegeboard.com Inc., 2007a). To potentially receive credit in statistics, students take the AP Statistics exam, which consists of multiple choice and free response questions. Each year, secondary AP Statistics teachers and college statistics instructors meet in a central location (the *AP Reading* site) to score students' free response solutions (R. Peck, Personal communication, May 21, 2007).

The College Board requires that secondary mathematics teachers have a minimum of three years of experience in teaching AP Statistics or the equivalent before they can attend the AP Reading (Collegeboard.com Inc., 2007c), although exceptions have been made when the demand for readers has exceeded the available pool of readers. As part of the application process, teachers provide information about their educational experiences and submit a curriculum vita and course syllabus (Collegeboard.com Inc., 2007c). Subsequent to examining the application materials, the *Chief Reader*—a college statistics professor who has served various roles at the AP Reading and who is hired to be a "content arbitrator"—extends invitations to the Reading (R. Peck, former Chief Reader, personal communication, May 21, 2007). The teachers who attend the Reading are the *readers*.

At the conclusion of each AP Reading, readers are evaluated and must exhibit proficiency in evaluating student responses for statistical completeness and correctness to be invited back to the AP Reading in subsequent years (R. Peck, Personal Communication, May 21, 2007). To evaluate student responses for students from multiple states and countries—responses that include both conventional and unconventional methods—readers need to have and exhibit flexibility in their understanding of statistics. Attending the AP Reading also presents a unique learning experience for teachers. As noted by one first-time AP Statistics reader, "[t]he greatest gain that readers will take away from an AP Reading … is an increase in their own knowledge and skills within their chosen fields" (Rees, 2007).

With the selection criteria for AP readers and the educational benefit from attending an AP Reading, secondary teachers who attend the AP Reading pass a screening process that suggests some level of competence in the area of statistics, and they participate in potentially powerful professional development. Table leaders pass through an additional screening process. To become a table leader, a teacher typically must serve as a reader for six years, currently teach AP Statistics or an equivalent course, and exhibit characteristics of leadership (R. Peck, Personal Communication, May 21, 2007). They also must be recommended by an existing table leader or serve the AP program in a particular type of leadership role (Collegeboard.com Inc., 2007d). In general, table leaders have taught statistics for a minimum of six years and exhibit their statistical knowledge through the selection process for readers and table leaders.

Selection Criteria

With AP readers' and table leaders' wealth of background experiences, these teachers are more likely to exhibit robust understandings of variation than a random sample of AP teachers (e.g., R. Peck, Personal Communication, May 21, 2007) and thus are more likely to have experienced the phenomenon of interest—a key consideration for participant selection in phenomenological studies (Moustakas, 1994). For that reason, AP readers and teacher leaders formed the population of interest for this study. In particular, teachers who attended the 2006 and 2007 AP Statistics Readings as readers or table leaders comprised the initial group of teachers considered for this study. Of the 16 teachers ultimately selected for participation, 15 attended a minimum of two AP Readings,[3] including one teacher who attended nine. On average, teachers attended 4.6 AP Readings, with a median of 4 AP Readings. Six of the teachers currently serve or served in the past as table leaders at the AP Reading and averaged 5.167 years of service as a table leader, with a median of 5 years of service.

To achieve a participant pool that represented some diversity in experiences, I selected teachers who differed in the number of years they taught statistics. My belief was that individuals who have more recently begun to teach statistics may be able to recount the activities and actions that contributed to their current understanding of variation better than veteran teachers, whereas veteran teachers may have a greater variety of activities and actions that contributed to their understandings. The teachers in this study taught statistics for as few as 3 years and for as many as 30 years. The mean number of years that teachers taught statistics at the time of data collection was 10.75 years, with a median of 9.5 years.

Secondary considerations for participant selection included selecting teachers with a variety of educational backgrounds and statistical experiences. Many teachers who have served as readers and table leaders have attended or conducted professional development. In many cases, attendance is self-initiated and less formal than an undergraduate or graduate-level statistics course. Professional development workshops are likely to include pedagogical strategies for teaching statistics along with discussions related to statistical content. My belief was that

---

[3] One of the teachers in the study is not an AP Statistics reader; however, she was recommended by an AP teacher-leader because of her participation in leadership institutes and her leadership in providing professional development in statistics for teachers at local and state levels.

selecting teachers with a variety of professional development backgrounds would help to isolate characteristics of effective professional development. The 16 teachers attended between 1 to 30 professional development sessions, inclusive, in statistics, with a mean of 9.44 sessions and a median of 6.5 sessions. Nine teachers conducted professional development in statistics, with the number of sessions varying from 1 to 60, inclusive. The mean number of sessions conducted was 22.56 sessions, and the median was 23 sessions. The mean and median values, however, are deceptive because five teachers conducted 23 or more professional development sessions, and the remaining four teachers conducted 7 or fewer sessions.

In addition to displaying a variety of professional development experiences in statistics, teachers in this study have a mixture of formal, course-related experiences in statistics. Whereas no teacher with a statistics degree participated in this study, one teacher has a minor concentration in statistics, and seven teachers completed three or more formal courses in probability and statistics at the secondary, undergraduate, or graduate level. Only one teacher never completed a formal probability or statistics course. Two of the pilot study teachers had minors in statistics. Twelve teachers have undergraduate degrees in mathematics or mathematics education, and ten have graduate degrees in mathematics or mathematics education. Several teachers have undergraduate or graduate degrees in other fields, including marketing and advertising, chemical engineering, counseling, psychology, and varying exceptionalities. Two teachers have undergraduate degrees in science or engineering and completed coursework to obtain secondary mathematics certification. By selecting teachers with a wide variety of informal and formal educational experiences, my belief was that I could isolate characteristics of both formal and informal experiences that may have contributed to teachers' development of robust understandings of variation.

I selected teachers using general characteristics typically considered in phenomenological research, including gender (Moustakas, 1994). I selected equal numbers of male and female teachers from fourteen different states across the continental United States and Washington, D.C., including teachers from the North, East, South, West, and Midwest, in an attempt to select teachers that obtained their undergraduate and graduate degrees from different universities, taught in different school systems, and experienced different professional development programs. Three teachers have assumed leadership roles in the College Board organization in ways beyond table leadership, and seven have statistics-related publications, including textbook and textbook-related publications, magazine or web-based articles and activities, and workshop-related publications. Serving in leadership positions or publishing statistics-related work provides additional learning opportunities for teachers—opportunities not necessarily duplicated in other criteria used to select participants. As the diversity of teachers' experiences might suggest, my main goal was to have teachers with as many varied learning opportunities as possible.

Selection Process

To select participants, I began by contacting a table leader to obtain the names and e-mail addresses of current and past table leaders and readers. Using e-mail communication, I contacted approximately 125 secondary teachers in the continental United States who attended the 2006 and 2007 AP Readings and who voluntarily included their e-mail addresses on a list generated at the AP Reading. Teachers who were interested in hearing more about the study included contact information in their responses to the e-mail. I called or e-mailed approximately 45 teachers who expressed interest, described the study to them, and ascertained their level of interest. When teachers stated a preference for e-mail, I e-mailed the introductory script approved by the Office of Research Protections. I read the script to everyone else. As part of my conversations with teachers, I received recommendations and contact information for three additional teachers, who I then contacted in the same manner as described. After my initial contact, if a teacher expressed an interest in participating in the study, I e-mailed a password-protected questionnaire and asked him or her to return the completed questionnaire to me as soon as possible. The questionnaire, displayed in Appendix A, was the primary tool I used for participant selection.

To obtain participants who represented a wide variety of experiences, I waited to review questionnaires until I received fifteen. Initially I believed that I would have sufficient diversity of experiences with ten participants; fifteen responses seemed to be sufficient for making preliminary choices. From this pool of fifteen teachers and with input from another mathematics education researcher, I selected male and female participants who exhibited diversity in years of teaching mathematics in general and teaching statistics in particular; years as readers and table leaders; formal degree work; and experiences with attending or conducting professional development. In particular, I selected a pool of participants with a variety of experience configurations while ensuring that not all of the teachers had the same experiences. If two teachers had a similar configuration of experiences, I gave priority to the teacher in closer geographical proximity to me.

An example of the types of decisions I made while selecting participants can be illustrated with the experiences of Blake,[4] Eden, and a gentleman not selected for participation. These three teachers were among the fifteen who initially expressed interest in the study. The gentleman not assigned a pseudonym had 29 years of teaching experience and taught statistics for eight of those years. He attended five AP Readings, had undergraduate and graduate degrees in mathematics education, attended approximately 40 professional development sessions related to statistics, and conducted five professional development sessions in statistics. In comparison, Blake attended a total of seven AP Readings, one for which he was a table leader. He has an undergraduate degree in mathematics and a graduate degree in mathematics education. He attended approximately 30 professional development sessions related to statistics. Eden taught for

[4] Blake, Eden, and all other participant names are pseudonyms.

30 years, including statistics for 10 of those years. She attended seven AP Readings, conducted two professional development sessions, and served as a statistical consultant for her high school. Arguably, when Blake and Eden are considered together, their experiences overlap with the unnamed gentleman's experiences to a large extent. Additionally, Blake and Eden add unique experiences to the study—experiences that potentially affected their statistical understandings. Eden has an undergraduate degree in chemistry, with a minor in physics. Blake has been teaching statistics for 18 years, including seven years before the AP Statistics course existed and is a table leader. Eden and Blake added unique characteristics to the study, while also having experiences similar to those of the gentleman. Other participants also had experiences that overlapped with those of the gentleman. Because they had unique experiences, Blake and Eden were selected to participate in the study; because the gentleman had common experiences that were experienced by others with additional unique experiences, he was not selected for the study.

From the initial pool of questionnaire respondents, I selected ten participants and e-mailed the teachers to inform them of their selection and to request additional information from them. Nine of the ten teachers returned a signed consent form to me and participated fully in the study. After these ten individuals were contacted, I continued to receive questionnaires. I expanded the number of participants if an individual had a unique characteristic or experience that may have related to statistics learning and that was not represented or experienced among those teachers already selected. Additional experiences included engineering coursework, degree work for teaching students with varying exceptionalities, degree work for an MBA, unique statistical publications, leadership separate from the AP Reading and the College Board, and leadership roles in AP statistics different from table leading. Adding participants with these experiences brought the total number of participants to sixteen. Three teachers participated in a pilot study, and a total of thirty teachers returned signed informed consent forms stating a willingness to participate in the study. Teachers who agreed to participate in the study and who were selected for participation were compensated monetarily for their participation.

Seidman (2006) suggests two criteria for determining sample sizes for qualitative studies: sufficiently large yet not exceeding saturation. Using these criteria for this study, the sample size is sufficiently large to reflect the variety of experiences for teachers who experience the phenomenon of coming to understand variation yet sufficiently small so that the point of saturation is not exceeded. These 16 teachers, eight male and eight female, studied statistics both formally and informally and had a variety of experiences in learning and teaching statistics, a variety of educational and cultural experiences, and a variety of leadership experiences in AP Statistics.

Data Collection to Address Research Question One

To obtain information about teachers' conceptions of variation and to establish whether they have robust understandings of variation, I conducted a 90- to 120-minute semi-structured content interview with each of the 16 teachers. During the interview, teachers responded to a set of tasks that required them to reason about variation from data-centric, modeling, and design perspectives. Each task statement was purposefully vague to allow teachers to approach the task from multiple perspectives, allowing insights into aspects of variation most prominent for each teacher. Each task was designed to allow teachers to exhibit formal, abstract reasoning from individual perspectives and integrated reasoning across perspectives through the process associated with the design of tasks using the SOLO Model. As a collection, the tasks are extremely open-ended and not the kind of tasks that many teachers are likely to have encountered previously; however, the tasks are approachable with introductory-level statistics knowledge.

The SOLO Model in Relation to Data Collection

The SOLO Model provides a tool for investigating teachers' understandings of variation through analysis of their responses to the interview tasks. The model aids in classifying different levels of responses—responses that provide insight into statistical understandings (Biggs & Collis, 1982). The interview tasks and the lines of questioning related to each task in this study were designed to elicit reasoning in the formal mode and to allow determination of levels of reasoning for both cycles of levels of the formal mode.

Figure 4-1 shows an abbreviated description of the four elements (variational disposition, variability in data for contextual variables, variation and relationships among data and variables, and the effects of sample size on variability) from each perspective that emerged from analysis of the pilot interviews and that were subsequently used to establish levels of reasoning in the SOLO Model.[5] Integrated reasoning involving the four elements from a particular perspective is indicative of relational reasoning within that perspective. In general terms, the four elements correspond with an expectation for and consideration of variation in statistical problem solving, consideration and exploration of variation related to specific contextual factors, consideration of variation to reveal relationships among data and variables, and consideration of the effects of sample size on data analysis. A description of the analysis that led to identifying these four elements appears in the section titled "Data Analysis to Address Research Question One." Discussion of how the interview tasks were designed to educe teachers' understanding of variation uses the shorthand notation for elements of perspectives shown in Figure 4-1 (e.g., DP1, DCP1, MP1, etc.).

[5] Chapter 6 contains detailed descriptions of each element.

Elements and Reasoning Indicative of Robust Understanding of Variation			
Perspective / Element	**Design Perspective**	**Data-Centric Perspective**	**Modeling Perspective**
Variational disposition	DP1: Acknowledging the existence of variability and the need for study design	DCP1: Anticipating reasonable variability in data	MP1: Anticipating and allowing for reasonable variability in data when using models
Variability in data for contextual variables	DP2: Using context to consider sources and types of variability to inform study design or to critique study design	DCP2: Describing and measuring variability in data for contextual variables as part of exploratory data analysis	MP2: Identifying the pattern of variability in data or the expected pattern of variability for contextual variables by
Variability and relationships among data and variables	DP3: Controlling variability when designing studies or critiquing the extent to which variability was controlled in studies	DCP3: Exploring controlled and random variability to infer relationships among data and variables	MP3: Modeling controlled or random variability in data, transformed data, or sample statistics
Effects of sample size on variability	DP4: Anticipating the effects of sample size when designing a study or critiquing a study design	DCP4: Examining the effects of sample size through the creation, use, or interpretation of data-based graphical or numerical representations	MP4: Anticipating the effects of sample size on the variability of a sampling distribution

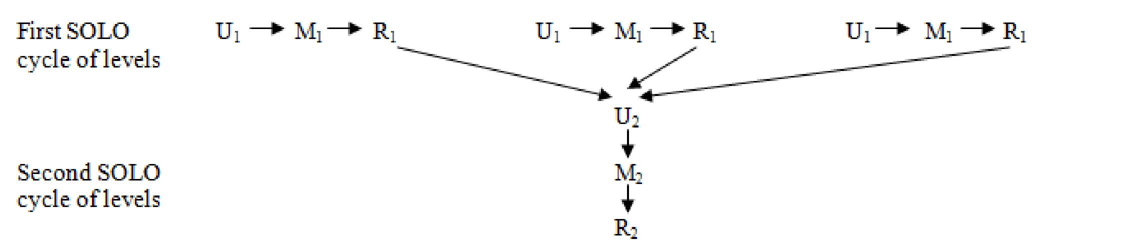

Figure **4-1**: Elements and Indicators of Robust Understanding as Two Cycles of Levels in the SOLO Model.

The Consultant Task

The first task presented to teachers was the Consultant Task[6] shown in Figure 4-2. Although I describe the task by attending to the design perspective first, the order of questions used with any one teacher was determined by the direction taken by the teacher in response to the task statement. I anticipated that the task statement could elicit reasoning from any of the three perspectives. By providing no information about how administrators selected exams, teachers could respond that no conclusion is possible because the samples might be biased, which would lead into reasoning from the design perspective. Because the only summary measures included in the statement were the average scores for each sample, teachers could respond that they needed additional information about the data to form a conclusion, leading into reasoning from the data-centric perspective. Finally, by presenting information about means and asking for a comparison between consultants, teachers could respond by suggesting that they would conduct a test of inference to form a conclusion, which would lead into reasoning from the modeling perspective.

To improve students' test scores on state assessments, administrators from a large school district require students to take practice exams. Two outside consultants create and score the open-ended questions from these exams. Although both consultants use the same rubric to score student responses, the administrators suspect that the consultants do not interpret and apply the rubric in the same way, resulting in differences in scores between the exams scored by the two consultants. The consultants' contract with the district is up for renewal, and the administrators are trying to decide if they should renew the contract. They decide to use the most recent practice exam to compare the scores assigned from each consultant and to decide whether there is a difference in the way the exams were scored. The administrators select 50 exams scored by the first consultant and 50 exams scored by the second consultant. They find that the average score for the 50 exams scored by the first consultant was 9.7 (out of a possible 15 points), while the average score for the 50 exams scored by the second consultant was 10.3 (out of a possible 15 points). What should the administrators conclude about the scores assigned by these two consultants?

Figure **4-2**: The Consultant Task.

[6] The Consultant Task is an original task created by the researcher for the content interview.

Design Perspective

Teachers tended to offer critiques of the data collection methods used by the administrators. In particular, when teachers noted missing information such as sampling technique, I asked them to describe why the information was important, how the information would enable them to answer the administrators' question, and what conclusions they could draw in the absence of that information. Asking why additional information was important provided opportunities for teachers to express recognition of the omnipresence of variability (DP1), to offer concerns related to the effects of potential sources of variation specific to the context (DP2), and to express a need to know how those potential sources of variation were controlled (DP3). By considering how the additional information would help and probing for justifications to support responses, teachers could reason about design elements in general terms that transcend multiple contexts and provide information of their reasoning in the formal mode. Lastly, through attention to what conclusions could be drawn strictly from the given information, further insight of teachers' variational dispositions could be obtained.

After teachers commented on the design implemented by the administrators, I asked them to describe and defend the design they would have instituted to answer the administrators' question. Through this question, teachers were given an additional opportunity to reason about the DP1, DP2, and DP3 elements offered in their critiques as well as an opportunity to consider additional controlling strategies (DP3) and to describe the effects of sample size in their proposed designs (DP4).

Data-Centric Perspective

When teachers addressed how they would analyze the administrators' data, they typically began by requesting information about the variation in scores to complement the given information about centers. I asked them to describe why they needed information about variation, which offered insights into their variational disposition (DCP1). I provided values for measures of variation and dotplots of the data separately and only after teachers requested the information. The summary values and dotplots are shown in Figure 4-3. From the summary measures, I asked teachers to describe the distributions they would expect to see associated with the given summary values to inform how they described variation and interpreted standard deviation (DCP2). When teachers examined the dotplots, they tended to notice a discrepancy between Consultant Two's summary measures and dotplot.[7] I asked teachers to estimate values for the mean and standard deviation of the data displayed in the dotplot and to explain how they estimated the values to inform how they used data to reason about variation (DCP2). I also asked teachers to reason about what the administrators could conclude from the given information, which allowed them to reason about variation within each distribution (DCP2) and to compare variation between distributions using summary measures and graphical representations of the data (DCP3).

[7] The discrepancy in Consultant Two's scores appeared from one misentered test score value, 150, in place of a score of 15. The summary measures for Consultant Two's scores were calculated using the value of 150. The dotplot of Consultant Two's scores only displays scores on the interval from 0 to 15.

Random 50

	Consultant		Row Summary
	Consultant_1	Consultant_2	
Value	9.7 3.3760864	10.3 20.205324	10 14.4152

S1 = mean ( )
S2 = stdDev ( )

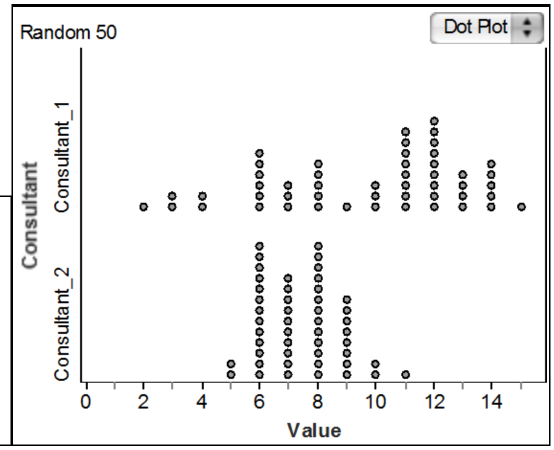

Figure **4-3**: Summary Values and Dotplots for Sample Exam Scores.

Teachers had additional opportunities to reason about variation from the data-centric perspective by addressing all four elements of the data-centric perspective in response to questions such as those asked of the original data for the corrected summary measures and dotplot for Consultant Two's scores, which are displayed in Figure 4-4.

Random 50

	Consultant		Row Summary
	Consultant_1	Consultant_2	
Value	9.7 3.3760864	7.6 1.7261494	8.65 2.8687624

S1 = mean ( )
S2 = stdDev ( )

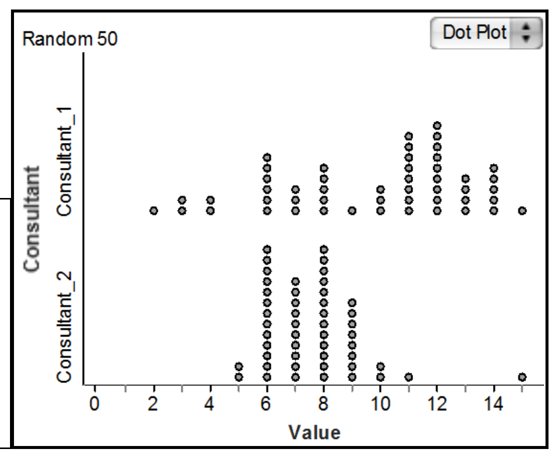

Figure **4-4**: Corrected Summary Values and Dotplots for Sample Exam Scores.

As a second part of the Consultant Task, I asked teachers to describe expected differences between size-15 samples and either size-30 or size-50 samples to ascertain their perceptions of the effects of sample size on variability (DCP4). I then gave teachers the scores of 15 exams randomly selected from those scored by each consultant and asked them again to determine whether there appeared to be a difference in the way the consultants scored the exams. The scores are displayed in Table 4-1. Because these sample data were presented in tabular form and available in lists on a TI-84 graphing calculator if teachers expressed a desire to use a calculator, teachers chose summary measures, graphical representations, or strategies to use in analyzing the data and reasoning about issues of representing, measuring, and describing variation (DCP2) for each consultant and to compare variation between the two consultants (DCP3). Through the variety of situations and questions present in the Consultant Task, teachers could represent, measure, describe, and reason about variation in multiple ways.

Table **4-1**: Exam Scores for Randomly Selected Exams Scored by the Two Consultants.

Consultant 1	Consultant 2
8	14
4	13
3	11
7	13
6	9
4	12
3	11
10	7
8	6
3	8
15	1
5	12
3	13
5	10
2	11

Modeling Perspective

As part of their comparison of scores for the two consultants, many teachers discussed formal inferential comparisons of the two groups, which provided opportunities for reasoning about distributions to model the pattern of variability in the data (MP2) and to model sampling distributions for sample statistics from samples of a given size (MP3). I also asked teachers to respond to questions related to their expectations for additional samples of different sizes selected from the same population (MP4), which informed how they balanced the ideas of sample representativeness and sample variability (Rubin, Bruce, & Tenney, 1990). The way teachers expressed their conclusions provided insight into their variational disposition (MP1). Although teachers had the opportunity to reason from the modeling perspective in the Consultant Task, the Caliper Task was specifically designed to elicit reasoning about variation from the modeling perspective.

The Caliper Task

Figure 4-5 and Figure 4-6 show the graph and question initially presented for the Caliper Task.[8] By failing to mention any kind of context for the data, I anticipated that teachers would initially reason from the design perspective or from the modeling perspective. Teachers who first attended to considerations for making a prediction exhibited reasoning from the modeling perspective, whereas teachers who expressed a need to know the context of the data to consider the nature of or expected pattern of variability for the data reasoned from the design perspective.

[8] The Caliper Task is an original task created by the researcher for the content interview.

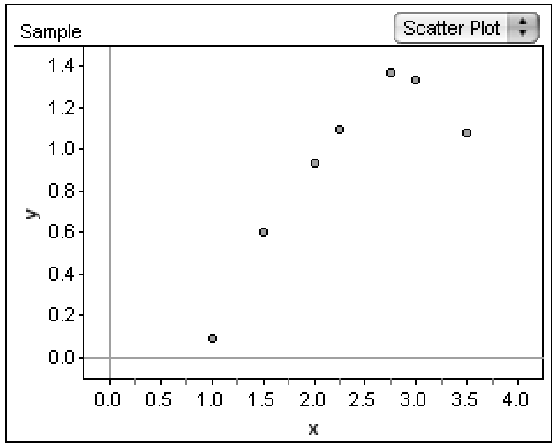

Figure **4-5**: Initial Graph for the Caliper Task.

Imagine that one of your students asked you to look at this graph of data their lab partners collected during a science lab. The student's partners did not give the names of the variables represented by *x* and *y*. The student asks you how they might use this graph to predict a value for *y*, given a value of 4 for *x*. What would you say to the student?

Figure **4-6**: Initial Question for the Caliper Task.

Design Perspective

Teachers who expressed a need to know the context were asked to articulate the reasons behind their requests and to justify the legitimacy of their concerns related to context. Responses to these questions tended to provide information about a teacher's variational disposition (DP1) or how the teacher used context to consider the variability expected in data for each variable and in the relationship between variables (DP2). The small sample size offered additional opportunities for teachers to reason about the effects of sample size on data variability, as did a larger sample presented later in the Caliper Task (DP4). I gave teachers information about the context after they reasoned about the data absent context: The data in the scatterplot are measurement data for an object manufactured to have a specified length measured in centimeters as the explanatory variable. The corresponding response value is a student's Vernier caliper measurement of the same object to the nearest thousandth of an inch. I asked teachers to describe reasons behind why the data did not exactly match the pattern of the known theoretical relationship between inches and centimeters, which provided additional opportunities for teachers to reason about sources of variability in the given context (DP2) and to offer suggestions for how design strategies could have controlled the variability from those sources (DP3).

Data-Centric Perspective

When teachers speculated about various models the student could use to make a prediction from the given data, I encouraged them to describe the conditions under which each response would be appropriate and to describe how the student could decide which model was the best to use to make a prediction. Their responses suggested how teachers focus on the aggregate of the data to reason about the pattern of the variability in the data (DCP2) and how they use the correlation, the coefficient of determination, or a residual plot to evaluate each model (DCP2) and select the best model (DCP3).

Modeling Perspective

In order to make a prediction for y in response to the student's question, many teachers offered a function they would use to model the data. In addition to asking how they formulated the model, I asked teachers to defend their choice of model, to describe the goodness of their models' fit, and to articulate their prediction along with why their prediction made sense. Articulating their predictions allowed me to observe teachers' reasoning about reasonable variability (MP1). Defending their choice of model and describing goodness of fit informed how teachers identified patterns of variability in the data and model use to explain variability in data

(MP2) as well as how they reasoned about random variability by examining deviations from the model (MP3). If teachers suggested models that were not linear, I asked them how they could fit a line to the data to inform their considerations about transforming data to improve fit and explain more variability (MP3).

After I told teachers the context, I asked them to reconsider their responses and their reasons for any changes in their responses in light of the context. In particular, I asked probing questions to elicit reasoning about how context affected model selection (MP2) and reasoning relating to the amount of variability allowed in predictions (MP1). To establish how teachers reasoned about the effects of sample size on bivariate data (MP4), I asked questions similar to those for the original scatterplot in response to the scatterplot shown in Figure 4-7.

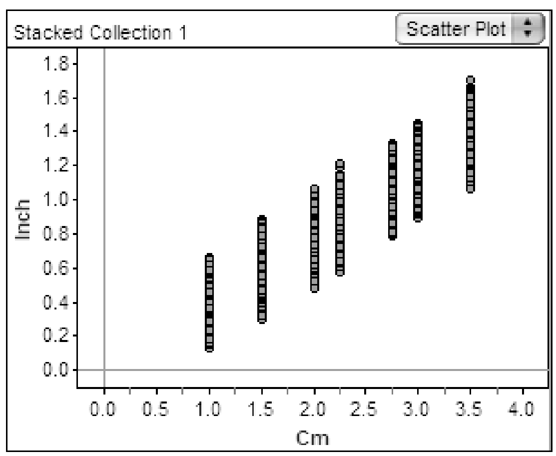

Figure **4-7**: Scatterplot Resulting From a Larger Data Set.

In addition to describing and justifying a model to best fit the bivariate data, I asked teachers to describe a reasonable model to fit the univariate distributions formed at the seven discrete centimeter measurements to gather further information about the extent to which context influences pattern expectation (MP2). I also asked teachers to consider the pattern of variability in univariate distributions under conditions that would produce different lines of best fit, which informed the meaning they associated with “good” fit (MP3).

If the content interview took no more than seventy-five minutes at this point, I asked teachers to describe different measures of variation displayed in the linear regression output shown in Figure 4-8. In particular, I asked teachers to reason about the coefficient of determination in light of the apparent deviation of data from the line. I also asked teachers to reason about *s*, the standard deviation of the residuals. Asking these questions provided information about how teachers reasoned about summary measures for bivariate data. I provided a residual plot if teachers indicated they needed one in order to reason about goodness of fit.

Regression Analysis: Inch versus Cm

```
The regression equation is
Inch = 0.000329 + 0.394 Cm

Predictor        Coef    SE Coef        T      P
Constant    0.0003289  0.0006014     0.55  0.584
Cm           0.393590   0.000248  1586.27  0.000

S = 0.0748784   R-Sq = 94.7%   R-Sq(adj) = 94.7%

Analysis of Variance

Source              DF     SS     MS           F      P
Regression           1  14108  14108  2516267.73  0.000
Residual Error  139998    785      0
Total           139999  14893
```

Figure **4-8**: Regression Output for the Data Displayed in Figure 4-7.

The Handwriting Task

The Caliper Task was designed for teachers to reason from the modeling perspective, and the Handwriting Task was designed to elicit reasoning from the design perspective. There are two parts to the Handwriting Task,[9] as shown in Figure 4-9 and Figure 4-10. The first part relating to

[9] The first part of the Handwriting Task stemmed from an idea posted on the AP Statistics electronic discussion group by Joshua Zucker on October 13, 2006 . The second part of the task stemmed from an idea posted by Floyd Bullard on October 11, 2006. Both ideas originated from an article published in the Washington Post on October 11, 2006 (Pressler, 2006).

the quote shown in Figure 4-9 typically was not discussed if less than ten minutes remained in the interview. Figure 4-9 presents a quote from a newspaper article in which the author describes parts of an experimental study but fails to describe key features of the study. I asked teachers to describe the apparent or missing features they would expect their students to notice and to describe any features they noticed that they would not necessarily expect their students to notice. For each feature, I asked teachers to describe the importance of the feature and the benefit or detriment of the features to gain insight into the types of variables teachers would attribute to the setting (DP2) and the aspects of control they would stress for the given setting (DP3), including aspects of randomization and sample size (DP4).

"In one of the studies, Vanderbilt University professor Steve Graham, who studies the acquisition of writing, experimented with a group of first-graders in Prince George's County who could write only 10 to 12 letters per minute. The kids were given 15 minutes of handwriting instruction three times a week. After nine weeks, they had doubled their writing speed and their expressed thoughts were more complex. He also found corresponding increases in their sentence construction skills" (Pressler, 2006).

Figure **4-9**: Excerpt to Critique Design.

"When adults are given the same composition written in good handwriting and poor handwriting, 'they still give lower grades for ideation and quality of writing if the text is less legible,' he said" (Pressler, 2006).

Figure **4-10**: Excerpt to Create Design.

In the quote displayed in Figure 4-10, the author conjectures that a relationship exists between handwriting quality and composition scores assigned by adults. I asked teachers to describe how they would design a study to test the stated conjecture. As teachers created their designs, I asked them to explain the decisions they made and their reasons for making those decisions. The questions targeted issues of replication, randomization, and control in relation to variation to gather information about teachers' reasoning about their expectation of variation (DP1), consideration of variation sources (DP2), attempts to control variability (DP3), and reasoning about sample size (DP4). If teachers did not mention sample size or blocking, I asked them to consider each as part of their design to provide further information about their consideration of control in study design (DP3) and control related to sample size (DP4).

As the preceding descriptions of the tasks and the lines of questioning related to each task suggest, aspects of reasoning from the design, data-centric, and modeling perspectives were touched upon in several of the tasks, thereby providing opportunities for teachers to integrate reasoning about variation from the three perspectives throughout the entire problem-solving session and to reason about variation in different contexts.

Data Analysis to Address Research Question One

The content interviews were the primary sources of data for determining teachers' conceptions of variation and for determining those teachers whose reasoning was indicative of robust understandings of variation. To analyze the interview data, I used annotated transcripts of interviews with each teacher and examined teachers' statistics course syllabi[10] and content-focused excerpts from their context interviews. The analysis consisted of multiple stages, with each successive stage building on the previous one.

Pre-Pilot and Pilot Study Analysis

Prior to data collection for the 16 teachers in the study, I conducted a pre-pilot study with three teachers, followed by a pilot study with three teachers. I piloted the content interview schedule and tasks with AP Statistics teachers who were not considered for inclusion in the study. Some are friends who I have known for a number of years, and although it seemed inappropriate to include them in the participant pool for the main study, I hypothesized that including them in

[10] Fifteen of the 16 teachers in this studied provided either paper or electronic copies of their AP Statistics syllabi.

the pilot studies could inform the design of tasks without compromising the participant pool. Other pilot study teachers were AP Statistics teachers who are not currently leaders in the AP Statistics program (using “leader” as it is defined in the “Selection Process” section).

The main purpose of the pre-pilot and pilot studies was to determine whether the content interview tasks and prompts could elicit evidence of teachers’ reasoning about variation from the design, data-centric, and modeling perspectives and elicit evidence of relational-level reasoning in the formal mode. Using insights about what it means to understand variation extrapolated from existing expository and research literature that mainly draws from my synthesis of the writings of Franklin and colleagues (2007); Garfield and Ben-Zvi (2005); Garfield, delMas, and Chance (2007); Moore (1990); Reading and Shaughnessy (2004); Reading and Reid (2006); and Wild and Pfannkuch (1999), I created a list of indicators for each perspective that suggested reasoning indicative of someone who understands variation. Table 4-2 contains the original list of indicators.

Table **4-2**: Initial List of Indicators of Robust Understanding.

Perspective	Indicators
Design	D1: Anticipation of the omnipresence of variability and acknowledgement of natural variability, particularly when designing a study and making conclusions from the study D2: Anticipation of possible sources of variability (such as measurement variability) in the context of the study and description of the differences in the magnitude of the effects various sources may have on the variability in measured characteristics D3: Anticipation of the effects of sample size on both the variability of the sample and on the statistics characteristic of the sample (statistics used to make inferences about parameters) for designing a study or in consideration of a study design D4: Acknowledgement of controllable and uncontrollable variability, such as explicating the benefit of using random assignment or random selection of observational/experimental units in the context of a particular study, and the need for control to be able to isolate systemic variation from random variation
Data-Centric	DC1: Creation, use, or interpretation of various representations of data to highlight patterns in the variability of the data and to focus on the aggregate features of the data DC2: Calculation of summary statistics values or acknowledgement of the utility in having summary measures for measuring the variability in the data or the use of and interpretation of appropriate summary statistics (including measures of variation such as range, interquartile range, and standard deviation for univariate sets of data or correlation and the proportion of variability for bivariate sets of data) to describe holistic features of the distribution DC3: Estimation of measures of variability for a set of data based upon characteristics of the data distribution, including shape, center, and the presence of outliers for univariate sets of data, or correlation and the proportion of variability for bivariate sets of data DC4: Use of summary statistics measures, including measures of variation, to make group comparisons and to examine the variability within and among groups
Modeling	M1: Use of a normal distribution to model patterns of variation for symmetric, bell-shaped data distributions (along with the corresponding use of other probability distributions for nonnormal distributions) and use of the characteristics of a normal distribution (based on center and spread and the effects of sample size on spread) to examine characteristics of the data, including invocation of the empirical rule to estimate variability by using knowledge that approximately 68% of the data lies within one standard deviation of the mean, approximately 95% of the data lies within two standard deviations of the mean, and approximately 99.7% of the data lies within three standard deviations of the mean M2: Use of appropriate models or transformations to account for the variability in data and to isolate the signal from the noise (i.e., variation in the data from the signal or expected pattern of data) for univariate or bivariate sets of data M3: Use of deviations from the model fit to variable data that deviates from the expected pattern to describe the goodness of fit of the model M4: Use of models to make predictions or statistical inferences from the data while allowing for variability with predictions or when interpreting results

Pre-pilot and pilot content interviews were videorecorded, transcribed, and annotated prior to analysis. Analysis consisted of matching passages of individuals' reasoning about variation to the indicators in Table 4-2 to determine whether the tasks evoked reasoning aligned with indicators and to determine if additional indicators needed to be added to the table. One of the changes to the interview schedule occurred subsequent to analysis of Consultant Task data with respect to the indicator listed as DC1 in Table 4-2. The indicator, "creation, use, or interpretation of various representations of data to highlight patterns in the variability of the data and to focus on the aggregate features of the data," suggests that teachers might create or use multiple data representations to analyze data. However, the original task gave teachers dotplots for both size-50 samples and size-15 samples, with most teachers reacting by reasoning strictly from the dotplots. Figure 4-11 displays the dotplots used in the pilot studies for the size-15 samples. To provide a higher probability for evoking reasoning about multiple representations, the data were presented in tabular form for the main study (see Table 4-1), with the result that some teachers created and interpreted dotplots, boxplots, and summary measures of the data.

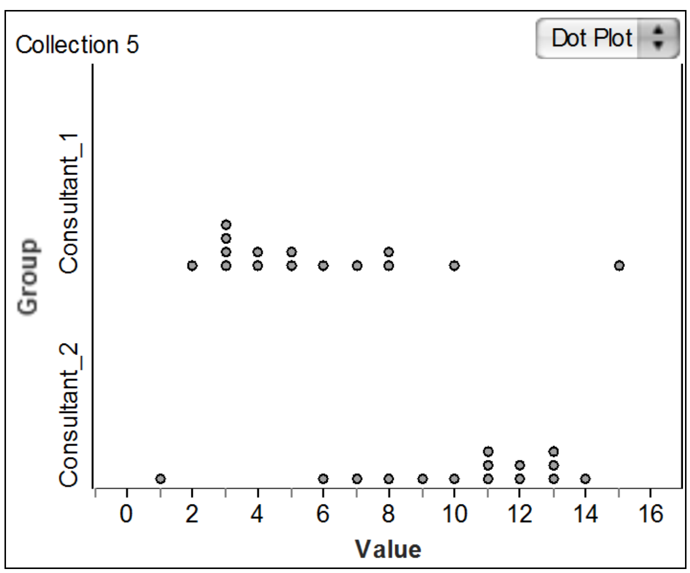

Figure **4-11**: Original Presentation of Size-15 Samples in the Consultant Task.

Analysis of pre-pilot and pilot data also led to the development of a framework consisting of a set of lists of observable *indicators* for perspectives crossed with four considerations or aspects of variability that transcend perspectives, hereafter referred to as *elements*. (See Table 4-

3, which appears at the end of this section.) To exemplify both how indicators, elements, and understandings are related and how the framework for robust understanding of variation developed from its roots in extant literature and empirical data, I summarize the development of several indicators related to one element. Similar evolution led to the final collection of elements and indicators, which I discuss in detail in Chapter 6.

The element of a *variational disposition* originated from statistics education literature that focuses on the role of variation in statistics and in statistical problem solving (e.g., Cobb & Moore, 1997; Franklin et al., 2007). In particular, much discussion attends to the "omnipresence of variability" and the importance of recognizing it. Statisticians opine that the discipline of statistics arises from a need to deal with the "omnipresence of variability" (e.g., Cobb & Moore, 1997; Moore, 1990; Snee, 1990; Wild & Pfannkuch, 1999). Most expository literature and curricular recommendations related to the learning and teaching of statistics identify recognition of the omnipresence of variability as foundational for students' development of increasingly sophisticated understandings in statistics (e.g., Franklin et al., 2007; Garfield & Ben-Zvi, 2005; Garfield & Ben-Zvi, 2008). The authors of the GAISE report articulate additional needs related to recognizing the omnipresence of variability in relation to study design—needs for anticipating variability when formulating statistical questions and for acknowledging variability when considering methods of data collection (Franklin et al., 2007). This body of literature led to the identification of an indicator for anticipating and acknowledging variation in study design.

Although anticipation and acknowledgement of variation were originally associated only with reasoning from the design perspective, analysis of pre-pilot and pilot data revealed related observable indicators in reasoning that included anticipation and acknowledgement of variation from data-centric and modeling perspectives. Teachers who participated in the pre-pilot and pilot studies displayed *variational dispositions*—states of mind in which variation is expected. Evidence of a variational disposition arose when one of the pilot-study teachers discussed the standard deviation value of 20 for Consultant Two's scores in the Consultant Task. The teacher considered both context and distributional characteristics of data to question the legitimacy of the value. She argued that scores ranging in value from 0 to 15 could not produce the given standard deviation value. Using the dotplot of Consultant Two's scores, she approximated the correct value to be between one and two. She conveyed a tolerance and expectation for variation but concluded that 20 was an unreasonable value for the standard deviation. She acknowledged the existence of variability and identified problematic characteristics of variability attributed to data in her reasoning about variation from the data-centric perspective. Her reasoning contributed to the identification of indicators for a variational disposition from the data-centric perspective. In particular, her reasoning contributed to the development of an indicator for anticipating reasonable variability in data by considering the context of data and an indicator for anticipating reasonable variability in data by recognizing unreasonable variability in data. A third indicator related to anticipating reasonable variability, recognizing that data descriptions should include descriptions or measures of variability (and center), originated from Garfield and Ben-Zvi's (2005) framework for teaching and assessing reasoning about variability. This indicator was grouped with the others through its connection to anticipating variability with respect to data.

A second teacher expressed anticipation of variability and allowance for variability in response to the Consultant Task as he reasoned about whether there was a difference in scoring based on the values of the means. He anticipated that consultants' scores are likely to differ, but he noted that means of 9.7 and 10.3 may not indicate a true difference in scoring. He articulated a need for information about the spread of scores to determine whether the difference was significant. In his reasoning, he showed evidence of anticipating and expecting variation while reasoning about variation from the modeling perspective as he talked about a *t*-distribution from which to make inferences. His reasoning contributed to the identification of an indicator of a variational disposition from the modeling perspective: anticipating and allowing for reasonable variability in data when using models for making inferences from data. A second indicator, anticipating and allowing for reasonable variability in data when using models for making predictions from data, stemmed from discussion in the GAISE report about allowing for variability in looking beyond data (Franklin et al., 2007).

The reasoning of these two teachers illustrates a variational disposition from different perspectives. Through analysis of teachers' responses and considerations of statistics education literature, a total of four considerations of or aspects of variability that transcend perspectives, hereafter referred to as *elements*, emerged from the data, along with detailed characteristics of indicators for each element. The four elements that elicit reasoning about considerations or aspects of variation across perspectives are: variational disposition, variability in data for contextual variables, variability and relationships among data and variables, and the effects of sample size on variability. I discuss these elements in detail in Chapter 6.

In-Depth Analysis of Data in Response to Research Question One

Analysis of the data for the main study began with a preliminary stage in which I examined each content interview and syllabus for each teacher before I conducted his or her first context interview and again before the second context interview to determine if content-focused questions needed to be asked in an attempt to assure a high probability for obtaining evidence of teachers' reasoning about variation for each element across each perspective. The content interview tasks were created through the alternative process of SOLO, and elements of reasoning that tasks were designed to elicit were used in the analysis of interview data because, "the SOLO taxonomy not only suggests an item writing methodology, but the same taxonomy can be used to score the items" (Hattie & Purdie, 2003, p. 17). In this preliminary stage, I listened to the audio recording of the teacher's content interview and noted evidence of reasoning that corresponded with particular elements for each of the three perspectives. If evidence was lacking for any element-perspective pair, I incorporated questions intended to elicit reasoning about the missing element in the context interviews. As an example, consider Haley's Content interview and Context I interview. Haley's work with the Consultant Task did not naturally lead to her reasoning about the effect of an outlier on the variation of a distribution. During her Context I interview, I made sure to ask about her learning experiences related to the effects of an outlier. In particular, I asked about her experience in learning "how the outlier affects the standard deviation" (Haley, Context I, Lines 627-628). By asking context questions focused on particular content, I was able to gain insights into how each teacher thought about that content. In Haley's case, I was able to learn more about her reasoning about variation from the data-centric perspective for the element of variability in data for contextual variables.

For my first pass through the data after I completed data collection, I created a matrix for each teacher with columns labeled by perspective and rows labeled by element. Any time the teacher exhibited reasoning that included evidence of an element or reasoning related to an element from one or more perspectives, I copied and pasted the passage from the annotated interview transcript to the appropriate cell(s) of the matrix and spaced interview passages in temporal sequence across perspectives for each element. I also wrote a summary of the indicators evidenced by each passage and made note of whether the teachers' reasoning was prompted by my questions. Figure 4-12(a) displays a portion of Everett's matrix for the variational disposition element and begins with an entry for reasoning with a variational disposition from the data-centric perspective, which is shown in Figure 4-12(b). The next occurrence of Everett's reasoning with a variational disposition included aspects of reasoning from the design perspective and from the data-centric perspective, whereas the next passage offered reasoning from the modeling perspective. The intent behind creating matrices for each teacher was to provide a coherent picture of the teachers' reasoning from which to obtain an image of dominant elements or perspectives in reasoning as well as a sense of teachers' integrated reasoning across perspectives. Annotated interview transcripts and matrices for some teachers were discussed with another mathematics education researcher until agreement was reached on the placement of evidential passages.

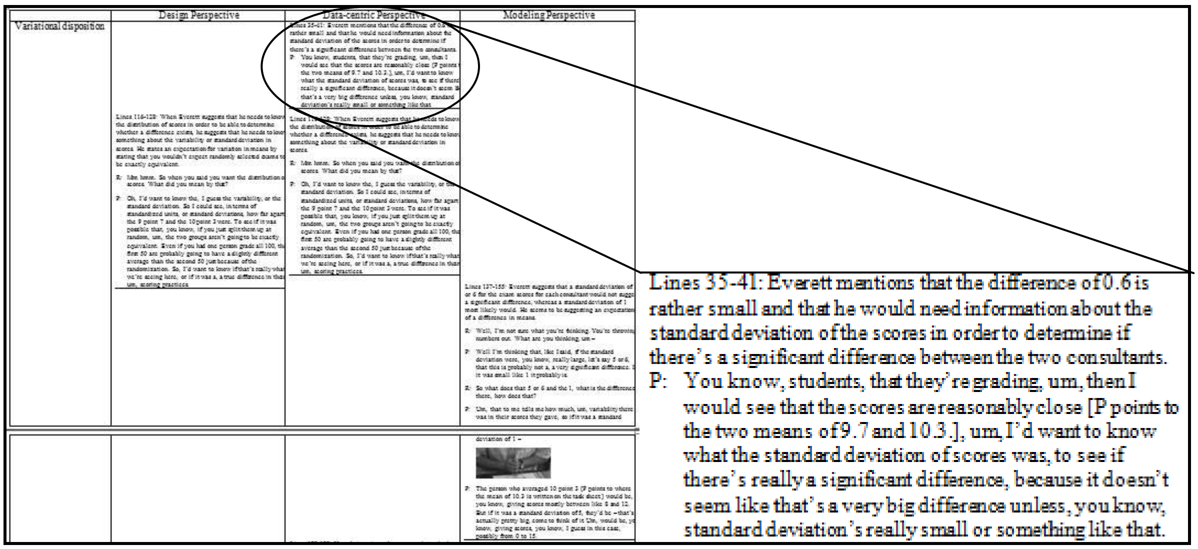

Figure **4-12(a) and 4-12(b)**: Example of the Matrix Layout (a) and Matrix Entry (b).

After I created matrices with evidence related to each teacher's reasoning about variation, I compared and contrasted passages illustrative of teachers' reasoning for each of the four elements and from each of the three perspectives to then look for commonalities or differences in reasoning among elements and perspectives and patterns in reasoning. For each teacher, I made conjectures about their conceptions of variation, identified evidence of understandings of variation, and recorded summaries. Figure 4-13 shows a portion of a summary written to describe Everett's reasoning about variation from the data-centric perspective and illustrates Everett's tendency to focus on finding signals, such as the average score assigned by consultants in the Consultant Task, within the noise of data.

After Everett is satisfied that appropriate methodology generated data representative of a larger population or populations via randomization, he explores data to identify any potential signals that can be observed through the noise of the data before attempting to establish the significance (or not) of the signals. In general, there are two types of signals to which Everett attends–signals in the form of statistics representing one or more populations and signals about the relationships that exist among two or more variables. Everett reasons from the data-centric perspective when he calculates, represents, or interprets characteristics of data or makes informal inferences from a sample or samples.

Figure **4-13**: Sample Summary Statements of Everett's Data-Centric Reasoning.

As I analyzed teachers' data, I mildly revised descriptions of the indicators based upon characteristics of teachers' responses that were not accounted for in the initial descriptions. For example, the final MP2 indicator of "considering or creating distribution-free models to explore contextual variables" was modified from "considering models to explore contextual variables" because of teachers' suggestions to use randomization tests to determine whether consultants scored exams differently in their responses to the Consultant Task.[11] I reread participants' responses and updated participants' summaries based upon the revised level descriptions using the constant comparative method articulated by Glaser and Strauss (1967), and I considered the need for further refinement of the list of indicators. Although indicators required revision, the main elements of the matrix remained constant throughout data analysis. The table of elements and indicators for each perspective allowed responses to the interview tasks to be compared against indicators to determine levels of understanding of variation using SOLO from teachers' reasoning about variation. The complete list of indicators for each element that emerged from the data is shown in Table 4-3.

During the course of revisiting characteristics of reasoning, different patterns of reasoning associated with different conceptions of variation began to emerge from the data. As I made continued comparisons through multiple additional passes through matrices and summaries of teachers' reasoning, distinguishing features of different conceptions were delineated. Analysis

[11] For example, Everett described a randomization test as taking combinations of the original 100 consultants' scores to form two new size-50 samples for each consultant. He therefore randomly selects 50 tests from the combined 100 tests, and considers characteristics of the newly formed samples in comparison with the observed difference in means of 0.6 for the size-50 samples.

continued until there no longer existed any conflicts for describing teachers' conceptions of variation. Details about the conceptions appear in Chapter 5.

Table **4-3**: Indicators of Robust Understanding of Variation.

	Design Perspective	Data-Centric Perspective	Modeling Perspective
Variational disposition	DP1: Acknowledging the existence of variability and the need for study design in • controlling the effects of variation from extraneous variable(s); • including considerations of variation for variable(s) of interest during data analysis; or • using sample statistics to infer population parameters for the variable(s) of interest	DCP1: Anticipating reasonable variability in data by • considering the context of data; • recognizing that data descriptions should include descriptions or measures of variability (and center); or • recognizing unreasonable variability in data (e.g., that which could result from a data entry error)	MP1: Anticipating and allowing for reasonable variability in data when using models for • making predictions from data or • making inferences from data
Variability in data for contextual variables	DP2: Using context to consider sources and types of variability to inform study design or to critique study design by • considering the nature of variability in data (e.g., measurement variability, natural variability, induced variability, and sampling variability) or • anticipating and identifying potential sources of variability	DCP2: Describing and measuring variability in data for contextual variables as part of exploratory data analysis by • creating, using, interpreting, or fluently moving among various data representations to highlight patterns in variability; • focusing on aggregate or holistic features of data to describe variability in data; or • calculating, using, or interpreting appropriate summary measures for variability in data (e.g., measures of variation such as range, interquartile range, standard deviation for univariate data sets; correlation and coefficient of determination for bivariate data sets)	MP2: Identifying the pattern of variability in data or the expected pattern of variability for contextual variables by • modeling data to explain variability in data or • considering contextual variables in the formulation of appropriate data models or in • modeling data to describe holistic features of data or • considering or creating distribution-free models to explore contextual variables

Variability and relationships among data and variables	DP3: Controlling variability when designing studies or critiquing the extent to which variability was controlled in studies by • using random assignment or random selection of experimental/observational units to (in theory) equally distribute the effects of uncontrollable or unidentified sources of variability or • using study design to control the effects of extraneous variables (e.g., by incorporating blocking in experimental design or stratifying in sampling designs) to isolate the characteristics of the variable(s) of interest or to isolate systematic variation from random variation	DCP3: Exploring controlled and random variability to infer relationships among data and variables by • using and interpreting patterns of variability in various representations of data; • focusing on aggregate or holistic features of variability in data to make comparisons; • using or interpreting appropriate summary measures of the variability in data to make comparisons (e.g., transformed versus untransformed data); or • examining the variability within and among groups	MP3: Modeling controlled or random variability in data, transformed data, or sample statistics for • making inferences from data (e.g., isolating the signal from the noise for univariate or bivariate sets of data or formally testing for homogeneity in variances) or • assessing the goodness of a model's fit by examining deviations from the model
Effects of sample size on variability	DP4: Anticipating the effects of sample size on the variability of • a sample or • statistics used to characterize a sample (e.g., mean, proportion, median) when designing a study or critiquing a study design	DCP4: Examining the effects of sample size on the variability of • a sample or • statistics used to characterize a sample (e.g., mean, proportion, median) through the creation, use, or interpretation of data-based graphical or numerical representations	MP4: Anticipating the effects of sample size on the variability of a sampling distribution to • model the sampling distribution or • consider the practical and statistical significance of inferences

Data Collection to Address Research Question Two

To investigate the second research question, I incorporated data gathering methods that allowed me to examine teachers' perceptions and recollections of variation-related activities and actions. The best way to determine an individual's perceptions of experiences to then extrapolate the nature of the experiences is to engage in dialogue with the person to obtain his or her first-person accounts of the experiences (Moustakas, 1994).

Self-Report Methods

Phenomenological studies often include the use of semi-structured, in-depth interviews to elicit individuals' feelings about and experiences with the phenomenon under study (Seidman, 2006). Because participants in these studies have already experienced the phenomenon, participants are able to provide a retrospective recall of their experiences and feelings through self report. Because of the reliance on participants' memories and the accuracy of those memories, the collection and analysis of retrospective data brings issues of reliability and validity into question (Martyn & Belli, 2002). However, to study a phenomenon as it is happening presumes that one can create conditions under which the phenomenon will occur and that subjects for whom these conditions would bring about the intended phenomenal experience can be selected. Even if the experience could be provoked, research techniques for collecting and analyzing data for such a phenomena would be extremely time consuming and costly (Freedman, Thornton, Camburn, Alwin, & Young-DeMarco, 1988). Additionally, since it is not possible to observe an individual's feelings about experiences, self-report interviews would still need to be conducted retrospectively to obtain affective information about the experiences (Cuddapah, 2005; Loftus, 2000). Given the paucity of research that has been conducted to study advanced knowers' development of understandings of variation, I chose to use self-report methods to study the nature of experiences that provoked teachers' constructions of robust understandings of variation by examining the experiences of individuals who already have robust understandings. Concerns about the conclusions that can be drawn from this type of study will be addressed in the next three sections.

Recall Effects

Research work using self-report data has uncovered several characteristics that may inhibit individuals' abilities to recall or report events accurately. As noted by Ross (1989) in his review of the self-report literature, the tendency of an individual to present a favorable image to an interviewer brings into question the accuracy of descriptions. However, there are people who would argue that an individual's motivation to misreport experiences should be considered before discounting the accuracy of self-report data (e.g., Baldwin, 2000; Loftus, 2000). For example, it is posited that little motivation for misreporting exists when individuals participate in interviews where anonymity is preserved through the use of pseudonyms (Baldwin, 2000). Specific to this study, teachers were informed that pseudonyms would be used in reports from this study. Additionally, the sixteen teachers in this study seemed to be motivated to collaborate with me to provide information about how teacher education in statistics might be improved.

Research reports and syntheses of research literature suggest other characteristics that may affect the accuracy of reported events. These characteristics include the recency of events to be recalled (Eisenhower, Mathiowetz, & Morganstein, 1991; Tourangeau, 2000; Van der Vaart, Van der Zouwen, & Dijkstra, 1995), the saliency of the effects of events from the perspective of the individual (Brewer, 1994; Eisenhower, Mathiowetz, & Morganstein, 1991; Van der Vaart, Van der Zouwen, & Dijkstra, 1995), the number of times events have occurred (Brewer, 1994;

Van der Vaart, Van der Zouwen, & Dijkstra, 1995), differences in experiences among all event occurrences (Eisenhower, Mathiowetz, & Morganstein, 1991; Tourangeau, 2000; Van der Vaart, Van der Zouwen, & Dijkstra, 1995), and the affective state of the individual when the events occur (Brewer, 1994; Eisenhower, Mathiowetz, & Morganstein, 1991; Tourangeau, 2000). Strategies that can reduce the impact of these characteristics include the use of event history calendars and critical incidents.

Event History Calendars

As noted by Tourangeau (2000) in his review of the research literature on the self report of autobiographical data, “no single variable seems to have such a profound impact on the accessibility of a memory than its age” (p. 36). Individuals are less likely to recall events, particularly dated events, when retrieval cues are not incorporated in study design (Tourangeau, 2000). Research suggests that event history calendars (EHCs), sometimes called life history calendars, provide a means for individuals to accurately and completely reconstruct past events and experiences through the use of cues for significant past events (Martyn & Belli, 2002). The format of the EHC is a matrix, with columns containing timing cues for recording behaviors and rows containing behaviors—significant activities or events related to the goals of the research—that can help individuals to frame the occurrence of important events (Freedman, Thornton, Camburn, Alwin, & Young-DeMarco, 1988). Event history calendars have been shown to provide significant agreement about the timing of events when compared with survey results acquired one year, five years, and eighteen years earlier (Freedman, Thornton, Camburn, Alwin, & Young-DeMarco, 1988; Martyn & Belli, 2002). Research suggests that an orderly review of events enables greater recall by participants than when participants are asked to recall events in a haphazard fashion (Eisenhower, Mathiowetz, & Morganstein, 1991). Additionally, landmark events can be included in the calendar to aid participants in recalling events that occur in close temporal proximity to landmark events (Eisenhower, Mathiowetz, & Morganstein, 1991; Tourangeau, 2000). Event history calendars can help to avoid recall concerns related to the recency of an event and the number of times the event occurred.

In the context of learning statistics, the calendar included landmark events of the initial release year for commonly used statistics education resources, including textbooks and technology items, as these events were perceived to be events that might enhance teachers' recall ability (Means, Swan, Jobe, & Esposito, 1991). Additional events listed on the calendar included the landmark events of the first year of the AP Statistics examination, the location of the AP Reading for each year, and the context for the most talked about and notable free response questions for examinations in each year, as determined from archived discussions on an electronic discussion group monitored by many statistics teachers (Collegeboard.com Inc., 2007b). A partial sample of an EHC is shown in Appendix B; teachers were asked to complete a similar calendar in electronic form to provide the details of their personal histories and to expedite the process of returning the completed calendars to me. (I provided teachers with a password-protected, electronic template of the calendar that could be returned to me via e-mail.) The events displayed in the sample are events from my life that align with my personal transformative learning experience. In completing this portion of the EHC, the year of the first AP Statistics Reading served as my reference point for much of the professional development training I attended prior to that Reading as well as the professional development workshops I conducted subsequent to that Reading.

I asked teachers to record their personal information on the calendar, including the times in which they were taking statistics classes and the time(s) when they had obtained employment in education. A sample of the template that teachers used can be found in Appendix C. While the EHC shown in Appendix C contains no active links, teachers were able to use links to navigate through the document when completing their EHCs. I also asked teachers to record events and experiences related to their statistics education—events that are either already listed on the calendar or events that needed to be added in the “other” category. The calendar has events related to AP Statistics already listed on the calendar, so teachers only needed to record the timing of these experiences. The precise timing of key events in teachers’ experiences is less important than teachers’ ability to recall information about each experience. For each experience, I asked teachers to record details of the event and people, places, and feelings associated with the experience in prose form (Cuddapah, 2005). I also provided the sample event history shown in Appendix B to serve as a resource to aid teachers in completing their individual event histories. Teachers were encouraged to contact me if any questions arose during completion of the EHCs. Of all of the events recorded by teachers, some were more salient than others. Teachers elaborated further on their salient events, or critical incidents, to enhance their recall of those events and to explore affective dimensions of their experiences.

Critical Incidents

Research suggests that individuals tend to remember well unique events that evoke emotion at the time of occurrence or events that mark a transition point in their lives (Eisenhower, Mathiowetz, & Morganstein, 1991). Critical incidents are defined to be these unique events that are significant in the lives of individuals (Brookfield, 1990). There exists a long history for the use of critical incidents in research (Butterfield, Borgen, Amundson, & Maglio, 2005; Flanagan, 1954), with recent usage in retrospective self-report accounts of incidents that include aspects of the thoughts, feelings, and reasons behind each individual’s actions taken in response to the incidents (Butterfield, Borgen, Amundson, & Maglio, 2005). Critical incidents provide participants with an opportunity “to highlight particular, concrete, and contextually specific aspects of people’s experiences” (Brookfield, 1990, p. 180) because participants choose which incidents to discuss. Recent research uses critical incidents as a window for inferring people’s assumptions and beliefs (Butterfield, Borgen, Amundson, & Maglio, 2005), under the rationale that participants’ assumptions are likely to be implicit within their descriptions of the incidents (Brookfield, 1990). While individuals may struggle to articulate the underlying assumptions and beliefs that guide their actions, critical incidents provide a means for researchers to infer participants’ assumptions and beliefs from the experiences and actions of the incidents recorded by participants (Brookfield, 1990).

In his use of critical incidents with adult learners, Brookfield (1990) gave instructions for the types of experiences to describe, and he encouraged individuals to write brief descriptions of the critical events in their lives, detailing the time, place, and other people involved. He also asked individuals to write explanations for their selection of critical events. To provide an even fuller picture of a person’s assumptions, researchers (e.g., Cuddapah, 2005) have asked participants to focus on one successful, or positive, critical incident and one failure, or negative, critical incident. Some researchers suggest that participants should also include the actions they took in response to the critical incident, the thoughts and feelings they had about the event, and their actions in response to the event (Kennedy & Wyrick, 1995). In their study that incorporated the use of critical incidents with teachers, Kelchtermans and Vandenberghe (1994) found that “teachers will mention these moments as important for their professional development” and that “as a result of some critical incidents the teacher has to change a habitual approach to cope with new challenges” (p. 48), suggesting that critical incidents may be triggers for significant, transformational learning experiences for teachers. Kelchtermans and Vandenberghe acknowledge that while certain events may be critical incidents for some teachers, resulting in a change in professional behavior, the same events may not be critical events for others. Because teachers may experience professional growth as a slow and gradual process, they may struggle to find a single, salient event related to a particular aspect of their professions. In such cases, however, the teachers can still identify and describe in great detail experiences that influenced their professional development (Kelchtermans & Vandenberghe, 1994).

I asked teachers to recount two critical incidents related to their study of variation or statistics—one positive experience related to their informal or formal study and one negative experience. Limiting the number of critical incidents focused teachers on their most salient experiences; other experiences were included in the EHC and discussed in subsequent interviews. The two critical incidents were used to identify potential disorienting dilemmas. Teachers were asked to write brief descriptions of these critical events, detailing the time, place, and other persons involved in the events. Research suggests that participants who can provide detailed information about their critical incidents provide valid descriptions of those experiences (e.g., Eisenhower, Mathiowetz, & Morganstein, 1991). Additionally, asking details about the timing, location, and individuals involved in the incident helped to place the critical incidents temporally in teachers' event histories. I also asked teachers to provide explanations for selecting the critical events they chose to write about and to describe the actions they took subsequent to the events. The information provided by the teachers yielded insight into the existence of disorienting dilemmas as well as details about the characteristics of events that created the dilemmas. Lastly, I asked teachers to express the thoughts and feelings they had about the event and their feelings about the actions they took in response to the event. As discussed in the section on transformation theory, disorienting dilemmas can precipitate self-examination, which may be accompanied by strong emotion. Thus, the descriptions teachers provided about their critical incidents provided me with valuable information about events or experiences that may provoke disorienting dilemmas for statistics teachers as well as information about actions teachers might take to resolve their dilemmas. The critical incidents questions that teachers were asked to consider are contained in Appendix D. I gained further insight into these events by interviewing teachers about the finer details of these events.

Teacher Interviews

Three interviews are common in phenomenological research (Seidman, 2006). In a phenomenological study, the main goal is to have participants reconstruct experiences related to the topic of the research. A series of three interviews gives participants an opportunity to reflect on the events and to provide detailed accounts of the events, as evidenced by the rich descriptions resulting from those who have employed the use of three interviews (Seidman, 2006). Seidman (2006) suggests that the first interview should "establish the context of the participants' experience" (p. 17), while others use the first interview to also establish rapport with study participants (Cuddapah, 2005). This study required an initial interview to establish that teachers experienced the phenomenon of developing robust understandings of variation; I worked towards establishing rapport with each teacher during our introductory conversations and during the content interview. Further, the teachers in this study established the context of their experiences by completing event history calendars and critical incident descriptions outside of my presence. They electronically completed and returned these documents to me. It was therefore possible to collect sufficient retrospective data for analysis with two context interviews for each teacher. Although EHCs are typically completed during the course of an interview (Freedman, Thornton, Camburn, Alwin, & Young-DeMarco, 1988; Martyn & Belli, 2002), research exists to suggest that successful retrieval of event details takes considerable time (Schwarz, Hippler, & Noelle-Neumann, 1994; Tourangeau, 2000)—time that individuals would not necessarily have if they completed the EHCs as part of an interview. Seidman (2006) acknowledges that deviations from his recommended course of three interviews can occur "as long as the overall structure is maintained" (p. 21). I provided teachers with clear directions for completing the EHCs and for describing their critical incidents and I made myself available for any questions that arose in completing the documentation, thus maintaining the structure of data collection for this phenomenological study. The pre-pilot and pilot studies were used to establish whether the directions for completing the documents were clear, to establish the viability of teachers completing the documents prior to the first context interview, and to establish whether the questions contained in the interview schedule provided evidence of individuals' perceptions of characteristics of experiences that enhanced their learning related to variation using transformation theory as the lens through which to view the experiences.

I conducted a face-to-face interview with each teacher to reconstruct the finer details of the experiences he or she listed in the completed EHC and described in the critical incidents reports (Seidman, 2006). The content interview and this first context interview were conducted on consecutive days, with the content interview occurring first. In most cases, teachers had returned their completed documents to me well in advance of the context interview, and in all cases, I had a chance to review the documents before the first context interview took place. I perused the documents to gain a sense of the temporal positioning of educational experiences, to become familiar with experiences listed as pivotal or influential, and to construct a preliminary set of questions unique to each individual, in consideration of transformation theory, and guided by the general questions contained in the interview schedule. An abbreviated form of the interview schedule for the first context interview appears in Appendix E. Individualized questions also included questions to connect context with content to gather additional evidence for any facets of variation that were not addressed thoroughly in the content interview and to clarify unclear statements written in the EHC and critical incident descriptions.

To illustrate the type of individualized questions I asked during the context interviews, I describe part of a context interview with Faith, a teacher who participated in the main study. Faith completed her first statistics course in her undergraduate program and disliked it. She wrote that she could not “use calculators and all I remember is the constant crunching of numbers. Thought statistics was incredibly boring and was glad never to take another course” (Faith, EHC). Questions I developed based on Faith’s comments were similar to the following.[12]

- You mention that you had only taken one statistics course, and you thought statistics was incredibly boring. What caused you to change your mind about statistics?
- How did you think about standard deviation upon completing this course?
- What, if any, value did you see in this course?
- How, if at all, did this course help you to learn the statistical content you needed to teach in AP Statistics?

These questions were developed in consideration of an apparent subsequent change in Faith’s beliefs about statistics and statistics teaching, which suggested a possible disorienting dilemma that may have caused Faith to reconsider the assumptions and beliefs she formed in response to this course.

Individualized questions were asked in addition to some of the general questions outlined in the interview schedule in conjunction with related ideas from other experiences. With Faith, I initially asked her to describe the experience she found to be most valuable for her learning of statistics. She responded, “I didn’t care to learn statistics at all until a teacher at my school asked me to teach AP Statistics” (Faith Context I, Lines 11-12). Her comment allowed for discussion about this positive learning experience, using the questions contained in the interview schedule as a guide, as well as the course that initially removed any motivation Faith had to learn statistics.

In general, I asked teachers to describe experiences that were valuable for their learning of statistics and variation, the statistics learned during the course of their experiences, their belief

---

[12] The questions written here differ slightly from those asked during the interview in order to remove personal information and to maintain the anonymity of the teacher.

about why the experiences benefitted their learning, their emotions associated with the experiences, influential people associated with the experiences, actions taken in response to the experiences, and how, if at all, the experiences changed the way they thought about statistics and variation. In all cases, teachers referred to their critical incident experiences in their responses to these questions. I used the information teachers provided in the critical incident descriptions similar to the way I used the information from their EHCs. Throughout the interview, teachers' descriptions moved from a general description of events to a more detailed accounting of the events (Tourangeau, 2000), and the interview provided details about memorable events and actions taken subsequent to the events (Peters, 1991; Seidman, 2006). Further details about why the events transpired in the way they did were reserved for a follow-up interview (Peters, 1991; Seidman, 2006), the Context II interview. An abbreviated interview schedule for this second context interview is contained in Appendix F.

Within several weeks of conducting the first context interview, I conducted this third and final interview with each teacher via telephone. The span of time between the two context interviews provided an opportunity for teachers to reflect on the reasons for their actions and on the meaning of their experiences—elements that comprised the focus of the third interview (Peters, 1991; Seidman, 2006). In the time between interviews, I asked teachers to record journal entries of their reflective thoughts related to their statistical experiences and events, and I provided them with a small tablet upon which to record their thoughts. During the same interval of time, I reviewed both the content and the context interviews for each teacher to determine any remaining questions I might have from the first two interviews.

I began the final interview by asking teachers to describe any experiences that needed to be added to their descriptions of events. I then asked teachers probing questions to explore what they perceived to be the meaning behind the events and actions they saw as valuable in their journeys towards understanding variation. During the interviews, I asked teachers questions about the experiences that were most valuable for their learning in the areas of exploratory data analysis, study design, and inferential statistics—areas that align well with the development of data-centric, design, and modeling perspectives. Questions focused on extracting features of the experiences that teachers found to be valuable and those found to be ineffective for their learning about variation and the reasons they attributed to the effectiveness of particular features. Because meaning making "requires that the participants look at how the factors in their lives interacted to bring them to their present situation" (Seidman, 2006, p. 18), I also asked teachers to describe how they believed their collective group of experiences contributed to their learning about variation and their reasons for that belief. Questions that guided the last interview appear in the interview schedule and align well with characteristics of transformative events using the lens of transformation theory. As with the content interviews, I transcribed both context interviews from each teacher and annotated the first context interview for my subsequent analysis of the data.

Additional Data Sources

To obtain additional information about teachers' experiences and understandings, I asked teachers to provide their most recent resume or curriculum vita if it was available in paper or electronic form. Ten of the 16 teachers in the study were able to fulfill this request. In addition to this documentation, teachers' recording of events and description of critical incidents, followed up in subsequent interviews, provided a form of triangulation for their retrospective accounts of experiences, actions, and events with as much accuracy as possible. The data collection and analysis schedule I followed for the pre-pilot, pilot, and main phenomenological studies is displayed in Table 4-4.

Table **4-4**: Data Sources and Data Collection.

Data Source/ Background Work	Timing	Purpose	Data Format
Pre-pilot EHCs and critical incident (CI) descriptions with 3 teachers	May-June 2007	Establish the viability of • the directions and categories of information on the EHC to determine usefulness of the document and • the directions and information about participants' critical incidents to determine usefulness of the document	Electronic files
Pre-pilot 1.5-hour Content interview	May-June 2007	Establish the viability of the instrument for ascertaining robust understandings of variation	Video and audio recordings
Pre-pilot 1.5-hour Context I interview	June-July 2007	Establish the viability of the instrument for ascertaining characteristics of actions and activities that led to an understanding of variation in conjunction with the EHCs and critical incident descriptions	Video and audio recordings
Pre-pilot 1.5 hour Context II interview	July 2007	Establish the viability of the instrument for ascertaining characteristics of actions and activities that led to an understanding of variation in conjunction with the first context interview	Audio recordings
Pilot study	July-August 2007	Fine-tune instruments and probing questions with the goal of establishing that • tasks and questions from content interview can elicit reasoning about variation from the data-centric, design, and modeling perspectives • the data for research question one can be analyzed using the SOLO Model to frame data analysis • documentation and context interviews can elicit detailed descriptions of important learning experiences related to variation • the data for research question two can be analyzed using the lens of transformation theory	Electronic files, video and audio recordings
Participant selection	July-August 2007	Establish initial pool of potential participants Initiate e-mail contact with potential participants Contact volunteers and e-mail questionnaires Select participants from questionnaires returned E-mail participant files for EHC and CIs	Electronic files

		Establish dates for content interview and first context interview with teachers	
1.5-hour Content interview	August-December 2007	Establish teachers' understanding of variation: • from the data-centric perspective • from the modeling perspective • from the design perspective • from the integration of these three perspectives	Video and audio recordings, transcribed, and annotated
Preliminary analysis of Content interviews	August-December 2007	• Establish that evidence exists to examine teachers' reasoning about variation from, data-centric, modeling, and design perspectives • Determine content questions to ask during context interview in areas for which content evidence is weak or missing	
EHCs, descriptions of CIs, and preliminary analysis of documents	August 2007-January 2008	Establish potential • disorienting dilemmas • opportunities for rational discourse • significant events that contributed to understanding • actions taken in response to critical reflection Establish teacher-specific questions for first context interview	Electronic files of EHC and CI
1.5-hour Context I interview	August-December 2007	Establish teachers' view of • the significance of events listed on their EHC • characteristics of events from which participants learned • organizational support, including resources, for events listed on the EHC • interactions with others that supported learning • reasons for their selection of critical incidents • experiencing Mezirow's different elements of transformation	Video and audio recordings, transcribed and annotated
Preliminary analysis of Context I interview	September 2007-January 2008	Establish participant-specific questions for second context interview	
1.5-hour Context II interview	September 2007-January 2008	Establish teachers' view of • improper characterization of events • the meaning of any disorienting dilemmas • the meaning of events identified by them as contributing to their understanding of variation	Audio recorded and transcribed

Data Analysis to Address Research Question Two

My analysis of teachers' documents and interviews to address the second research question, characterizing statistics teachers' perceptions and recollections of activities and actions that contributed to their understanding of variation, followed the detailed, systemic procedures recommended for analysis in phenomenological studies (Moustakas, 1994). Throughout the process, I attempted to put aside my preconceived notions about how individuals might come to understand variation, to the extent possible, so that I could envision the experiences of participants without bias, a process called bracketing (Moustakas, 1994; Stanage, 1987). One recommended bracketing method suggests that repeated reflection allows me to disconnect from my experiences (Moerer-Urdahl & Creswell, 2004) so that my experiences are no more or no less important than the experiences of others. Dialoguing with other statisticians and mathematics educators, including members of my thesis committee, provided further means to accomplish bracketing, as did conducting and analyzing the context interviews from the pre-pilot and pilot studies.

Pre-Pilot and Pilot Study Analysis

In addition to piloting the content interview process, my pre-pilot and pilot studies also focused on the viability of the instruments and context interview schedules in consideration of research question two. The same teachers agreed to participate in this second part of the pilot studies. The main purpose of the pre-pilot and pilot studies was to ensure that directions for completing the EHC and CI descriptions were clear and could elicit sufficient information to allow in-depth conversation about learning experiences and allow the second research question to be answered.

The pre-pilot and pilot Context I interviews were videorecorded, transcribed, and annotated prior to analysis, and the Context II interviews were audiorecorded and transcribed prior to analysis. Analysis consisted of matching passages of individuals' descriptions of perceived learning experiences to the elements of transformative learning listed in Table 4-5.

Table **4-5**: Elements of Transformative Learning.

	Elements of Perspective Transformation
Critical Reflection	Disorienting dilemma or sequence of transformed meaning schemes
	Self-examination, accompanied by emotions
	Critical assessment of assumptions related to epistemic, sociolinguistic, or psychological perspectives
Rational Discourse	Recognition that others have experienced similar discontent with their perspectives
	Exploring new roles, relationships and actions through engaging in rational discourse with others—learning in the communicative domain
Action	Planning a course of action
	Constructing the knowledge and skills needed to enact the plan—learning in the instrumental domain and possibly in the communicative domain
	Experimenting with new roles
	Building a sense of competence and self-confidence for new roles and relationships
	Reintegration into life based on the transformed perspective

From the identified elements of transformative learning for each pilot-study teacher who exhibited robust understandings, I wrote a summary of characteristics of experiences that they perceived to be valuable for their learning. The intent was to describe a coherent and cohesive picture of the teacher's learning. Interview transcripts and summaries for some teachers were discussed with another mathematics education researcher until agreement was reached on the viability of the instruments for providing sufficient evidence to respond to research question two.

In-Depth Analysis of Data in Response to Research Question Two

Subsequent to the completion of data collection, I examined teachers' context interviews and documents to find statements that provided information about experiences related to the development of a robust understanding of variation for those teachers identified as exhibiting reasoning consistent with robust understandings. This was accomplished by recording experiences the teachers identified as important for their development of understandings of variation, as well as their perceptions of characteristics that helped or hindered their development (Cuddapah, 2005). I sought evidence of elements related to transformative learning as well as evidence potentially refuting transformative learning. In particular, I sought evidence of events that triggered dilemmas, critical reflection, rational discourse with self or others, seeking additional knowledge related to statistics, experimenting with new roles, and changes in beliefs and assumptions related to the teaching and learning of statistics. Interview passages for each teacher were grouped in a table according to the elements of perspective transformation (Moerer-Urdahl & Creswell, 2004; Moustakas, 1994), and a separate table was created for learning experiences that occurred prior to and learning experiences that occurred subsequent to triggers of transformational experiences.

During the next phase of analysis, I organized statements from context interviews into themes and grouped them to produce a textual, or factual, description of the phenomenon for each teacher (Creswell, 1998; Moustakas, 1994) in relation to elements of transformative learning. For each element, I extracted the essence of the phenomenological experience by viewing the phenomenological descriptions from divergent perspectives, including some input from another mathematics educator, and continually reread the descriptions to gain further insight into teachers' experiences (Moustakas, 1994). Whereas textual descriptions of teachers' phenomena describe *what* their experiences were, the descriptions that result from this phase of analysis describe *how* teachers experienced the phenomena. The resulting description, a structural description that explicates the essential structures of the phenomenon, was recorded for the experiences of each teacher and for the larger group of teachers (Moerer-Urdahl & Creswell, 2004; Moustakas, 1994).

The final stage of analysis involved integrating the textual descriptions of teachers' experiences into a composite textural description. To do so, I combined individual teachers' evidence related to each element of transformational learning into a larger table that contained evidence of each element for the larger group of teachers. For example, Table 4-6 contains a list of disorienting triggers for the group of teachers. Detailed descriptions of the types of triggers teacher experienced appear in Chapter 7.

Table **4-6**: Disorienting Triggers for Learning.

Triggers
Whys behind hows and making connections between concrete and abstract and among concepts
Listening to statisticians' "arguments"
Awareness of questions and knowledge limitations from various sources: (a) Conversations with colleagues and/or statisticians (b) Listening to conversations at AP Reading (c) Solving AP free-response questions
Design and recognition of differences between mathematics and statistics
Language discrepancies
Subtleties
Limitations of only introductory-level understanding of statistics
Teaching AP Statistics; AP requirements
Participating in activities such as those used with students in introductory courses

Identification of commonalities was followed by integrating the structural descriptions of teachers' experiences into a composite structural description (Moustakas, 1994) to form a synthesis of these composite descriptions that characterize the overall essence of the experience of developing robust understandings of variation (Moerer-Urdahl & Creswell, 2004). This description appears at the end of Chapter 7.

Concluding Comments

As the preceding sections suggest, the 16 teachers who participated in this study formed a purposeful sample of secondary statistics teacher–leaders. To answer the first research question of what conceptions of statistical variation they exhibit, data collection and analysis used a protocol for conducting task-based interviews framed by the SOLO Model and for analyzing interview and artifact data with the constant comparative method. Chapter 5 details the answer that followed from implementing the processes outlined in this chapter. Through data analysis to answer the question of teachers' conceptions of statistical variation, an empirically derived framework consisting of indicators of robust understandings for statistical variation emerged. That framework is described in detail in Chapter 6. Learning experience data collected from the five teachers who exhibited reasoning consistent with robust understandings of variation were collected and analyzed following protocol for phenomenological studies. Chapter 7 details influential factors for learning that emerged from analysis in answer the second research question: "For those secondary AP Statistics leaders who exhibit robust understandings of variation, what are the activities and actions that contributed to their current understandings of variation as reflected in their perceptions and recollections of experiences?"

Chapter 5

Conceptions of Variation

Three types of conceptions of statistical variation emerged from analysis of the content and context interviews with 16 teacher-leaders: Expected but Explainable and Controllable (EEC), Noise in Signal and Noise (NSN), and Expectation and Deviation from Expectation (EDE). Individuals with EEC conceptions see variation as something that needs to be controlled and explained and hence tend to focus their attention on issues of design. In contrast, individuals who harbor NSN conceptions see variation as something that needs to be explored, which manifests in strong consideration of variation during exploratory data analysis. Lastly, individuals who conceive of variation as EDE see variation as something that can be expected and modeled, and their reasoning about variation is typified by a focus on models, and in particular, models related to inference. As their different statistical foci of design, exploratory data analysis, and inference might suggest, individuals with different conceptions view variation from predominantly different perspectives. Specifically, the design, data-centric, and modeling perspectives are prevalent for individuals with EEC, NSN, and EDE conceptions, respectively. Discussion of the conceptions is organized around the design, data-centric, and modeling perspectives to allow focus on the most salient characteristics of each conception while providing a means to make comparisons across the conceptions. This chapter describes each conception in detail, associates teachers in the study with each conception and provides examples from interviews to support the existence and nature of each conception. The chapter concludes with a comparison of the key similarities and differences of the conceptions.

Conception: Expected but Explainable and Controllable (EEC)

Individuals with EEC conceptions of variation see variation as not only omnipresent and unavoidable but also as explainable and controllable. Their sense of the omnipresence and unavoidability of variation leads them to expect variation in statistical settings, and their view of variation as controllable focuses on design strategies for both observational and experimental studies. Their view of variation as explainable aligns with their focus on context to identify factors that potentially contribute to variability in data and their attraction to designs that allow them to determine cause-and-effect relationships.

Individuals' search for causes or explanations is a phenomenon witnessed by Reading and Shaughnessy (2004). They noticed that primary and secondary school students in their study provided explanations and causal reasons in their responses to sampling tasks even though students were not asked to consider causes. Their study suggests that a search for explanations may be fairly typical of children, intimating that EEC conceptions may emerge early and potentially develop into conceptions similar to that of Isaac and Haley, the two teachers in this study who clearly viewed variation as EEC.

Throughout discussion of the EEC conception, I draw on examples from Haley's and Isaac's Content interviews to illustrate facets of their conceptions. Haley's and Isaac's general reactions to the interview tasks are strikingly similar but differ in terms of details. For example, they both consider multiple sources of variation when they design studies for the Handwriting Task, but the sources differ. Isaac mentions the quality of writing and scorer training as potential sources of variation, and Haley cites the reading level and subject matter of essays as potential contributors to variation. When Haley and Isaac reason similarly in response to a task, I describe the clearer and more succinct example. On several occasions, I note that neither Haley nor Isaac was prompted to address a particular aspect of design or analysis. My intention is not to imply that they were prompted in other cases but to draw a comparison with a number of other teachers who addressed the same issue only after they were asked to do so. Instances in which any teacher received prompting will be noted as such.

EEC and the Design Perspective

Isaac's and Haley's view of variation as EEC seems to lie at the heart of their privileging of the design perspective and their desire to know as much as possible about context. Statisticians note that obtaining background knowledge about context is a critical component of statistical problem solving (Pfannkuch, 1997; Pfannkuch & Wild, 2000). Whereas some preliminary research suggests that some teachers view considering sources of variation as one of the primary statistical areas in which they need support for teaching (Arnold, 2008), Isaac and Haley are extremely adept at using contextual information to consider a variety of variables that may confound their results. Their EEC conceptions also may be the source of their criticisms of studies whose designers do not anticipate the general presence of variation, do not consider potential sources of variation, or do not design studies with explanation and control in mind.

Expected

Individuals with EEC conceptions attend to context through anticipation. Their expectation of variation is most noticeable when they design studies or examine studies conducted by others. For example, when Haley first reacts to the Handwriting Task, she suggests designing a study that uses essays written on a reading level known to be understandable by most adults.

> Well, I would take, um, I would take something typed from *USA Today* because I know that's a sixth grade reading level. And I know most adults could

> comprehend a sixth grade reading level. And I would make sure that the adults at least could read a sixth grade reading level. (Haley, Content, Lines 1916-1924)

Haley's inclination to "make sure" adults can read the essays suggests that she anticipates reading levels could contribute variation to the resulting scores if left unchecked. Haley's expectation of variation presumably leads her to devise a strategy that may reduce variation from reading levels—that of selecting readings on a sixth grade level—resulting in a treatment that allows her to study the variable of interest while reducing the potential for large errors from scorers' reading levels. As someone who expects variation, Haley focuses on identifying potential sources of variation for a given context and ways to reduce the effects of the sources she identifies.

For individuals with EEC conceptions, their expectations for variation couple with their affinities for explanation and for control. They use context to identify potential sources of variation and use knowledge related to those sources to select and implement design strategies that allow them to explain as much variation in data as possible and to control variation. The effects of Haley's and Isaac's expectations of variation will be considered further in concert with their reasoning about variation in terms of explanation and control.

Explainable

Someone with a view of variation as explainable relies heavily on context to consider factors that may explain variability in data for the variable or variables of interest in a statistical study. That person attends to context before conducting any type of formal or informal data analysis. Consider Isaac's initial reaction to the Caliper Task. When he sees the scatterplot with axes labeled x and y, he does not outwardly attend to the pattern of variability in the seven plotted points. Instead, he seeks information about context to consider whether there is a known relationship or pattern between contextualized variables in the student's science lab. He notes, "the student asks you how they might use this graph to predict a value for *y*… Whew. Well, did – I'm – I think my first question of one of my students would be, um, the nature of the data" (Isaac Content, Lines 1298-1302). The intent of Isaac's request for the "nature of the data" may not be clear at first. He later offers situations for which he knows the "nature of the data", and he describes how that expectation affects the way he looks at data and the models he considers to fit the data. One example is found in his descriptions of a relationship between test anxiety and student achievement.

> Uh, if I knew what the data were, I'd want a model that was consistent with the current scientific understanding of the relations between the two variables…Um, if this were say, uh student achievement as a function of background anxiety – test anxiety…it makes sense that, um, a student has to have a certain amount of anxiety just to even take the test, but after, after some point that anxiety becomes debilitating and so hmm, in that case maybe a quadratic model makes sense. (Isaac, Content, Lines 1337-1370)

Isaac also offers a setting from which he might expect the data to exhibit an exponential pattern. For that setting, he suggests that the two rightmost points would have to be "errors." Considering Isaac's initial comments along with his later comments, it seems that Isaac wants to know more about context to gain a sense of expectation for a pattern of variability and to explain variation in the response variable based on a known theoretical relationship between variables. He offers examples of different relationships and the corresponding models for those relationships. He uses the models to explain variability in values for the response variable, thereby reducing the amount of unexplained variability. For Isaac and others with EEC conceptions, considerations of contextual factors are necessary precursors to using statistical methods of analysis.

Individuals with EEC conceptions seem to attribute value to patterns and relationships in data only if they are plausible within the context of the data—deliberations that are characteristic of statistical thinking (Cobb & Moore, 1997). Results that stem solely from the application of statistical procedures seem to have little meaning for them. Indicative of this stance is Isaac's reaction to the Caliper Task. In addition to desiring contextual information to consider possible data patterns, Isaac provides evidence that he would like contextual information when asked to make a meaningful prediction for *y* when *x* is four.

> Well, what you might do is assume an underlying linear relationship. Get a regression line and predict that way … What you might do is assume an underlying quadratic relation … And so, it seems to me, what I would – if their question is how might you use this graph, well here's a bunch of alternatives. If the question is how might you validly do this, then I think I'd want a little bit more information than what's here. (Isaac, Content, Lines 1311-1326)

Isaac does not make a prediction based strictly on a model that "best" fits the data; he seems to prefer combining contextual considerations with statistical procedures to help the student make a prediction. Although his language suggests that he can model the data to explain variation in the response variable, Isaac suggests that using context during model selection allows conclusions to be "valid." Individuals with EEC conceptions espouse Moore and Cobb's observation that "context provides meaning" in data analysis (Moore & Cobb, 2000, p. 615).

Context also allows for considerations of variation beyond association or theoretical relationships among variables towards a search for explanations or causes behind unusual variation in data. Haley suggests that a search for causes is normal, noting, "I think that's just part of human nature… is naturally to question variation. To think for a reason" (Haley, Content, Lines 2127-2129). Both Haley and Isaac seem to seek causal relationships even when analyzing data from non-experimental studies. For example, after Isaac is given centimeters and inches as the names of the variables for *x* and *y*, respectively, in the Caliper Task, he seeks to explain why the two rightmost points vary from the theoretically increasing and linear relationship between centimeters and inches. He notes the following.

> My suspicion would be that somehow or other when you get to that level of measurement, uh, maybe they've gone beyond what the caliper was designed for. Maybe the student's hand is too small or something, uh, makes it perhaps problematic to accurately measure above two point five. (Isaac, Content, Lines 1479-1484)

Isaac's explanation resides in the tool used to measure objects of varying lengths. Haley attributes the error in the rightmost point not to the physical apparatus as Isaac does but to student error, immediately reacting to the variable names with "somebody made a little flaw here. [*Haley points to the rightmost point*.] I would think" (Haley, Content, Lines 1686-1687). Both teachers identify contextual factors that could feasibly contribute variation to the measurements displayed in the scatterplot, and both suggest a causal relationship for their identified factors. Moreover, they each present a conjecture to explain the deviation of the rightmost points.

A view of variation as explainable also may be at the heart of privileging experiments over observational studies. A fundamental advantage of experiments is the capacity to establish cause-and-effect relationships, which provides a stronger explanation than is possible from association alone. Haley, in particular, has a strong affinity for experimental design. Although the Consultant Task presents data from an observational study, Haley seems to be dissatisfied with the limited conclusions that she can draw. After she reads the task statement, she observes, "I don't understand, if you do a difference of two means, what's that going to prove?" (Haley, Content, Lines 44-46). Haley seems to expect the administrators to want more information than a comparison of means will allow—she may be looking for a potential cause for improved scores or alternatively for a potential cause for changes in scores. Specifically, she notes the administrators' stated goal of improved scores, and she suggests that their design will yield little information towards achieving their goal.

> They want—what—what is their goal…they're trying to get to improve students' test scores on the state assessment…The consultants' contract—see I'm not quite sure how that's going to improve, how that's going to show improvement. There's no treatment there. (Haley, Content, Lines 58-69)

Haley notes that the administrators have not designed an experiment—no treatment exists for determining how to improve scores. Comparing average scores for consultants seems to make little sense to Haley if the ultimate goal is to improve scores. She seems to struggle with the administrators' use of an observational study that from the task description appears to provide no explanatory power for how to improve scores. She spends a considerable amount of time questioning their methods before she addresses the question posed in the Consultant Task statement. Even after she focuses on analyzing the collected data, she remains critical of the design. Haley's analysis of the methods employed by others and consideration of alternative designs that achieve greater explanatory power are hallmark characteristics of those with EEC conceptions.

The search for both causal and associational relationships and explanations is not unique to those with EEC conceptions of variation. Some statisticians even suggest that statistics education should focus on the search for causes and emphasize how statistics can aid in the pursuit of causes (Wild & Pfannkuch, 1999). What seems to separate Haley and Isaac as teachers with EEC conceptions of variation from other teachers in this study is the extent to which they search for explanations, particularly causal explanations, and the extent to which they focus on context and issues of design. Closely tied to their selection of design strategies is their utilization of methods that control variation in variables of interest.

Controllable

A second characteristic theme in the statistical reasoning of teachers with EEC conceptions of variation is control of variation through design. For example, Haley and Isaac seek to control variation (or to evaluate the extent to which designers of studies with published results controlled variation) and to seek potential explanations for uncontrolled variation. They design experiments to determine the significance of induced, systematic variation in response values in comparison with naturally expected random variation. They use design strategies such as blocking to combat the effects of variables that are likely to contribute variation to the response variable(s) but are not particularly of interest to the study. Haley's (and Isaac's) design for the Handwriting Task incorporates blocking as one method to control variation. She suggests using essays with varied subject matter based on her expectation that subject matter makes a "difference" in scoring.

> Then I would take a, a story —I'd take a sports, I'd take something about fashion, something about news… Block design for sports, block design for fashion, block design…The block would, um, hopefully minimize the variation because I really think—I honestly think sports, fashion, and news—I really think there's a difference...So the variation between this, somebody might score fashion a lot higher than they do sports because they hate sports. (Haley, Content, Lines 1924-2056)

Based on her belief that adults may score essays according to personal interest, Haley suggests blocking by subject matter to control variation in scores assigned by the adults. Unlike many of the other teachers in this study (but not unlike Isaac), Haley suggests blocking without prompting from the interviewer. Haley's efforts to control variation from sources that might interfere with her ability to address the Handwriting Task are closely tied to her expectations. For her, expectation and control are intimately related. Her use of context to identify potential sources of variation that then can be controlled typifies the reasoning of individuals with EEC conceptions of variation.

Given EEC conceptions, strategies to control variation are not limited to experimental design. Haley and Isaac also recommend control strategies for observational studies, including the observational study analog to blocking in experimental design: stratified sampling. For example, when Isaac is asked how he would design the study described in the Consultant Task, he considers selecting a stratified sample in order to sample exams over the entire interval of scores from 0 to 15. He notes that one advantage of a stratified random sample over a simple random sample is precisely this dispersed effect. Using a stratified sample, he controls variation by imposing greater variation on each set of exams but (presumably) reduces variation overall by considering each stratum separately. Isaac even states that his goal is control: "If I could get a stratified sample, then I could in a sense control that [sample] distribution" (Isaac, Content, Lines 287-288). As these examples suggest, Haley and Isaac use design strategies for experiments and observational studies for the purposes of controlling variation. Their strategies for controlling variation align with the designs they suggest based on their expectations and their attempts to find explanations.

Haley's and Isaac's desires to control variation in data and to explain as much variation in data as possible, characteristic of EEC conceptions of variation, may elucidate their dismay with less-than-ideal designs employed by others. In particular, they look for designs that are appropriate for the research and statistical questions under consideration. For example, when Isaac first reads the Consultant Task, he believes that the administrators are "looking for the reliability of the, uh, the interrater reliability" (Isaac, Content, Line 32-33), which he seems to associate with correlation, between the consultants. He suggests there is a mismatch between the administrators' professed goals and their design when he says, "I don't see how they're going to get that from looking at two independent samples" (Isaac, Content, Lines 33-34). Isaac suggests that a more informative approach would be to examine consultants' scores for the exact same sample of exams. He proposes considering the strength of the association between the two consultants' scores, noting that he would be looking for "some indication of a strong association between the scores" (Isaac, Content, Lines 37-38). By focusing on interrater reliability, he in a sense controls the natural variation in scores that can be expected on assessments taken by multiple students and focuses his attention on the variation in scores due to consultants' scoring. Isaac's approach would provide information about *how* the scores differ rather than simply *if* they differ by calculating a measure that reveals the consistency of agreement between consultants.

When Haley is asked how she might use the data from the two samples, she proposes a design somewhat similar to the paired approach of Isaac. Because she reasons about scores from two independent samples, though, she does not suggest comparing the two sets of consultants' scores. Instead, she suggests comparing consultants' scores against some known standard by analyzing the same students' results from a previously administered standardized exam to determine whether distributional characteristics are similar between the consultants' scores and the standardized scores. Haley controls variation in the data she uses for comparison by removing the variation in scores that could be expected regardless of scorer (i.e., students' standardized scores). This allows focus to remain on the variation in scores contributed by consultants. Like Isaac, Haley suggests a design method that provides information beyond whether there was a difference in means or a mean difference for the consultants' scores. Characteristic of their EEC conceptions, they incorporate design strategies to control variation for greater explanatory power from data.

Although individuals with EEC conceptions show some creativity and ingenuity in implementing strategies to control variation, they also employ relatively common strategies of control, including strategies for randomization and sample size. For example, Isaac suggests incorporating random assignment into his design for the Handwriting Task. He notes that: "I'm really, um, bringing any difference between… these into a probabilistic model as, as opposed to depending on me making the judgment" (Isaac, Content, Lines 1892-1898). Isaac recognizes random assignment as a method to control variation by theoretically distributing variation from uncontrolled sources equally among treatment groups, clarity largely unseen in a study by Derry, Levin, Osana, Jones, and Peterson (2000). He mentions that without randomization, he might believe that he creates equal groups, but there might be some underlying cause that creates bias in the way he selects groups. Isaac also mentions the advantage of increased sample size in relation to the reduced variability in sampling distributions, as does Haley. Their recognition of the effects of sample size stands in contrast to a common misconception that sample size is irrelevant (Fischbein & Schnarch, 1997).

Summary of Design Perspective and EEC Conceptions

Individuals with EEC conceptions exhibit telltale signs of their conceptions in their reasoning, including explicit and thorough consideration of contextual factors that might be sources of variation—a key consideration that some would argue distinguishes statistics from mathematics (e.g., Cobb & Moore, 1997). They seek to collect data in ways that allow them to discover patterns and relationships in data, with a preference for establishing cause-and-effect relationships. They use their knowledge of context to implement design techniques that allow them to control and to explain variation to yield meaningful results. Finally, yet importantly, they tend to do each of these things naturally and without prompting. These signs of an EEC conception of variation align with factors seen as necessary for understanding variation: recognizing the omnipresence of variability, considering potential sources of variation and distinctions among the types of variation, explaining variation based on context and current knowledge of sources of variation, and considering unexplained variation (Pfannkuch, 1997). No single factor or combination of factors appears to be exclusive to those with EEC conceptions, but the totality of and tightly interwoven nature of reasoning about design issues is unique to those with EEC conceptions of variation.

EEC and the Data-Centric Perspective

When individuals with EEC conceptions reason from the data-centric perspective, they tend to view data through a lens of expectation—an expectation that if they properly control and explain variation, what remains will be random variation. In the absence of apparent random variability, they seek explanations for aberrations. As a result, when they reason from the data-centric perspective, their reasoning often contains elements reminiscent of reasoning from the design perspective. In working with data, they gather information about variation by exploring data through graphical representations and measurements, and they compare the characteristics and relationships they see in data with their expectation of randomness.

In bivariate settings, individuals' EEC conceptions become visible in their expectation for patterns of random variability for data in the form of residuals. After they fit a model to data, they expect the resulting residual plot to display a random scattering of points. For example, Isaac suggests having consultants score the same 50 exams as he reasons about the Consultant Task. He notes that he would expect a strong association between scores and also notes that "what I basically want is, uh, a high interrater correlation with a small confidence interval around it… in my dream world, I could get reliabilities in excess of point seven" (Isaac, Content, Lines 336-343). Isaac mentions that he would like to see a tight confidence interval around a high correlation coefficient value, suggesting that the data would be tightly grouped about a linear pattern. Isaac expects to explain most of the variation in consultants' scores with the linear relationship, and he presumably expects the remaining unexplained variation to be revealed in random patterns. In contrast with Isaac, Haley is more explicit in her articulation of an expectation for random variability in residual plots. In describing her expectation, she notes, "if your residual plot shows a pattern… like say it goes in a pattern like this. [*Haley draws a residual plot. See Figure 5-1.*] Some kind of pattern. Then these distances [residuals] are not random" (Haley, Content, Lines 1614-1618). Haley associates the pattern in her residual plot with a model that does not provide a good fit to data, which suggests to her that variation has not been adequately explained. Characteristic of their EEC conceptions, Haley's and Isaac's reasoning about data from the data-centric perspective reveals considerations consistent with their views of variation as explainable and controllable.

Figure **5-1**: Haley's Sketch of a Nonrandom Residual Plot.

Individuals with EEC conceptions also consider random variability in bivariate settings through their considerations of outlying values. As noted previously, Isaac reacts to being given centimeters and inches as the names of the variables for *x* and *y* in the Caliper Task by focusing on the rightmost two points and seemingly treating the points as outliers.

> So these ought to line up straight. Uh, if we look at these first, uh, *n* minus 2 points as kind of representative [*Isaac traces a path back and forth over the first five points in the scatterplot.*], then we're left to worry about well what the heck is going on here at the end? [*Isaac points back and forth between the two rightmost points in the scatterplot.*] (Isaac, Content, Lines 1473-1479)

Isaac evaluates the scatterplot with labeled axes by noting the representative pattern in the leftmost five points. He reacts to the rightmost two points by searching for an explanation for why the points are not representative of the known relationship, suggesting that he considers the variation in the rightmost points to be more than random variation. Haley also seeks an explanation for the unrepresentative nature of the data pattern. Unlike Isaac, she focuses only on the rightmost point. She ponders reasons behind why the rightmost point differs considerably from what the theoretical relationship would predict, wondering, "how we got one point oh, instead of one point three seven, unless he was just rounding" (Haley, Content, Lines 1735-1736). Like Isaac, she seems to be concerned about magnitude of the residual and immediately tries to find an explanation for why the rightmost value varies so far from the theoretical value. Neither Haley nor Isaac comment on the deviation of the leftmost five points from their theoretical values, suggesting that both can tolerate random variability. When the magnitude of the residual reaches a certain level, however, they seek explanations for variation that is greater than what chance would suggest.

In univariate settings, an expectation of random variability appears in recognition of reasonable variation in data and explanations for outlying values and extreme measures. For example, after Isaac looks at the sheet upon which the means and standard deviations for consultants' scores are written, he immediately reacts to the standard deviation value of 20.

> Isaac: Okay, so I've got, um, let's see. [*Isaac picks up the task sheet, and he reacts to something while looking at the sheet*.]
> R: Okay, I just saw your eyes get –
> Isaac: Twenty?
> R: Really big.
> Isaac: Uh, huh, huh, huh. [*Isaac laughs*.] (Isaac, Content, Lines 451-456)

Isaac's reaction to the standard deviation value of 20 suggests there is a limit to how much variation he expects in this context. He begins to ponder conditions that would produce a standard deviation of that magnitude.

> Isaac: The, the—it would seem to me clear that the, that the difference in variances in these two samples aren't occurring—this isn't occurring by chance alone. There's something else operating here. Um, a standard deviation of 20 in a scale from 0 to 16. I think—wow. How would you get a standard deviation that big?
> …
> R: So how could you explain that?
> Isaac: [*Pause 5 seconds*.] The explanation that leaps to mind is that somebody's just flipping a coin here. It could be a mis—it could be a misscoring. I mean, something like that can happen. Um, so maybe that would be my – I'd say look, let's go back and see the original data. This just really looks fishy. (Isaac, Content, Lines 478-498)

Shortly after he notices the value of the standard deviation for the second consultant's scores, he questions how that standard deviation might be obtained. Although Isaac suggests possible explanations at the prompting of the researcher, it appears that he had begun to consider explanations prior to the prompt. Isaac mentions that the difference between the standard deviations does not appear to occur by chance—the value is larger than what he would expect from random variation given the restrictions that exist in this context. He appears to immediately look for an explanation or cause for the unusual measure. He suggests that the value is suspicious, and his quest for explaining the variation seems to lead to his request for the actual data. Characteristic of EEC conceptions, even though Isaac is not reasoning about design but reasoning about data and characteristics of data, his reasoning remains consistent with a view of variation as expected but explainable and controllable.

Although the main examples presented in this section come from Isaac's interviews, the lesser focus on examples from Haley is not intended to suggest that Haley's reasoning from the data-centric perspective is less consistent with an EEC conception. Haley also expects random variability, and she suggests contextual explanations for the descriptions, patterns, and relationships she sees in data during exploratory data analysis. True of both Haley and Isaac is an expectation for patterns of random variability and a desire to explain apparent nonrandom patterns and unusual data values when they reason about data.

EEC and the Modeling Perspective

When individuals with EEC conceptions reason about variation from the modeling perspective, they tend to view models through a relationship lens. They use models to capture relationships among data or among variables and evaluate models according to the extent to which the models capture relationships. They also use models to determine or confirm the strength or significance of the relationships among data or among variables.

Individuals with EEC conceptions of variation seek to fit models to data with a goal of explaining variation by capturing the nature of the relationships seen in the data and doing so in ways consistent with any theoretical relationships that exist within a particular context. For example, lack of context prevented Isaac from fitting a model to the seven points in the Caliper Task. Even after he is told the context, he does not model the data based strictly on the known theoretical relationship between centimeters and inches. Although he may consider using the theoretical relationship to model the data or a line that “best” fits the seven data points, Isaac verbalizes an alternative model he would use to fit the data.

> I would probably say okay, here’s what you do. Don’t tell your stat teacher, but toss these out. [*Isaac covers up the two rightmost values in the scatterplot. See Figure 5-1.*]… And then phew, fit that. [*Isaac quickly traces a linear path over the first five points. A segment has been superimposed in Figure 5-2 to illustrate the path.*] And use that for a prediction. (Isaac, Content, Lines 1487-1494)

Isaac notes that the data should be linear, and supposing that the five leftmost points form a representative sample, he recommends fitting a linear model to those points and using that model to make a prediction. Presumably, Isaac combines his consideration of context with his observations of the data to model the data in a way that allows the student to make a meaningful prediction. Consistent with her view of variation as EEC, Haley offers a solution that is nearly equivalent to Isaac’s. Because Haley focused on the rightmost point only, her recommendation is to ignore the rightmost point and fit a linear model to the remaining six points. Both Haley and Isaac fit a model to the data that captures the nature of the relationship among the points they use and that is consistent with the theoretical relationship between the variables.

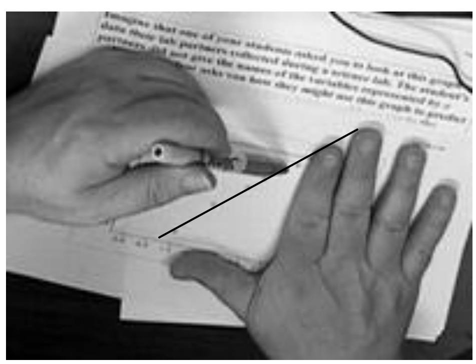

Figure **5-2**: Isaac’s Recommendation for the Student.

Regardless of whether they have EEC conceptions of variation, data analysts should consider context when fitting models to data. What seems to be different for Haley and Isaac as teachers with EEC conceptions is the extent to which context influences their decisions. Like others, Isaac suggests the best-fitting line for the full scatterplot of data for the Caliper Task would be one that passes through the approximate midrange for each vertical grouping of points. He posits that the patterns of variability in the residuals for vertical groupings would be approximately normally distributed. Where he differs from others[13] is in his reaction to whether it would be possible to have a different line as the line of best fit—a line parallel to, but below, the line he indicated. (See Figure 5-3.)

> Isaac: I suppose it could, if you—I mean you'd have to have—you'd have to have pretty seriously skewed errors. [*Isaac draws a curve by the leftmost vertical grouping of points. See Figure 5-3*.]
> R: Okay. And why do you say that?
> Isaac: Well, I'm thinking that when I do a regression, essentially what I'm doing is coming up with a model for the means at the different levels. And if I'm going to have a regression line go through there [*Isaac points to a value near the intersection of the newly drawn line and the leftmost vertical grouping of points*.], then the means are going to be closer over here [*Isaac points to a value near the bottom of the leftmost vertical grouping of points*.], which would suggest to me that those are skewed. I'm—I'm having a devil of a time trying to figure out how you would get skewed errors, though. (Isaac, Content, Lines 1616-1631)

Although Isaac describes how data would need to be distributed for the alternative model to be the "best", he seems to hesitate in embracing the alternative model as a viable alternative. Because he later attributes error to the measurement context, noting, "there shouldn't be any factors other than, um, measurement error that would account for the *y* value" (Isaac, Content, Lines 1735-1737), it seems reasonable to believe that he is considering context when he reasons between best-fitting lines and struggles to explain how the (measurement) errors could be skewed. His reaction was unlike that of any other teacher in the study and suggests his desire to have context-based explanations for the variation he sees in data.

[13] Due to the length of time Haley spent on the Consultant Task and the time limitations of the interview, she was not asked any questions about the full scatterplot of points for the Caliper Task nor was the graph ever shown to her.

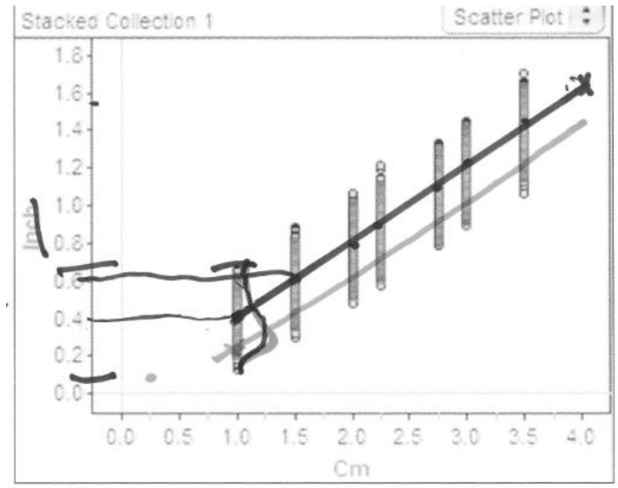

Figure **5-3**: Scatterplot With the Lines of Best Fit Suggested by Isaac and by the Researcher.

Summary

The preceding examples illustrate how conceptions of variation as expected but explainable and controllable influence Haley's and Isaac's reasoning not only from the design perspective but also from the data-centric and modeling perspectives. Although they reason from the three perspectives, their focus on reasoning from the design perspective reveals identifiable and consistent differences in the way they view variation from the views of others. They view design through dual lenses of explanation and control. Their affinity for explanation becomes evident through their privileging of experimental studies to determine cause-and-effect relationships. They view models through a relationship lens, hoping to determine or confirm the strength or significance of relationships among data and variables. They view data through a lens of expectation, expecting that if their models fit data well, what remains is data that exhibits patterns of random variability. They view the purpose of data exploration as gathering information about variation to explore and compare data characteristics and relationships.

Conception: Noise in Signal and Noise (NSN)

At the heart of Noise in Signal and Noise (NSN) conceptions of variation is a view of summary measures, data patterns, and relationships among variables as signals that are sometimes lost within noisy data. As the two teachers in this study with NSN conceptions of variation, Everett and Cheyenne see variation as the noise in data for data that does not precisely match underlying parameters, patterns, and relationships and thus interferes with identifying signals. Their view of variation as noise focuses their attention on exploring data; their quest to find patterns and relationships focuses their attention on aggregate features of data distributions while simultaneously considering individual datum that do not clearly fit the patterns and relationships. Their focus on individual and aggregate features of data is indicative of sophisticated reasoning about data and distribution (Hancock, Kaput, & Goldsmith, 1992; Konold & Higgins, 2002). Their view of variation and "the examination of data for interesting patterns and striking deviations from those" (Moore, 1997, pp. 3-4) aligns well with what Moore (1997) considers to be a major focus of contemporary statistics. Throughout discussion of the NSN conception, I draw on examples from Cheyenne's and Everett's Content interviews to illustrate facets of their conceptions, selecting the clearer and more succinct examples for description.

Everett's and Cheyenne's views of variation as noise are consistent with the idea of noise implicit in mathematics and statistics educators' descriptions of data as noise and signal. Konold and Pollatsek (2002) describe one interpretation of average as signal in noise and associate measures of center, including mean and median, with signal. It follows that measures of variation such as standard deviation and interquartile range describe the noise in data, and a data distribution becomes a "'distribution around' a signal" (Konold & Pollatsek, 2002, p. 262). Wild and Pfannkuch (1999) apply the notion of data as signal and noise more broadly than Konold and Pollatsek and describe statistics as existing to isolate and model signals in the presence of noise.

They note that making sense of statistical data “begins by trying to find patterns in the data” (p. 240). This search for patterns and relationships using summary measures and data representations during exploratory data analysis is of utmost importance to Everett and Cheyenne. As a focus on data might suggest, the data-centric perspective is prominent in their reasoning about variation.

NSN and the Data-Centric Perspective

Both Everett and Cheyenne are adept in reasoning about variation from the data-centric perspective. When they explore data, they view data through the lens of distribution (Wild, 2005), viewing data without regard for individual case information beyond the values for the variable(s) under study. Everett and Cheyenne employ what Bakker and Gravemeijer (2004) describe as an upward and downward view of data and distribution. Through an upward view of data, Everett and Cheyenne see a data set as a frequency distribution—a pointwise collection of individual values from which they can calculate summary measures (Bakker & Gravemeijer, 2004). Through downward views of data, they see an idealized distribution that they can characterize with aggregate features such as shape, center, and spread (Bakker & Gravemeijer, 2004). When they reason about patterns and trends using aggregate views and reason about individual cases such as outliers from pointwise views, they engage in what has been called distributional reasoning (Ben-Zvi, Gil, & Apel, 2007). When Everett and Cheyenne look at data, they see both the idiomatic trees *and* the forest with their pointwise and aggregate views, respectively.

Noise With Measure of Center as Signal

Individuals with NSN conceptions of variation see exploration of data as a necessary precursor to employing inferential methods. They explore data to identify potential signals and to gauge the magnitude of noise in data before attempting to establish the significance (or not) of signals. In response to reading the Consultant Task, Cheyenne and Everett both acknowledge the insufficiency of making decisions from average scores alone. Cheyenne states that:

> I would have liked to have taken a look at, um, I guess I'm a graphical person. I like to see the, the spread of the distribution to see what it is. Just looking at the means without knowing anything else about the distribution isn't gonna help an awful lot in making the decision. (Cheyenne, Content, Lines 61-65)

Through stating a need to see the "spread of the distribution," Cheyenne indicates that she needs to see the "distribution around" (Konold & Pollatsek, 2002) a signal in order to compare distributions by "reading between the data" (Curcio, 1987). Everett also notes that he would "need to know about the distribution of scores" (Everett, Content, Line 105). Both Cheyenne and Everett mention that a difference of 0.6 does not seem to be indicative of a problem; without additional information, they would be loath to state that any difference exists. What distinguishes Everett's and Cheyenne's reasoning is their request for information about the distributions rather than information about specific characteristics of the distributions, namely values for measures of variation in general or standard deviations in particular.

Everett and Cheyenne choose not to reason about data from summary measures alone. Rather, they seem to use summary measures to obtain an aggregate view of data, and they use the actual data values in the form of tables, dotplots, or stemplots to view data pointwise. Cheyenne's allusion to her graphical disposition and her presumed request for a graph to see the "spread of the distribution" suggests a desire to gain a pointwise and aggregate view of the consultants' data. Cheyenne's desire becomes clear when she is given values for the standard deviations in addition to the means.

> They [the administrators] wouldn't have done anything like five number summaries or graphical displays or anything like that? Because, again here what's missing is—this could be as simple as one huge score. [*Cheyenne points to the values of the standard deviations.*] (Cheyenne, Content, Lines 180-184)

Cheyenne's reaction suggests that she cannot get a clear image of a distribution from only the mean and the standard deviation. Her desire to examine multiple representations reveals evidence of transnumeration: a fundamental element of statistical thinking in which the thinker represents data in numerous forms to develop a better understanding of the system in which the data is embedded (Wild & Pfannkuch, 1999).

Cheyenne's reactions also present evidence of a desire to engage in reasoning characteristic of distributional reasoning (Ben-Zvi, Gil, & Apel, 2007; Reading & Shaughnessy, 2004; Shaughnessy, Ciancetta, and Canada, 2004). In particular, Cheyenne struggles with Consultant Two's standard deviation of 20, noting, "there's something strange going on there. I can't visualize that distribution" (Cheyenne, Content, Lines 204-205). Using pointwise logic, she suggests that the large standard deviation value could result from one outlier. She struggles to visualize a distribution with the noted aggregate features, suggesting that an outlying value might contribute so much noise that the summary measure of the mean produces a signal too weak to characterize the entire distribution of scores adequately.

Like Cheyenne, Everett fluently reasons about data using a lens of distribution. At times he combines aspects of reasoning from the data-centric perspective with reasoning from the modeling perspective. If he reasons about data as an aggregate collection that approximates a typical distribution pattern and uses known properties of the distribution to reason about the data, he is simultaneously reasoning from both perspectives. When Everett is given the standard deviations for the consultants' scores, he reacts to the large standard deviation for Consultant Two's scores, noting that "if it's out of 15 points, that seems unrealistic for a standard deviation" (Everett, Content, Lines 187-188). Unlike Cheyenne's initial reaction but similar to her later reasoning, Everett identifies a problem with the standard deviation when he focuses on aggregate features of the distribution. He notes that his "experience with test scores, is that they tend to be reasonably normally distributed" (Everett, Content, Lines 191-192). As a result, he suggests that a standard deviation of 20 with a mean of 10 "would indicate that scores could go from negative 10 to 30, which is way out of the range of 0 to 15" (Everett, Content, Lines 198-200). He indicates that Consultant Two's data do not produce the data pattern, or signal, one typically finds with test scores. There is too much noise for the data to be normally distributed. To justify his conclusion, he uses the interval bounded by one standard deviation above the mean and one standard deviation below the mean to argue against a normal pattern for the data. He later indicates his belief that the standard deviation of 20 is not possible for data on an interval from zero to 15. His aggregate view of the data (and the model he would normally associate with test scores) proves insufficient for adequately describing the data, and he asks to see either a stemplot or dotplot of the data, pointwise representations, so that he can identify any aberrations that exist in the data. He moves away from an aggregate view in favor of a pointwise view to explore the data further. Everett and Cheyenne both provide evidence of considering pointwise and aggregate views of data and of variation and offer no conclusions based on the conflicting messages they get when they compare their images to the summary values provided. Their flexible and seemingly effortless movement between viewing data pointwise and viewing data as an aggregate collection, which provides evidence of what some would term "expert" distributional reasoning (Bakker & Gravemeijer, 2004), typifies the data-centric reasoning about variation and patterns of variability for those who have NSN conceptions of variation.

Everett is particularly adept at simultaneously combining procedural and conceptual aspects of center and variation, or signal and noise, and combining pointwise and aggregate views to reason about data. Everett's reasoning through the Consultant Task highlights his fluency in reasoning from the data-centric perspective and his ability to draw on knowledge of procedures and conceptual properties to argue for the value that resulted from a data-entry error. Initially Everett estimates the missing score for Consultant Two by considering the signal—using calculations and properties for the mean. Everett is asked to address whether the standard deviation value of 20 could result from entering a large score for Consultant Two. Everett uses his description of a “casual understanding” (Everett, Content, Line 552) of standard deviation as the average deviation from the mean to form his response.[14] Although other teachers expressed similar conceptualizations of standard deviation, none of them used their informal characterizations to reason about the data in the Consultant Task. The absence of informal characterizations of standard deviation in reasoning is a result seen elsewhere (Clark, Kraut, Mathews, Wimbish, 2007) and suggests the novelty of Everett’s reasoning. Everett estimates the average deviation from the mean for the 49 values displayed in the dotplot shown in Figure 5-3 and combines the squared deviations for these 49 values with the squared deviation from the mean for the proposed missing score. He essentially calculates a weighted average for the 50 squared deviations, and he takes the square root of this result to confirm the value of the misentered score (Everett, Content, Lines 480-596). In his reasoning, Everett uses his informal characterization of standard deviation as a measure of noise in combination with the procedural formula for calculating a value for standard deviation to confirm the value of the 50th point. He combines pointwise reasoning related to the outlier and estimates for deviations from the mean with aggregate, conceptual reasoning about the average absolute deviation to form his conclusion. Everett’s fluency in reasoning about standard deviation stands in stark contrast to previous research that suggests many “successful” introductory statistics students have not developed appropriate process conceptions[15] of standard deviation (Mathews & Clark, 2005).

Although Cheyenne was not asked the same question that spurred Everett’s thoughts about standard deviation, Cheyenne displays proficient reasoning about variation and data in

---

[14] Everett notes that the mean absolute deviation describes the actual average absolute deviation from the mean and not the standard deviation.

[15] Mathews and Clark (2005) and Clark, Kraut, Mathews, and Wimbish (2007) use APOS theory (Asiala, Brown, DeVries, Dubinsky, Mathews, & Thomas, 1996) to suggest levels of understanding for standard deviation. Students with action-level understandings are unable to compute or to discuss standard deviation without the standard deviation formula. Those with process-level understandings are able to describe standard deviation in terms of a distance measurement from the mean.

other ways. Cheyenne seems to want as much information as she can possibly have to reason about data. She prefers having multiple representations of data and multiple summaries measures of data to make decisions from data. When she is given a table of values for the size-15 samples in the Consultant Task, Cheyenne uses technology (Minitab) to calculate descriptive statistics values, including five-number summaries, means, and standard deviations. In addition to the columns of data values she is given, Cheyenne displays boxplots and dotplots of the data for each consultant. In doing so, she examines more representations of this data and its variation than does any other teacher in the study. From the information she generates, she has multiple opportunities to examine pointwise and aggregate features of distributions in general and variation in particular. By examining the boxplots and summary values, for example, Cheyenne is able to develop an aggregate view of the data, whereas the dotplots and the table provide her with a pointwise view. Her reasoning combines "rule-based comparisons" (Rubin, Hammerman, Campbell, & Puttick, 2005), calculating a statistic and using it for comparisons, with "value-based comparisons" (Rubin, Hammerman, Campbell, & Puttick, 2005), comparing salient features tied to the context from which the data originate. Although Cheyenne reasons about data and variation in ways that differ from Everett, both Cheyenne and Everett reason through the pointwise and aggregate lenses of data and distribution. Neither draws conclusions from data before they carefully and thoroughly consider both signal and noise for data. By the time they turn to formally determining the significance of signals, they have already formed conclusions on an informal level. Their informal inferential reasoning, in which considerations of variation or noise play a crucial role, is particularly visible when they compare sets of data.

Noise With Data Patterns/Distributions as Signal

In a setting where data are properly collected and comparisons between sets of data are desired, some might succumb to the temptation to turn to inferential methods immediately. Those individuals might start with assumptions that samples originate from the same population and differences in samples are due to sampling variability. In contrast, individuals with NSN conceptions of variation turn to comparisons of data and reasoning from data to consider whether samples could plausibly be drawn from the same population. In this way, they are reading beyond the data to make predictions and inferences from the data and displaying advanced comprehension of data structure in their reasoning from data representations (Curcio, 1987; Friel, Curcio, & Bright, 1991). Their comparisons of two sets of data could include comparisons for variation within each distribution as well as comparisons of the variation between distributions as steps towards determining the relationship between data sets. For example, after Cheyenne has the graphs and summary values for the size-50 samples in the Consultant Task, she notes that:

> There is a tendency for both of them to score within the 6 through 8 range. Um, this Consultant One has a—the distribution is more spread [*Cheyenne points to the extreme values for Consultant One's scores.*], um, on both ends, um, but more so on the high end. Um…it makes me a little more cautious that it's not quite the same. (Cheyenne, Content, Lines 377-382)

Cheyenne leans towards saying that the consultants do not score the tests in the same manner based partially on the larger range shown in the dotplot for the first consultant's scores in comparison with the other consultant's scores. When Everett was given the initial dotplots, he commented on not just the difference in variability seen within each distribution but also the difference in means between the distributions, displaying what Makar and Confrey (2002) term a "tolerance for variability" (p. 2).

> I see, um, two differences, at least. One is that, um, the center, or the average score for Consultant One definitely seems higher, uh, than Consultant Two. Also, Consultant One has quite a bit more variability in his scores, uh, than Consultant Two. (Everett, Content, Lines 305-309)

Neither Cheyenne nor Everett saw the initial difference in means of 0.6 as problematic and neither was initially willing to state that a difference existed. It seems Cheyenne is beginning to think “that it’s not quite the same.” Although we cannot be sure about her meaning of “it”, she does seem to be suggesting that the signal for “it” might differ between the two distributions based on the difference in “spread”, or noise, in the two distributions. Similarly, Everett suggests that a difference exists in both “the center” and the “variability” for consultants’ scores based on the noise he sees within each distribution and the difference in means he sees between the two distributions. For both Cheyenne and Everett, it is plausible that they see the differences between consultants’ scores as more than just noise about a single signal but as a true difference in signals.

Informal inferences using data-based considerations are characteristic of individuals with NSN conceptions. They reason not only from sample distributions, but they also make comparisons that include considering the plausibility of selecting distributions with similar characteristics in repeated samplings. For example, after Cheyenne has the graphs and summary values for the size-50 samples in the Consultant Task, she sits quietly. When asked what she is thinking about, she says “I’m just looking at—well, I’m thinking about, um, if I repeat this, this 50 again and again, um, how likely would it be to get these two differences” (Cheyenne, Content, Lines 387-389). Like Cheyenne, Everett turns to comparing two distributions to determine whether there is a difference between populations. He considers what he calls a randomization test to compare the size-50 consultants’ samples. He describes the test as follows.

> We could throw all 100 of these scores into, you know, one set and then split them up into groups of 50, um, find the averages for both groups. See what the difference is. And then do that a bunch of times, over and over and over and over again. And see if a difference of point 6 [*Everett points to the mean values of 9.7 and 10.3 displayed in the summary table*.] is likely to come up just due to the random separation of the scores into two groups. (Everett, Content, Lines 364-374)

Everett’s method would allow him to test for multiple signals, including means and standard deviations, although he focuses on means. His method is based on data in that it essentially involves taking combinations of the original data to form two new samples. In this case, he randomly selects 50 tests from the combined 100 tests, and considers characteristics of the newly formed samples in comparison with the observed difference in means of 0.6. Everett's method allows him to determine if the noise he literally sees is noise that is consistent with a single signal for a population mean or consistent with noise for two separate signals or population means. The examples from Everett and Cheyenne are not intended to suggest that they never consider formal parametric tests of inference. They do—Everett considers a two-sample *t*-test for the size-50 samples and Cheyenne considers the two-sample *t*-test for the size-15 samples. These examples do, however, illustrate Everett's and Cheyenne's tendency to view variation (as well as center and distribution) through a data-centric lens. Their flexibility in reasoning about data from this perspective offers them the benefit of considering signal and noise when traditional probabilistic models do not apply. Everett also notes that his randomization approach "just helps me think about – get a first impression" (Everett, Content, Line 879) for the significance of a signal, leaving formal inferential methods to serve a somewhat confirmatory role. Everett's and Cheyenne's focus is on exploratory analysis, whereas formal inference is used in confirmatory analysis, a term introduced by Chatfield (1988).

Noise With the Relationship between Variables as Signal

When exploring relationships among variables, those with NSN conceptions seek to find a signal, if one exists, for the relationship(s). Although they consider context in their reasoning about data, context is not their main focus. They seem to be willing, although hesitant, to search for signals when no context is given. They may have a lesser sense of expectation or a lesser need to have an explanation for observed patterns of variability than others. For example, Everett and Cheyenne both express hesitation in making a prediction from the seven points plotted in the scatterplot from the Caliper Task. Although their reasons differ based on the signal of the relationship between variables that they see, both are willing to offer advice to the student. The noise, or scatter of data away from the observed relationship, does not appear to be sufficiently large to prohibit making a prediction. To do so, Cheyenne asserts that assumptions need to be made about the distribution, namely that the pattern seen in the distribution must continue beyond the rightmost point. In the end, Cheyenne notes that "if I were going to be forced to make a prediction, I would probably try to, um, do a best fit graph for the data…It looks like it might be, um, quadratic… it's where the flow of the data tends to be" (Cheyenne, Content, Lines 858-881). Initially Cheyenne suggests that the data could come from a linear, quadratic, or cubic distribution. She settles on quadratic based on the "flow," or pattern, of the data. Implicit in her selection seems to be a sense that the noise in the variability of the data from the quadratic pattern is less than the noise in the variability from linear or cubic patterns. She seemingly sees the strongest signal in the "flow" of a parabolic pattern.

For individuals with NSN conceptions of variation, their reasoning from the data-centric perspective at times overlaps with reasoning from other perspectives. To a large extent, looking for relationships between variables often includes a modeling component as the signals sometimes take the form of a model. Like Cheyenne, when Everett examines the scatterplot with seven points in the Caliper Task, he sees a signal for a quadratic pattern.

> It does seem like a quadratic [*Everett traces a parabolic path over* the *points displayed in the graph.*], but again, it's only 7 points, so I wouldn't—I wouldn't bet my life on it or anything like that. It could be that, you know, one, if I move this point up [*Everett points to the rightmost value in the graph.*], then I would think it's linear. [*Everett points to a location above the rightmost point. See Figure 5-4.*] (Everett, Content, Lines 964-971)

Everett does not state a quadratic fit with great certainty, but he suggests the student could use the quadratic relationship between the variables to make a prediction. Everett, however, does note that relocating just one point—the rightmost point—would change the signal that he sees emanating from the data from quadratic to linear. The amount of noise, or deviation of the points from the pattern, would be less for a linear model than for a quadratic pattern if the ordinate value of the rightmost point increased. The parabolic signal that Cheyenne and Everett see in data that do not all lie on the curve of a parabola is strong enough for them to offer advice, albeit hesitantly.

Figure **5-4**: Everett's Linear Pattern.

Summary of Data-Centric Perspective and NSN Conceptions

Individuals who view variation as noise in the signals and noise of data exhibit revealing indicators of their conceptions as they reason while exploring data. Those indicators include explicit, thorough, and flexible consideration of variation, in addition to center and shape, when they view data through a lens of distribution. They consider variation through pointwise and aggregate examinations of data and distributions. They use the same distributional characteristics to compare distributions, considering variation within and between distributions while contemplating the relationship between data and the populations from which the data are drawn. Finally, they look for signals in bivariate distributions to determine the relationship between variables. They examine noise pointwise in terms of outliers and influential values and examine noise in the aggregate of data, looking at residuals as part of their attempts to isolate signals. No single characteristic in their reasoning about data appears to be exclusive to those with NSN conceptions, but the totality of their facility in reasoning about and from data and their continued focus on data to reason about variation is unique to those with NSN conceptions of variation.

NSN and the Design Perspective

Cheyenne and Everett reason about variation from the design perspective using a lens of control. They seek to control variability in data to strengthen signals and thus enable identification of signals of interest. In their considerations of design, Cheyenne and Everett do not focus on expectation and explanation. Their desire is to reduce noise in data to isolate *a* signal in data rather than *the* expected signal. They also do not seem to have a strong need for determining cause-and-effect relationships.

Cheyenne and Everett employ some of the same design strategies to control variation as teachers with other conceptions of variation. One of those strategies includes using sample size to isolate signals of sufficient strength. Consider Everett's comments about the seven-point scatterplot from the Caliper Task discussed in the last section. One of the reasons Everett hesitates to make a prediction is the small sample size, noting that "it's only 7 points, so I wouldn't—I wouldn't bet my life on it or anything like that" (Everett, Content, Lines 966-967). Implicit in his words is the sense that if he had more points, he might get a stronger signal from the data. Consider this reaction with his reaction to the size-15 samples from the Consultant Task, where he notes that "it's obviously harder to draw conclusions with smaller sample sizes, um, because there is more variability that can be expected" (Everett, Content, Lines 758-760). Everett seems to associate smaller sample sizes with increased noise (more variability) and decreased opportunity for finding or isolating a signal or signals from the data. One way that he can control variation in the studies he designs is to select large samples, something he suggests for the study he designs in the Handwriting Task.

Individuals with NSN conceptions of variation design studies and offer justifications for the elements of their designs. Everett and Cheyenne consider contextual factors to speculate about potential sources of variation, and they consider explanations for variation based on the context of data. Compared to others, the difference in their reasoning lies in their primary rationale. For example, when Everett is asked how he would design the consultants' study, he reacts as follows.

> I would suggest that you pull a random sample of papers from the district and make two copies of each of those particular papers…and then distribute them to both consultants. So there's maybe a set of 50 papers that they both grade, and then since they're grading the exact same papers, we know there should be no difference in the scores other than their own personal scoring biases or scoring, um, decisions. (Everett, Content, Lines 63-71)

Everett does not offer the same design as the administrators reportedly used. Instead, he offers a matched-pairs design, presumably to control variation from the effects of variables different from consultants' scoring. He offers a design that benefits from his consideration of the context and that seems to focus on isolating a signal for the factor of interest—differences in scores. Although he briefly mentions scoring bias, he mentions no specific bias and seems to stop himself and focus instead on scoring decisions. From a focus on scoring decisions, he attributes differences in scores to the consultants, but he does not attribute any particular cause for the differences.

In addition to considering the general design of observational and experimental studies and the size of samples used in the studies, those with NSN conceptions of variation also consider the use of randomization in their designs. Neither Cheyenne nor Everett seems to see much value in finding or considering signals that come from non-random data. Everett's immediate reaction to the Consultant Task strongly suggests that he sees little meaning in results from biased samples caused by lack of randomization.

> I'd want to know first how the consultants were assigned to grade the exams. For instance, um, is it just sort of everybody's exam gets thrown in the pile and then they just grab half at random to grade—Or does one consultant get tests from one school in the district and the other consultant get tests from other schools in the district. If the sec—if that latter is the truth, then I don't think you can get

> anything meaningful out of the results that are here. (Everett, Content, Lines 21-31)

Although Everett's language is somewhat loose when he inquires about data collection, he is clear in making the point that samples collected without randomization and with an obvious possibility for bias will not produce meaningful conclusions from data analysis. Although she personalizes the example by referring to her own school, Cheyenne offers the same example in her argument for selecting random samples. Everett and Cheyenne are consistent in making sure that the data they examine are selected randomly. Neither is willing to analyze the data from the size-15 samples without some assurance of randomization. Even though their primary focus in data analysis seems to be finding signals in noise, their actions seem to suggest that the signal, no matter how much noise or how little noise, may be misleading without randomization as part of the design.

NSN and the Modeling Perspective

Cheyenne and Everett tend to reason from the modeling perspective either in conjunction with or subsequent to reasoning from the data-centric perspective. They tend to view data through a lens of relationships, searching for patterns and relationships among data values or among variables. Because of the close connection with their reasoning from the data-centric perspective, some of their modeling reasoning has been discussed. For example, in the section discussing noise when the measure of center is a signal, I noted that Everett's experience with test scores formed the basis for his generalized statement that scores tend to be normally distributed. He seems to consider a potential model for the test data—a normal distribution—which would be the signal he considers. A normal distribution has known properties against which he can compare characteristics of the data. As noted earlier, he suggests that a set of scores with a standard deviation of 20 and a mean of 10 "would indicate that scores could go from negative 10 to 30, which is way out of the range of 0 to 15" (Everett, Content, Lines 198-200). Although he does not specifically mention the property that approximately 68% of the data lies within one standard deviation of the mean for a normal distribution, Everett does consider the interval of values within one standard deviation of the mean. He compares the interval against that of possible values and notes that the two intervals do not align. Although he does not state that the data cannot follow a normal distribution, there is an implication that the data contain more noise than he would expect from a normal signal. Everett's reasoning simultaneously incorporates data-centric elements in reasoning about aggregate features of the data with modeling elements in reasoning about the feasibility of a model for the data.

Most often, Cheyenne and Everett reason from the modeling perspective after they carefully reason about data from the data-centric perspective. They tend to use models as signals or in determining the significance of signals, but they typically wait to do so until after they have thoroughly explored data. Formal inferential methods based on parametric methods that assume a normal model for a sampling distribution seem to serve a confirmatory purpose for Cheyenne and Everett. For example, Everett states a clear preference for comparing the size-15 samples in the Consultant Task with an empirical model rather than a theoretical model, particularly when the scores in the size-15 samples are labeled as scores of students from two teachers.

Everett: I was more thinking about putting all 30 of these in one pile—And seeing how they split up. That it'd be—this is an unusual split in my mind. [*Everett points to the two columns of values for the size-15 samples, with scores labeled as A's and B's for Teacher A and Teacher B. See Figure 5-5.*] More of a, sort of a randomization test.

R: Okay. And so when you originally looked at the 15 and 15, before you knew about the different teachers, that one you were thinking about the, um, two-sample *t*?

Everett: No, more of the randomization test also. Just looking at—there's way more double digits here [*column of values for Consultant Two.*] than here [*column of values for Consultant One.*], and it would be unusual if it really was equally likely to get them to both—I'd expect them to be one, two, three, four, five, six, seven, eight, nine, ten. [*Everett points*

> *from bottom to top at the double digit values in the column of scores for Consultant Two as he counts.*] There's twelve scores in the double digits, and I'd—I wouldn't expect it to be exactly six and six. [*Everett points to the two columns of scores.*] But maybe five and seven, or four and eight, but it seems like ten and two is pretty—an extreme split. (Everett, Content, Lines 847-876)

Even though the researcher believed that Everett compared the size-15 samples with a *t*-test, Everett clarifies that when he was comparing both the size-15 samples as originally presented and the size-15 samples with scores divided between the students of two teachers, he was basing his conclusions on a randomization test and the likelihood of achieving the observed division of scores if the 30 scores were repeatedly combined and randomly divided into two groups. He clearly explores the data and engages in modeling through use of his empirical test to think about whether the observed difference could result from the same population with a single signal for the mean or if each consultant produced separate and distinct signals for their average test scores. In this case, Everett suggests that there is too much noise to believe that the two consultants graded uniformly.

Consultant One		Consultant Two	
Score	Teacher	Score	Teacher
8	A	14	A
4	B	13	A
3	B	11	A
7	A	13	A
6	B	9	B
4	B	12	A
3	B	11	A
10	A	7	B
8	A	8	B
3	B	8	B
15	A	1	B
5	B	12	A
3	B	13	A
5	B	10	A
2	B	11	A

Figure **5-5**: Consultants' Size-15 Samples Broken Down by Teacher.

As mentioned earlier, context does not play the same prominent role for Cheyenne and Everett as it does for others; Cheyenne and Everett are willing to reason about data without context and without a need to know the causes behind their observations. As mentioned earlier, Everett was willing to help the student from the Caliper Task. To test whether Everett would be less likely to make a prediction from a linear model, the researcher tells Everett that the student later stipulated that he needed to use a line. Unlike others who were asked the same question,[16] Everett did not fit *a* line to the data. Rather, he fit a piecewise function consisting of two linear pieces.

> Okay, what I would have them do is I would have them divide it into two different data sets and look at the first five. [*See Figure 5-6(a).*] Create a linear model going like this. [*Everett traces a linear path over the first five points. A segment has been added to the Figure 5-6(b) to show the path of the trace.*] And then use the last three, this middle point [*Everett points to the value vertically highest on the graph.*], a second time. Have them create a linear model that goes like that. [*Everett traces a linear path from left to right over the rightmost three points. A second segment has been added to Figured 5-6(c) to show the path of the trace.*] (Everett, Content, Lines 1000-1012)

Figure **5-6(a)**: Everett's First Data Set.

Figure **5-6(b) and 5-6(c)**: Everett's First (b) and Second (c) "Pieces."

[16] Due to time limitations for the length of the content interviews, not every teacher was asked this question. In particular, Cheyenne was one of the teachers who were not asked the question.

Although context does not seem to overly influence Everett when he makes decisions from data, signals do seem to influence him. Everett earlier acknowledged that the movement of the rightmost point would suggest a signal different from a parabola. Even with the tenuous signal of an increasing followed by decreasing pattern, it seems to be sufficiently strong for him to control noise by fitting two linear segments to the data. Everett also shows less concern for context than other teachers when he considers the full scatterplot of points for the Caliper Task. He draws a line for the theoretical relationship but suggests that the caliper tends to under-measure objects. When he draws the line for the theoretical relationship, he actually creates a line with a *y*-intercept greater than 0. (See the line with a positive slope in Figure 5-7.) When asked if the theoretical line could be the line of best fit, he responds as follows.

> Uh, no. The, the real—the least squares regression line would be, uh, shifted lower and pretty much trying to go, it would go through the—or attempt to go through the means of all of these little distributions. [*Everett traces a path from left to right through the approximate middles of each vertical grouping of points.*] So I would suspect that, for the most part, it would just be shifted a little lower. (Everett, Content, Lines 1324-1330)

Even though Everett previously stated the theoretical relationship between centimeters and inches, the signal he sees in the data does not seem to contain enough noise for him to consider a different signal. Everett disagrees with the theoretical model as a possibility for best fit despite the context that might suggest agreement to others.

Figure **5-7**: Complete Scatterplot for the Caliper Task and Everett's Theoretical Relationship.

Summary

The preceding examples illustrate how conceptions of variation as noise in signal and noise influence Cheyenne's and Everett's reasoning not only from the data-centric perspective but also for the design and modeling perspectives. Although they reason from the three perspectives, their reasoning about variation is dominated by the data-centric perspective. Their reasoning about variation from the three perspectives closely relates to their search for signals. They reason about variation from the design perspective using a lens of control, with a goal of controlling variability to strengthen signals in data. They search for those signals through a lens of distribution when reasoning about variation from the data-centric perspective. To model signals or to confirm the significance of outcomes hypothesized from their data-centric explorations, they tend to reason about variation from the modeling perspective using a relationship lens. Their reasoning reveals identifiable and consistent differences in the ways they view variation from the ways in which individuals with other conceptions of variation view variation.

Conception: Expectation and Deviation from Expectation (EDE)

A view of variation as expectation and as deviation from expectation was the most prevalent form of variation conception among the teachers in this study. Eight teachers—Blake, Carl, Dana, Dustin, Frank, Gavin, Hudson, and Ivy—provided evidence to support interpretations of their conceptions as a form of Expectation and Deviation from Expectation (EDE). The most distinguishing feature of their conceptions is a view of variation juxtaposed with expectation—either an a priori expectation or an expectation acquired from exploratory analysis. They often approach statistical situations with some hypothesized expectation—expectation for particular outcomes or measures (including measures of variability), particular parameter values, particular patterns of variability, or particular relationships among variables. Their expectations stem from the statistical questions they wish to answer or from the context in which their statistical problems are set. In settings for which they have no a priori expectation, they explore data to develop a sense of expectation. In addition to expected amounts of variation, they view variation as deviation from expected outcomes or measures, deviation of statistics from parameters, deviation of observed data from expected patterns, or deviation of observed data from expected relationships.

Aspects of the preceding descriptions of expectation and deviation from expectation can be seen in the levels of appreciation for expectation and variation articulated by Watson, Callingham, and Kelly (2007). Watson and colleagues speculate that increased appreciation of the interaction between expectation and variation matches a developmental progression of increasingly sophisticated statistical ideas for students in grades three through nine, forming a critical foundation for statistical understanding. They describe expectation in terms of "probabilities, averages, 'caused' differences, and random distributions of outcomes," which somewhat align with variation as "uncertainty," "outliers," "anticipated change," and "unanticipated change" (Watson, Callingham, & Kelly, 2007, p. 84). Watson and colleagues' descriptions of variation roughly align with variation for those with EDE conceptions: deviation of statistics from parameters, observed data from expected variation or patterns of variability, observed data from expected outcomes or relationships, or expected variation in the form of random variation, respectively. In general, the descriptions of reasoning from individuals with EDE conceptions of variation in this study are intended to provide a broad overview of reasoning about expectation and deviation from expectation, with expectation referring to a broader range of measures, patterns and relationships and a priori expectation than Watson and colleagues.

Even though eight teachers provided evidence of EDE conceptions, I mainly limit my discussion to examples from three teachers (Blake, Dustin, and Hudson) with EDE conceptions. I introduce examples from other teachers only when their reasoning provides a contrast to examples discussed previously. Focusing on examples from three teachers provides the reader with a more complete image of reasoning for individuals who view variation as EDE than would be possible using examples from all eight teachers.

EDE and the Modeling Perspective

As individuals who view variation as expectation and deviation from expectation, Blake, Dustin, and Hudson tend to reason about variation from the modeling perspective through a lens of expectation. In general, in familiar context settings, individuals with EDE conceptions determine the extent to which models for relationships among data or among variables conform to expectation, and they use models to determine if deviation from expectation is greater than chance would predict. Their views of variation align with the view of statistics articulated by a statistician in Pfannkuch and Wild's (2000) study: "statistics as the fitting of models, formal analysis, and the 'measuring of evidence'" (p. 198). These three articulated characteristics of statistics align with the main statistical foci of individuals with EDE conceptions: modeling patterns and relationships, using formal inferential methods, and interpreting *p*-values in context. Individuals with EDE conceptions also reason from the modeling perspective to develop a sense of expectation in settings for which they have no a priori expectation and then reason about deviations from their newly formed expectations.

Deviation of Statistics From Parameters

Individuals with EDE conceptions of variation approach inferential settings with expectations for the relationships between statistics and parameters. Their expectations for parameters are determined by the statistical questions they hope to answer, and they attempt to determine if observed statistics are reasonable given their stated expectations. With the dominance of traditional parametric methods in introductory statistics courses (Cobb, 2005), it is not surprising that many of the teachers in this study tended to approach settings relating means and proportions with parametric methods that necessitate comparisons of observations with theoretical models. Each teacher with an EDE conception of variation suggested conducting an inferential test in their initial analytical reactions to the Consultant Task. For example, after Blake reads the Consultant Task description, he notes the following.

> You can see that it's nine point seven versus a ten point three, we can obviously, uh, do some sort of *t*-test or something like that on it to, to see if that result is significant…Obviously we got the one score was ten point three. We could, everybody could see that that was higher. The issue and the statistic—from a person who's trained in statistics—is it significantly higher. (Blake, Content, Lines 56-66)

Blake's initial reaction is to compare the means using a *t*-test to establish if one mean is significantly higher than the other mean. For Blake, the question is not whether a difference in means exists—he expects to see a difference—but whether the observed difference in means deviates from expectation by more than what is probable. Although he does not explicitly acknowledge that he is using a theoretical model to model the situation, he later notes that "the *t*-test is a nice approximation to the model we're seeing" (Blake, Content, Line 170), and he clarifies that to him significance means "not reasonably attributed to chance" (Blake, Content, Line 70). Focus on the difference in means of 0.6 and determining whether the difference deviated significantly from the expected difference in means of zero seems to dominate Blake's initial considerations for analysis and the initial analysis considerations for others with EDE conceptions. Their considerations invoke comparisons of sample characteristics with theoretical models.

For those with EDE conceptions of variation, determining whether observed results deviate from expectation by more than chance would predict tends to be at the forefront of their considerations of data. Blake describes his view of variability in statistics as something that he expects to see in a "sampling distribution-type thing." He clarifies what he means by suggesting that variability is his "expectation given the confines of the problem that has been presented, uh, in sort of random repetition of the, um, of the event" (Blake, Content, Lines 1939-1942). It seems plausible that his "sampling distribution-type thing" refers to an empirical sampling distribution built from sample statistics measured from repeated random samples of a population and for which there is an expectation that some or all sample statistics will differ in a predictable manner from the parameter of a larger population. This distribution of statistics resulting from random repetition provides Blake with an image of expectation and a model against which he can determine the plausibility of a particular event occurring based on deviation from expectation for a characteristic of the event. In contrast to Blake, Hudson states a clear relationship between variation and chance, noting that "often we're asking questions about whether it [sample result] could have been due to chance variability" (Hudson, Content, Lines 2206-2207). For both Blake and Hudson, their views of variation connect tightly with formal or informal inferential methods—methods that rely on the use of models to determine whether observed data or statistics differ from expectation by more than chance would predict—and reasoning from the modeling perspective. Each teacher who views variation as EDE introduced the idea of chance variability either early in their response to the Consultant Task or in response to the question of their associations with the word variation.

Although a suggestion to use significance tests in situations typically associated with parametric methods is neither unusual nor unique to individuals with EDE conceptions, their immediate suggestions to employ formal inferential methods subsequent to checking conditions or subsequent to making assumptions about conditions being met differs from the reactions of those with other conceptions. Individuals with EDE conceptions offer inferentially-linked reactions to statistical situations through the ideas of expectation and deviation from expectation. The exact ways in which they do so may differ from individual to individual and task to task. For example, Blake notes that he does not have sufficient information to conduct a *t*-test for comparing consultants' means. In response to a question of why he needed the standard deviation values he requested, Blake introduces the idea of expectation for standard deviation.

> If everybody in the whole world, uh, in every measurement that we ever did, if the variability, you know, from the mean, from person to person, was consistent in all measurements, then yes, we could take some universal variation amongst people, but I have no knowledge of how kids typically vary in their scores on this test. (Blake, Content, Lines 226-232)

Whereas most of the teachers with EDE conceptions commented on needing to know how the difference in means of 0.6 related to the spread of scores associated with the specific samples from both consultants, Blake observes that he has no general expectation or model for reasonable standard deviation values. His request centers on expectation for standard deviation in a way that parallels the use of sample means as estimates of expectation for population means. In contrast to Blake, although Hudson does not state a particular expectation for standard deviation, he has some idea about values that would be outside his realm of expectation for concluding a difference exists. He tests his conjecture by conducting a two-sample *t*-test with hypothesized standard deviation values and then proceeds to conduct additional tests to determine standard deviations at the threshold for declaring a difference exists. Although Blake and Hudson reason differently in the absence of standard deviation values, their reasoning contains considerations of expectation for variation based either on students' historical performance on the assessment or the limited interval of possible scores. Additionally, although both Blake and Hudson seemingly approach standard deviation with little sense of expectation, their approach to the overall task is based on expectation—expectation that the deviation in means for consultants' scores will not differ significantly from an expectation of zero.

Deviation of Observed Data From Expected Patterns

In addition to characteristically using models to reason inferentially, individuals with EDE conceptions of variation tend to use models to reason about data and expectation. They use properties of the models to develop a sense of expectation for characteristics of data or to decide if observed data deviate from expectation. For example, consider Dustin's method for estimating the standard deviation of Consultant Two's size-50 sample. After Dustin notices a mismatch between the value of the standard deviation and the dotplot for the second consultant's size-50 sample, he estimates the standard deviation from the dotplot to be around two.

> Here's a spread of 6 [*Dustin points to the minimum value and the maximum value on the dotplot for Consultant Two's scores. See Figure 5-8. A line has been added to portray Dustin's estimate for the mean of Consultant Two's scores*.] Reasonably symmetric, um, given some variation, so you've got 6 wide…let's divide it by four rather than six, uh, for two standard deviations on each side, six divided by four is… one and a half. So let's say around 2 for just want of anything—a nice number to use. (Dustin, Content, Lines 844-853)

Dustin calculates the range of Consultant Two's scores and divides that range by four. Although he does not specify here that he considers an interval of values within two standard deviations of the mean, he later notes, "the vast majority of the data really falls within two on each side" of the mean (Dustin, Content, Lines 871-872). He uses that information to estimate the standard deviation for this "reasonably symmetric" distribution (Dustin, Content, Line 861). Dustin compares the observed distribution of values against an expected pattern with known characteristics—a normal distribution model. He earlier implied an expectation of normality for test scores, and he acknowledges that these data deviate from the idealized normal pattern. He rounds his calculation to two to account "for a couple of extreme values" (Dustin, Content, Lines 875-876). Although not a perfect fit, his model provides him with a sense of expectation and estimates for the mean and standard deviation of Consultant Two's scores. Not every teacher with an EDE conception of variation reasoned from a normal distribution model to estimate values for

Consultant Two's mean and standard deviation, although they all reasoned about the values based on some sort of expectation. For example, Frank uses his knowledge of Chebyshev's Theorem and the fact that "almost all the data is plus or minus three standard deviations, generally, in a distribution, unless it's a really wild distribution" to estimate whether the given standard deviations match his expectation and to estimate values if his expectation is not met (Frank, Content, Lines 395-397).

Figure **5-8**: Dustin's Interval for Estimating Consultant Two's Standard Deviation.

Empirical rules are often invoked by individuals with EDE conceptions to reason about characteristics of normal or approximately normal distributions and to reason about how data are distributed. For those who reason about expectation and deviation from expectation, it is not unusual for them to appeal to expectations of known distributions, such as normal distributions, to estimate values for all distributions. They seem view data as a combination of "pattern and deviation" (Borovcnik, 2005), with substantial deviation at times. For example, even when Dustin seems to examine data without expectation for characteristics of the data, he turns to models to help him develop a sense of expectation. When he is faced with a distribution in which data deviate far from expected patterns, such as the distribution of Consultant One's scores, his estimates are less accurate. For the data from Consultant One, Dustin uses characteristics of a normal model and considers how much he needs to adjust for nonnormality in order to estimate a value for expectation and a value for typical deviation from expectation. Even though he stated that the data were clearly skewed, he still chose to reason from a flawed model for the data, suggesting his need for models in characterizing distributions. Other individuals with EDE conceptions used models in similar ways to develop a sense of expectation for nonnormal data distributions.

Deviation of Observed Data From Expected Relationships

Relationships between variables seem to offer those with EDE conceptions a stronger sense of expectation than relationships in univariate data or between sets of univariate data, although context surely plays a role in determining their level of confidence for expectation. Even before the context of the Caliper Task is given to Blake, he uses patterns in the data to gain a sense of expectation for the relationship between *x* and *y*. When Blake responds to the task, he suggests there is a somewhat periodic trend or pattern in the data. He also suggests a somewhat linear trend in the data by suggesting a model that considers all but the rightmost point. Blake's implication is that the trends or models for the data can be used to make a prediction for *y* when *x* is four. His final suggestion, however, offers no prediction for the student. It is here that we see his sense of expectation.

> I'm not going to tell you anything because you're outside of the domain of the data that I had to make a decision on, and so, so really I can make predictions comfortably if…I'm gonna put some boundaries on my abilities between, say 1 and I guess about 3. [*Blake points to the values of 1 and 3 on the horizontal axis of the scatterplot.*] (Blake, Content, Lines 1062-1068)

Blake suggests he can comfortably make predictions for only a restricted interval on *x*; the value of four lies outside the domain for which he has some sense of expectation. Although he recognizes several possibilities for making a prediction when *x* is four, he objects to extrapolating those patterns beyond the given data as well as within the right-most portion of the given data. Blake seems to be willing to offer a prediction, but his prediction is limited to values of *x* between one and three, inclusive. Most teachers with EDE conceptions of variation expressed hesitation in making a prediction based on extrapolation. When asked if they would be willing to make a prediction from different values of x, such as a value of three or a value close to three, they offered predictions. For this data, they were willing to interpolate but not extrapolate to read beyond the data (Curcio, 1987; Friel, Curcio, & Bright, 2001). For example, Gavin notes that he would be willing to make a prediction when x is 3.25 "because it's, uh, it's within the range of our data… our x values. Um, and it seems like we have a pretty clear pattern of increasing" (Gavin, Content, Lines 1088-1091). The combination of a value within the domain of given data along with a "pretty clear" increasing pattern of variability provides Gavin an expectation for y if x is 3.25. For Gavin and others with EDE conceptions, their confidence in and willingness to make predictions seems to be based on their level of certainty for expectation.

One of the main reasons for modeling variation is for the purpose of prediction (Wild & Pfannkuch, 1999). Individuals with EDE conceptions of variation are more willing to make predictions when the value of the explanatory variable falls clearly within an interval about the expected value. Although not everyone with an EDE conception reacts to statistical situations in the same manner, they do reason in ways consistent with expectation and deviation from expectation. Although Blake and others were uncomfortable in offering the student a prediction for x equal to four in the Caliper Task, Hudson offers a different reaction. Hudson also hesitates to offer advice to the student based in the value of x not staying "strictly within the bounds of x" (Hudson, Content, Line 1273). Like Blake and Gavin, he seems to have no expectation for y when x is four. He suggests that information about the variables might help him to model the data based on a theoretical relationship that he "might bring…into the thinking" (Hudson, Content, Lines 1329-1330). By having information about the context, he suggests that he might be able to form a better sense of expectation for the relationship between x and y. Hudson discusses a variety of options for modeling the data, including a line, a piecewise function with linear pieces, or a parabola. He stops by concluding, "absent any idea of the relationship between x and y, we might just guess the mean *y* value" because "the mean is the point that minimizes the sum of squared deviations around it" (Hudson, Content, Lines 1370-1375). Although he never states a definitive conclusion, Hudson suggests that the average, or expected, value for *y* minimizes vertical deviations from the mean, making it (for him) as good for predicting *y* as the other models he mentioned. Unlike Blake and Gavin, he seems to come to some resolution for how the student might make a prediction based on expectation. His reasoning differs from Blake's and Gavin's and ultimately his conclusion differs as well, but despite the outward differences in their conclusions, at the core of their rationales are the ideas of expectation and minimizing deviation from expectation.

As Hudson's comments might suggest, context plays a role in enabling those with EDE conceptions to develop a stronger sense of expectation for patterns in data; their sense of expectation plays a prominent role in their reasoning about relationships. After Blake is given the context of the Caliper Task, he reacts by noting, "the relationship between centimeters and inches is going to be a linear function" (Blake, Content, Lines 1152-1153). When asked what he would suggest to the student, he seems to ignore the data and offers advice based strictly on expectation from the context.

> Well I would tell them to just take the formula, the, the 2 point 5, 4 centimeters equals 1 inch, and convert it. They could, they could um, if they know that, they can just make the conversion and fudge the whole thing. (Blake, Content, Lines 1184-1187)

Rather than fit a model that ignores the two rightmost points that he believed were mistakes, Blake suggests the student should make a prediction based solely on his expectation from the theoretical relationship. Blake does not fit a model that takes context into consideration; instead he suggests the theoretical relationship that *only* takes context into consideration. After he is told that the student needs to include the data as part of the assignment, Blake notes the following.

> I would still—now that I know this, you—I would insist on the linear relationship, insist well does he want the formula or to do some sort of regression

> on a linear relationship, it's got to go up here. [*See Figure 5-9.*]…If these are, if these [seven points] are locked in, I would go to the teacher and whoever you're dealing this with, this assignment and hope to make a—start over with a new graph. Because you know you're in error here. [*Blake points to the rightmost point.*] (Blake, Content, Lines 1203-1207)

Blake does not seem to consider legitimate reasons for why the two rightmost points might deviate from expectation. Instead, he suggests that the student use either the theoretical formula or a linear relationship that would produce a prediction in agreement with his expectation. In the end, he seems to be unwilling to have the student complete the assignment from the given data.

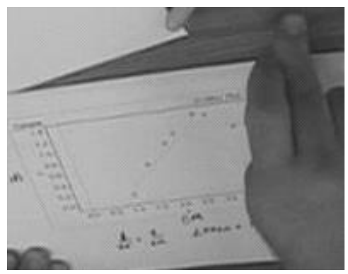

Figure **5-9**: Blake's Prediction When *x* is Four.

Although Blake might not react in the same manner if he were making a prediction for his own purposes, his sense of expectation for the relationship between the two variables along with the deviation from expectation for the rightmost point, perhaps coupled with his expectations about the student's grade, seems to be so strong that he is not willing to use the given data. To him, the data are flawed and do not adequately represent the situation and thus should not be used. He focuses on what Gould (2004) refers to as deterministic variation, that which has a "regular structure," to the exclusion of stochastic variation.

Unfortunately due to time constraints, not every teacher who views variation as EDE was asked to make a prediction after the context was given. Those who were asked either suggested using the theoretical relationship to make a prediction, tossing out the rightmost point and then using subsequent regression to make a prediction, or having the student ask their lab partners for more information about the rightmost point. In each case, their suggestions were based in their expectations within the context of measurement.

Context provides such a strong sense of expectation for those with EDE conceptions that it might unconsciously guide their development of models. Consideration of the data produced within the context matters for modeling data. For example, after Hudson is given the full scatterplot of data in the Caliper Task, he examines the data in contrast with his expectation from the context. He plots two points, one at the origin and the other at the point with *x*-coordinate of 2.54 (centimeters) and *y*-coordinate of one (inches), and draws a line through the two points to represent the theoretical relationship between variables. [See the darker line displayed in Figure 5-10(a) and the upper line displayed in Figure 5-10(b).] Because of the imprecise nature of the points he plotted, the line he draws has a slightly greater slope than that of the theoretical relationship, which in actuality corresponds with the line of best fit that passes through the center of each of the vertical groupings of points. Hudson also draws a second line through the centers and describes it as the best-fit line. [See the lower line displayed in Figure 5-10(b).]

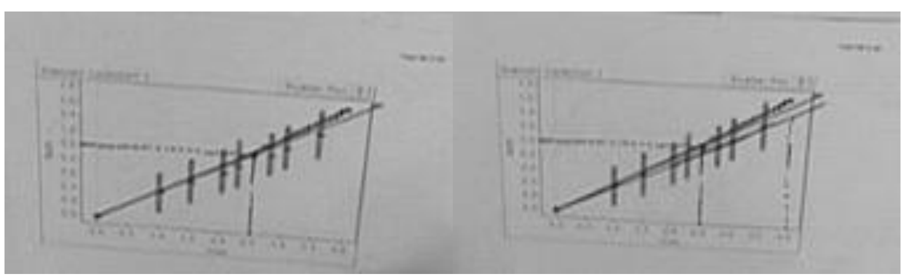

Figure **5-10(a) and 5-10(b)**: Hudson’s Theoretical Relationship (a) and Line of Best Fit (b).

Although he is aware of the theoretical relationship, Hudson does not seem to be so influenced by the context that he cannot consider a different model that provides a better fit to the data. Later, after he is given the regression output, Hudson observes that “I did something back here [graph of best-fit line] that there was actually no, no reason to do. I fixed the point at the origin because I would like, uh, something that has zero dimension to have a measurement of zero" (Hudson, Content, Lines 1743-1747). Even though Hudson was consciously aware that he was fitting a line to the data that did not match the theoretical relationship, the imprecise nature of his measurements and freehand drawing allowed him to draw a line through the centers of the vertical groupings of points and through the origin. Not until he saw the nonzero value for the intercept in the regression output did he seem to realize the effect context had on his sense of expectation. Although Hudson did not seem to be consciously aware of the powerful effect context had for his expectation, Dana recognizes the pull that context has on her thinking, noting that she was "being colored by the fact that I know the relationship between centimeters and inches" when she considered models for the data from the Caliper Task (Dana, Content, Lines 1057-1059). For Dana, Hudson, and others with EDE conceptions, their expectations from the known theoretical relationship between inches and centimeters influences their analysis. They are consciously or subconsciously influenced by the expected theoretical relationship.

Summary of Modeling Perspective and EDE Conceptions

Whether consciously aware of how their sense of expectation influences their reasoning or not, it seems clear that individuals with EDE conceptions of variation reason with a focus on expectation and deviations from expectation, particularly in deciding whether deviations go beyond what random processes should produce. Their consistent focus on models for developing a sense of expectation, use of models for examining deviation from expectation, and use of models to decide whether there is too much deviation from expectation characterizes their reasoning and distinguishes their views of variation from other views. Their reasoning seems to incorporate models naturally and without prompting. Horvath and Lehrer (1998) suggest that a "model-based perspective" aligns with the nature of statistics, which they describe as being "all about modeling data" (p. 147). No single characteristic appears to be exclusive to those with EDE conceptions, but the totality of their reliance on reasoning from models and expectation is unique to those with EDE conceptions of variation. Blake, Dustin, and Hudson are most adept at reasoning about variation from the modeling perspective. They are, however, capable of reasoning from other perspectives. When they reason from the design and data-centric perspectives, elements of their expectation and deviation from expectation view of variation appear in their reasoning.

EDE and the Design Perspective

As already mentioned, individuals with EDE conceptions of variation tend to view context with a lens of expectation. They use context to develop a sense of expectation for factors of interest to as well as factors tangential to studies. They view design through the lens of control. They attempt to design studies in ways that control variability in data in order to minimize deviation from expectation and to increase the probability that they will be able to detect significant deviations from expectation. For example, Blake, Dustin, and Hudson recommend matched pairs designs for the Consultant Task. Blake suggests pairing students by similar qualities for "every variable that I think would affect their scores" (Blake, Content, Lines 93-94), whereas Dustin and Hudson suggest having consultants score the same tests. Consider Dustin's reasoning for a matched-pairs design.

> Essentially what would be probably a better approach is to give 50 exams to Consultant A [Consultant One] and the same 50 randomly selected exams to Consultant B [Consultant Two], doing a matched pair and see if in fact there was a difference. Now you've controlled the variability of the tests between the two groups, where you're looking now strictly at, since you're giving the same test, is there a real difference between the consultants. (Dustin, Content, Lines 111-118)

Dustin essentially suggests that the matched-pairs design controls most, if not all, variation from sources other than the consultants and allows focus to be on any differences in consultants' scoring. Specifically, he notes that a matched-pairs design will allow him to determine "if in fact there was a difference." His words imply that he connects his design with a test of inference, with the design enabling him to determine if the average difference in scores deviates too far from his expectation of no difference to be plausible. By connecting design with inference, Dustin presents evidence that he considers models when he reasons about design. In general, Dustin and others with EDE conceptions desire design strategies that produce clear models for making decisions from data.

Individuals with EDE conceptions use sample size and randomization to control variation in observational studies and experiments, and they consider blocking as a strategy for controlling variation in experiments. The rationale behind their suggestions focuses on designing conditions conducive to determining significance. For example, when Blake designs a study to test the conjecture in the Handwriting Task, he mentions controlling for a variety of factors and then randomly assigning papers "after I'm done with everything that I think I can control" in the "hope that the rest of the stuff gets smoothed through" (Blake, Content, Lines 1746-1772). He focuses on control and seems to hope that the effects of any remaining confounding factors are divided evenly among observational or experimental groups. Blake also expresses concern about the effects of small samples, and he suggests that large sample sizes are more likely to produce "equal" groups. In particular, he notes, "small just means—like I said small just makes it harder to find a significant difference. Your difference typically has to be of a higher magnitude to make it significant" (Blake, Content, Lines 1798-1801). His words imply that small samples require deviations from expectation to be of larger magnitude for determining whether the deviation differs significantly from expectation. Not only does Blake suggest design strategies that allow him to make comparisons, but he connects his design to inference and the reasoning from the modeling perspective that would follow. By controlling variability, he and others with EDE conceptions reduce the variation in sampling distributions, making it easier to identify significant deviations from expectation.

EDE and the Data-Centric Perspective

Those who view variation as expectation and deviation from expectation tend to explore data to gain a sense of expectation or to explore whether data conforms to expectation. They view data and characteristics of data through a lens of expectation. For example, some sense of expectation for reasonable standard deviations in scores can be formed from reading the Consultant Task description. The range of values is 15, which even from a conservative stance suggests that the standard deviation needs to be smaller than 15. The strong reactions of Dustin and Hudson at seeing the standard deviation for Consultant Two's scores in the summary table certainly suggest that the value of 20.2 significantly deviated from their expectations. Hudson reacts with "holy moly!" (Hudson, Content, Line 563) and Dustin says "Yowza" (Dustin, Content, Line 326). In both cases, they confirm that their reactions were in response to the large value of the standard deviation. Their strong reactions and the reactions of several others with EDE conceptions were the strongest external reactions that were observed in response to the large standard deviation value.

In general, Blake, Dustin, Hudson, and others with EDE conceptions of variation reason about data with help from models and do not exhibit reasoning dominated by the data-centric perspective. They use theoretical models and characteristics of those models to reason about data when possible, and the models tend to match with their expectation for data. As a result, their reasoning from the data-centric perspective often reveals characteristics reminiscent of their reasoning from the modeling perspective. This connection between their reasoning from the modeling and data-centric perspectives was discussed in the section titled "Deviation of Observed

Data from Expected Patterns." Additionally, even when they are asked to think outside the context of statistics, just the word variation has connotations of expectation for them. When Blake is asked to describe what he thinks about when he hears the words variation or variability, he states that his view of variability "outside statistics" is "anything that's different" (Blake, Content, Lines 1950-1953). When he describes this view further, he likens his view to an outlier "that varies quite a bit from what I would expect" (Blake, Content, Lines 1949-1958). His follow-up suggests that he conceives of variability in terms of expectation and deviation–expectation for a distribution of values and the deviation of a data value from expectation of particular outcomes.

Summary

The preceding examples illustrate how conceptions of variation as expectation and deviation from expectation influence Blake's, Dustin's, and Hudson's reasoning not only from the modeling perspective but for the design and data-centric perspectives as well. Although only a few examples were drawn from the data from the other five teachers with EDE conceptions, their reasoning about variation is consistent with a view of variation as expectation and deviation from expectation. Examples to support this claim would look similar to examples presented throughout discussion of EDE conceptions. The dominance of the modeling perspective and consideration for expectation and deviation from expectation in their reasoning reveals identifiable and consistent differences in the ways they view variation from others.

Comparison of Conceptions

Similarities and differences can be seen throughout the preceding discussion of the three types of conceptions: expected but explainable and controllable (EEC) conceptions, noise in signal and noise (NSN) conceptions, and expectation and deviation from expectation (EDE) conceptions of variation. Those differences are elucidated further in this section. Because this section will include examples from individuals for each of the three conceptions throughout, a recap of the teachers and their conceptions is displayed in Table 5-1.

Table **5-1**: Teachers and Their Conceptions of Variation.

Conception	Expected but Explainable and Controllable (EEC)	Noise in Signal and Noise (NSN)	Expectation and Deviation from Expectation (EDE)	
Teachers	Haley Isaac	Cheyenne Everett	Blake Carl Dana Dustin	Frank Gavin Hudson Ivy

Summary of the Three Conceptions

Individuals with different conceptions of variation view variation in distinctly different ways. Individuals with EEC conceptions of variation primarily see variation as something they need to control and explain to uncover relationships among data and among variables. They control variation by implementing carefully selected data collection methods; they seek to explain as much variability in data as possible and privilege explanation of systematic variation contributed by causal factors. As a result, the design perspective dominates considerations of variation for those with EEC conceptions. Although design considerations are prominent in the reasoning of individuals with NSN and EDE conceptions of variation, their views of variation do not align as closely with design considerations as the views of those with EEC conceptions.

Individuals with NSN conceptions of variation see variation in data as something through which they need to sort to find signals. Variation is the noise in data that sometimes obscures the signals of interest. They prefer to explore variation in data through measures and multiple representations, which results in the data-centric perspective as most present in their reasoning about variation.

Unlike those with NSN conceptions, individuals with EDE conceptions of variation tend to explore data through the use of models. They see variation as something that can be expected in data, in statistics, and in patterns of variability for relationships in data for variables or among variables. They expect that data, statistics, and patterns can be modeled and at times have expectations for the form of the models. They also see variation in deviations or differences from expectation. They use models to describe variation and patterns of variability in data and to determine whether deviation from expectation is greater than what they would expect from random variability. Reasoning about variation through models results in reasoning dominated by the modeling perspective.

As differences in the dominance of perspectives and differences in the views of variation associated with the three types of conceptions of variation might suggest, the ways in which individuals reason about variation differ in relation to constructs associated with each perspective. Analysis of the 16 teachers' data revealed that the ways individuals look at design, data exploration, and models differ according to their conceptions. Succeeding discussion about the similarities and differences of the three conceptions of variation focuses on not only how individuals view design, data, and models but also their perceived purposes for these things. Additionally, contextual considerations affect reasoning about variation with regard to design, data exploration, and models. Because context transcends all areas of statistics, individuals' views of context and their perceived roles of context are discussed. Table 5-2 contains a summary of the lenses and purposes for design, data exploration, models, and context that are characteristic of each type of conception. Similarities and differences among conceptions in each of these areas are discussed next.

Table **5-2**: Dominant Lenses and Purposes for Models, Data Exploration, Design and Context for Each Type of Conception of Variation.

	EEC Conception	**NSN Conception**	**EDE Conception**
Relationship to variation	Explain it and control it	Sort through it	Expect it and model patterns of it
Dominant perspective from which variation is viewed	Design	Data-Centric	Modeling
Dominant lens through which design is viewed	Explanation and control lens	Control lens	Control lens
Purpose of design	Control variability in data to increase potential for determining and explaining relationships, particularly cause-and-effect relationships	Control variability in data to strengthen signals and enable identification of signal(s) of interest	Control variability in data to minimize deviation from expectation and to increase probability for detecting significant deviations from expectation
Dominant lens through which data are viewed	Expectation lens—look for patterns of random variability	Distribution lens—search for patterns and relationships as signals in noise	Expectation lens—look for whether data conform to expectation
Purpose of data exploration	Gather information about variation (in the form of descriptions, measurements, and representations) to explore and compare data characteristics and relationships	Find signals—summary statistics, data patterns, or relationships among variables—in the noise of data that do not precisely match the statistics, patterns, or relationships	Explore data to gain a sense of expectation and to explore whether data conform to expectation (with expectation taking the form of particular outcomes or measures, parameter values, patterns of variability, or relationships among variables)
Dominant lens through which models are viewed	Relationship lens—use models to capture relationships among data and among variables	Relationship lens—search for patterns and relationships among data and among variables	Expectation lens—determine the extent to which models for relationships among data and among variables conform to expectation
Purpose of models	Determine or confirm strength or significance of relationships among data and among variables	Model signals to explore characteristics of the data or to determine, quantify, or confirm significance of signals or of including factors in the models of signals	Determine if deviation from expectation is greater than chance would predict and determine the significance (or not) of expected relationships
Dominant lens through which context is viewed	Explanation and control lens—look for factors that are potential contributors of variability and that need to be controlled or explained	Anticipation lens—consider potential contributors to noise and reasonable variability in data	Expectation lens—develop a sense of expectation for or recall a priori knowledge about expectation for variation in factors of interest and factors tangential to studies
Purpose of context	Identify potential sources of variation; identify theoretical values or relationships among parameters or variables, respectively; and determine feasibility of conclusions	Identify potential contributors to noise and to identify reasonable variation in the factor or factors of interest	Develop expectation for particular outcomes or measures, parameter values, patterns of variability, or relationships among variables

Design

Controlled variation in data is an important outcome of well-designed statistical studies, and control is part of how individuals with each of the three conceptions view design. Individuals with NSN conceptions and individuals with EDE conceptions view design primarily through lenses of control, whereas individuals with EEC conceptions mainly view design through lenses of both explanation and control. Although each of the lens descriptors alludes to control, there are two ways in which the conceptions differ: purposes for viewing design and the role of explanation in design.

Individuals with different conceptions differ in their purposes for viewing design and do so in ways that reflect crucial aspects of their conceptions. Teachers with EEC conceptions of variation see the main purposes of design as both controlling variability in data to increase the probability for determining relationships among data and among variables, and explaining the relationships apparent in data. Those with NSN conceptions view the primary purpose of design as controlling variability in data to strengthen signals that emanate from data and to allow them to identify signals of interest for their research questions. Individuals with EDE conceptions view the main purpose of design as controlling variability in data to minimize deviation from expectation and to increase probability for detecting significant deviation from expectation. From these descriptions, we see that design is a tool for those with EEC conceptions to maximize their potential for describing relationships, particularly cause-and-effect relationships; for those with NSN conceptions to identify signals of interest by strengthening signals in noisy data; and for those with EDE conceptions to detect significant deviations from expectation by minimizing deviations in data.

Differences in reasoning aligned with differences in conceptions appear in individuals' reactions to considering study design and in designing studies. For example, after Everett, who has an NSN conception of variation, describes the signal he sees in the initial scatterplot for the Caliper Task, he hesitatingly uses the pattern to extrapolate a predicted value. He bases his hesitation on the small sample size of seven and indicates that a different signal emerges with a change in just one point. Implicit in his reaction to the small sample size and unclear signal is a sense that increased sample sizes amplify signals from data. In alignment with an NSN conception, he associates sample size as an element of design with control; a larger sample size should control noise in data and strengthen the signal for the relationship between variables. Blake also describes the controlling effects of sample size when he designs a study for the Handwriting Task. Typical of his EDE conception of variation, he focuses on minimizing deviations. He implies that small samples require deviations from expectation to be of larger magnitude to determine significant deviations from expectation. We see evidence that Blake controls variability by increasing sample size; he reduces variation in a sampling distribution and enables easier identification of significant deviations from expectation.

Haley describes the controlling effect of sample size on the variation of sampling distributions as she reasons about sample sizes for the Consultant Task. She notes that by increasing sample size, she "tightens" the sampling distribution, which allows her to draw "more accurate" conclusions (Haley, Content, Lines 693-694). True to her EEC conception, she considers not only how sample size affects her abilities to determine whether the consultants differ in their scoring but also how methods can be used to show how the consultants differ in their scoring. She suggests determining the relationship between consultants' scores by comparing their scores against some known standard. As an individual with an EEC conception, Haley evidences using sample size as one strategy to control variation for determining the relationship between consultants' scores and complementing her control strategies with strategies that allow greater explanatory power for the relationship between consultants' scores. From these three teachers with three different conceptions, we see evidence of using sample size as a design strategy to control variation. The reasons for controlling variability, however, differed in ways consistent with their conceptions.

For each conception of variation, control is part of how design is viewed. One definition of control is "changing the pattern of variation to something more desirable" (Wild & Pfannkuch, 1999, p. 236). The major differences among conceptions appear to be in what individuals consider "more desirable." For those with EEC conceptions, more desirable means controlling variation and achieving explanatory power. They privilege experimental design, as experiments provide much greater explanatory power than observational studies for observed patterns of variability. For those with NSN conceptions, more desirable means improved ability to sort through the variability in data by strengthening signals coming from data. Individuals with EDE conceptions desire increased ability to detect significant variation, or deviation, from expectation.

Data

To answer statistical questions, data collected from observational studies and experiments are typically explored before formal inferential methods are employed. Individuals with NSN conceptions explore data by primarily using a lens of distribution, whereas individuals with EEC conceptions and individuals with EDE conceptions explore data by predominantly using lenses of expectation. Differences among conceptions appear both in approaches to exploring data and in the purposes for exploring data.

Teachers with NSN conceptions explore data to search through the noise of data to find patterns and relationships, or signals, in data. They see the purposes of data exploration as finding signals in the form of summary values, data patterns, or relationships among variables. Those with EEC conceptions explore data to compare data characteristics and relationships. They expect that properly controlled and explained variation will produce data with random patterns of variability. Individuals with EDE conceptions explore data to gain a sense of expectation or to determine whether data conform to expectation. When they attempt to establish conformance, their expectation stems from a priori contextual knowledge, including knowledge formed from prior statistical study. Data exploration is a tool for individuals with EDE conceptions to develop a sense of expectation or determine whether data conform to expectation; for individuals with EEC conceptions to determine the extent to which variation has been explained, with an expectation that data will reveal random patterns of variability; and for those with NSN conceptions to search for signals in the noise of data largely without expectation for the signals that may underlie the data.

Differences in reasoning during data exploration are visible in individuals' reasoning about relationships. Those differences align with differences in conceptions. In her reasoning about the size-15 samples in the Consultant Task, Cheyenne, a teacher with an NSN conception of variation, looks for signals in data for the population distribution of scores and characteristics of the distribution. She initially examines descriptive statistics, dotplots, and boxplots of the data for each consultant to see what the data "[a]re telling me" (Cheyenne, Content, Line 605) about the distribution(s). She later comments that a sample of size 15 is "silly" because, with a larger sample size, she can "get a clearer picture of the distribution" (Cheyenne, Content, Lines 560-567). In alignment with her NSN conception of variation, Cheyenne approaches data to hear the signals told by the data through reasoning about data and distribution. She explores data to get a picture of the larger population distributions and parameters that characterize the populations by identifying patterns and relationships in data.

In alignment with her EEC conception of variation, Haley approaches data with a sense of expectation. For example, as Haley reasons about the initial scatterplot in the Caliper Task, she states that she expects to see a patterned residual plot when a model fit to data does not properly explain variation. Presumably, she would explore data to produce a residual plot and would expect random scatter in residual plots for models that provide a good fit to data. Compare Haley's expectation with the expectation articulated by Hudson, a teacher with an EDE conception of variation, in his reasoning about the same scatterplot. Hudson seemingly expects a linear relationship between inches and centimeters. He notes that proper measurement with a caliper results in points plotted close to what his model would predict, and he expresses surprise at how the points in the scatterplot deviate from his expectation. He seems to mentally explore the data in a residual plot to discuss how each point of data differs from his expectation. True to his EDE conception of variation, Hudson approaches the Caliper Task data with a sense of expectation and explores the data to reason about how the data deviate from his expectation of a linear relationship. In contrast, Haley approaches the Caliper Task data with an expectation for a residual plot that exhibits random variability. Representative of her EEC conception of variation, Haley reasons about the fit of a model by examining the pattern of variability in the plot of residual values.

To answer statistical questions, Cheyenne, Haley, and Hudson create and manipulate data representations, transform data as needed, and consider multiple summary measures to find signals in data or to consider whether data conforms to expectation. They explore the data collected from observational studies and experiments through what Wild and Pfannkuch (1999; Pfannkuch & Wild, 2000) call transnumeration. Transnumeration encompasses transforming data by manipulating graphical displays of data, transforming data by using different types of graphical displays, considering multiple summary measures, and using the displays and measures that best represent data for further analysis. Transnumeration is a foundational aspect of statistical thinking (Pfannkuch & Wild, 2000; Wild & Pfannkuch, 1999), and every teacher in this study, regardless of their conception, explored data through transnumeration for reasoning about variation. Differences among conceptions appear in the perceived purposes for transnumeration and in the ways in which data are approached. Individuals with EEC conceptions approach data with a sense of expectation. They transnumerate to determine the extent to which variation in data are explained, and they expect random patterns of variability for data in which variation in characteristics and relationships is explained. Individuals with EDE conceptions also approach data with a sense of expectation, but their expectation is tied to the characteristics and relationships they expect data to exhibit. They transnumerate to gain a sense of expectation or to explore whether data conform to expectation. Those with NSN conceptions approach data with an eye towards distributional characteristics. They transnumerate to search for patterns and relationships, or signals, in the noise of data. The signals for which they search include signals for summary measures, data patterns, or relationships among variables.

Models

Fitting models to data and reasoning from models are important aspects of statistical thinking (Pfannkuch & Wild, 2000). Individuals with EDE conceptions of variation view models primarily through lenses of expectation for relationships—they have some preconceived expectation for relationships among data or among variables and expect models to convey these relationships or to clarify how a situation deviates from expectation. Individuals with EEC conceptions and individuals with NSN conceptions view models predominantly through relationship lenses. Differences among the three types of conceptions appear in the degree of and focus of expectation with which data models are approached.

Individuals with EDE conceptions of variation view models with expectation; they determine the extent to which models conform to expectation, determine if deviation from expectation is greater than chance would predict, and determine whether the expected relationships are significant. Those with EEC conceptions approach models and modeling not with pre-existing expectations for relationships but rather with expectations that they can use models to capture relationships among data and among variables and determine or confirm the strength or significance of relationships. Individuals with NSN conceptions approach models with a sense of anticipation for the significance of patterns and relationships formed during their explorations with data. In combination with searching for patterns and relationships among data or variables, they model signals to explore characteristics of data and to determine, quantify, or confirm the significance of outcomes or of including factors in a signal's model.

Differences in reasoning representative of differences in conceptions are seen in teachers' reactions to using models to make predictions or making predictions for a model after being told the context for the Caliper Task. Blake, a teacher with an EDE conception of variation, describes an expectation for data to match the theoretical model of the known relationship between inches and centimeters. He suggests that the student use the theoretical model to predict the caliper measure in inches, essentially ignoring the data in the scatterplot. Blake implies that the rightmost point deviates too far from his expectation for the deviation to be a result of random variability. True to his EDE conception of variation, Blake bases the prediction on a model generated from his expected relationship between inches and centimeters. He seemingly concludes that an empirical model fit to the data deviates further from his expectation than chance would predict.

In his response to the student in the Caliper Task, Isaac, a teacher with an EEC conception of variation, explains how he could attribute an error in measurement to the rightmost two points as he recommends ignoring the rightmost two points.[17] He suggests that the student fit a model to the five remaining points and make a prediction from that model, presumably noting the residuals for a model of the targeted relationships would not be randomly scattered if the

[17] Isaac posits a reasonable explanation for the error in the rightmost two points, which he suggests would allow him to safely disregard the points.

omitted points were included. In alignment with his EEC conception, Isaac suggests a model that captures the relationships both in the data and between the variables and one that produces a residual plot pattern that suggests his model captured the relationships as well as possible. Everett, a teacher with an NSN conception of variation, constructs a linear model for the underlying signal of caliper measurements and not for the theoretical relationship between inches and centimeters (*inches = centimeters/2.54*), which may or may not match the caliper measurements. He uses context to suggest a linear model but fits a model to all seven points. He examines a few residual distances and concludes that the line he fit to the data is "reasonable" based on the amount of noise he estimates. Characteristic of his NSN conception of variation, Everett models the signal he sees for the relationship between variables.

Some statisticians describe statistics, in part, "as the fitting of models" (Pfannkuch & Wild, 2000, p. 138). As the preceding consideration of teachers' reasoning with and about models illustrates, characteristics of an individual's conception of variation can be seen in foci to which the person attends while fitting models and in evidence the person uses to make predictions and decisions about relationships from models. Individuals with EDE conceptions approach models with expectation and examine evidence for the extent to which data deviate from expectation. Their reasoning often incorporates formal inferential analyses to determine whether deviation from expectation is more than what chance would predict. Those with NSN conceptions focus on models for the relationships they hear through the noise of data as they make informal inferences from data and approach these models with expectation that forms during their explorations with data. The degree of their emerging expectation is less than that of the pre-existing expectation that dominates how individuals with EDE conceptions approach models to formally determine the strength of evidence for their hypotheses. Like those with NSN conceptions, individuals with EEC conceptions focus on relationships in their modeling activities. They fit models to data to capture the relationships among data or among variables. They employ formal inferential techniques to measure the strength of the evidence for the relationships they capture.

Context

Context plays a central role in statistical reasoning and forms the basis of distinctions between mathematical reasoning and statistical reasoning (e.g., Moore, 1990, 1997). The role of context in statistical reasoning is much richer than the brief mentions of it in the preceding comparisons suggest. Individuals with EEC conceptions of variation view context primarily through lenses of explanation and control when they look for factors that are potential contributors to variability that then needs to be controlled or explained. Those with EDE conceptions of variation view context through lenses of expectation—expectation for particular outcomes or measures, parameter values, patterns of variability, or relationships among variables based on context. Individuals with NSN conceptions of variation view context through lenses of anticipation by considering potential contributors to noise and considering reasonable variability for data. Context plays a lesser role for those with NSN conceptions than for those with other conceptions.

Isaac, an individual with an EEC conception, exemplifies using context to consider potential sources of variation that he then needs to control when he describes the design he would use for the Consultant Task. He suggests that various student characteristics, including gender, ethnicity, and socioeconomic status, might contribute variability to consultants' scores. He bases his suggestions on contextual considerations that include the subject matter of the assessments and typical ways that student data are disaggregated in reports from state assessments. He suggests selecting exams using a stratified sampling scheme to control the variability he expects from these sources. Characteristic of his EEC conception of variation, Isaac uses context to identify factors with potential to contribute variability along with a strategy for controlling the variation. Hudson's main focus in the Caliper Task is his expectation for the relationship between inches and centimeters. As often happens with those with EDE conceptions, prior knowledge indicates the expected pattern of variability for the relationship among variables. Everett, who has an NSN conception, uses context to anticipate reasonable variation in measurements for the Caliper Task. He anticipates that students' measurements will differ and that the underlying signal might be linear based on the linear relationship between inches and centimeters and reasonable variation in residual values. True to his NSN conception of variation, he uses context to consider contributors to the noise in data and whether the observed variability in data is reasonable for his anticipated contributors.

Isaac's, Hudson's, and Everett's uses of context differ according to their perceived roles of context. Individuals with EEC conceptions of variation perceive context to be a tool for explanation and control. Individuals with both NSN conceptions of variation and EDE conceptions of variation perceive the role of context in terms of expectation, but to different ends. Those with NSN conceptions see context as an aide in anticipating potential sources of variability, whereas those with EDE conceptions see context as a tool for forming particular expectations related to the variation contributed by different sources.

Conceptions and Teachers in This Study

There are four teachers whose names do not appear with any of the conceptions listed in Table 5-1. Each of these four individuals reasons without elaboration and in ways that make their conceptions difficult to identify. In general, they each reason in ways that are not inconsistent with one of the three conceptions, but they do not present adequate evidence to definitively identify their conceptions. Although Eden, Faith, Georgia, and Jenna do not exhibit a clear type of conception of variation, they each exhibit reasoning that suggests their conceptions may align with one of the EEC, NSN, or EDE conception types.

As an example, consider Jenna's reasoning about the Consultant Task. Upon reading the task description, she suggests conducting a test of inference to determine whether the consultants' scores differ. With every successive question she is asked, she describes steps for the inference test in a procedurally-sequenced order. She refers to using a normal model as she responds with reasoning from the modeling perspective. She draws a generic normal curve and describes how the curve relates to a *p*-value. A portion of her reasoning about the task is duplicated below. Bolded statements provide some evidence of Jenna's conception of variation, which will be discussed after the passage.

> Jenna: From a statistical point of view, I mean, point 6 could be nothing or it could be a lot. It really—
>
> R: So when you say that. It could be nothing. In order for it to be nothing, what would need to be true?
>
> Jenna: In order for it to be nothing, you would have to fail to reject the null hypothesis… they computed an average score. And so now you need a test to see whether or not these average scores—**if there's a difference between the two scores**. Difference between the **average scores**…you would fail to reject the null if your *p* value would be greater than the significance level...So to figure out a *p* value, you need to come up with the test statistic.
>
> …
>
> R: And what does this test statistic do?
>
> …
>
> Jenna: Okay, **so you want to base everything on the pretty normal curve**. And then you're going to have. Let's see, here, something like this [*Jenna draws a normal curve. See Figure 5-11.*]…Okay. **And we're testing to see, going extreme**… And so these are extremes. [*Jenna points to the two shaded regions under the curve she has drawn. See Figure 5-11.*]
>
> R: So you've drawn this picture. How does that relate to what's here? [*R points to the Consultant Task description.*]
>
> Jenna: Because the value of *x* down here [*Jenna points to x.*], which is the test statistic, okay, is actually computed using the information that is given to you here. [*Jenna points to the task sheet.*] (Jenna, Content, Lines 50-125)

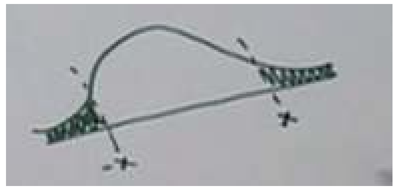

Figure **5-11**: Jenna's Normal Curve.

In Jenna's response to the Consultant Task, we see no evidence that she considers issues of design, or in particular, how data were collected, as she immediately suggests conducting a significance test. Nor do we see any consideration of the consultants' data. Jenna indicates that a difference of 0.6 could be "nothing." Unlike others, when she is asked what would need to be true for the difference to be nothing, she responds by stating a conclusion in terms of the null hypothesis. She indicates that she wants to determine if there is a difference between average scores based on an implied expectation that the difference is "nothing." Her reference to "going extreme" seems to coincide with determining whether the difference in means deviates further from expectation than chance would predict. Implicit in her consideration of a normal distribution may be some sense of variability, but her reasoning does not stray from modeling variability in sample statistics for making inferences from data [MP3]. We see that Jenna's reasoning about the normal distribution is consistent with the purpose of models described by those with EDE conceptions. She provides little evidence in this passage, however, for how she views design, data exploration, or context. Jenna's reasoning for the remainder of the Consultant Task as well as for the Caliper and Handwriting Tasks remains at a general level that is not inconsistent with an EDE conception but that does not fully evidence the characteristics of an EDE conception as shown in Table 5-2.

Eden's reasoning suggests alignment with a view of variation as EEC. She focuses on context throughout her consideration of variation for the three tasks. For example, she hesitates to draw any conclusions from the size-50 samples in the Consultant Task based on the design employed by the administrators. She seeks explanations for the variation she sees, including wanting an explanation for why the rightmost point in the seven-point scatterplot for the Caliper Task varied from the other six points. She seems to view the purpose of design in ways consistent with increasing her potential for determining and explaining relationships and in particular cause-and-effect relationships. Although she does not exhibit other characteristics of EEC conceptions, she also does not exhibit any of the defining characteristics of NSN or EDE conceptions.

Faith and Georgia reason about variation in ways consistent with one or more of the identifying characteristics discussed for NSN conceptions. For example, Faith engages in exploratory data analysis in ways consistent with NSN conceptions. As soon as Faith has the data for the size-15 samples in the Consultant Task, she uses technology to construct boxplots of the data and to calculate five-number summaries. Although she does not calculate values for means and standard deviations, she indicates that she thought she already had the values—the summary values for the size-50 samples. She seems to be viewing the data through a distribution lens as she searches for signals in the noise of data. At no point does Faith suggest conducting a *t*-test or any other test of significance to determine whether a difference in consultants' scoring exists. Instead, she seems to prefer making a data-based argument to suggest a difference exists. Both Faith and Georgia do not provide evidence of reasoning consistent with every characteristic of NSN conceptions as shown in Figure 5-2. They also do not provide evidence of reasoning inconsistent with NSN conceptions, such as reasoning characteristic of the other two types of conceptions.

These four teachers exhibit what appears to be superficial and at times faulty reasoning about variation. The reasoning they do exhibit, however, is suggestive of one of the three types of conceptions of variation. Their conceptions of variation appear still to be developing.

Concluding Comments

This chapter answers the research question, "What conceptions of statistical variation do secondary mathematics teachers who are recognized leaders in AP Statistics exhibit?" Three types of conceptions of statistical variation emerged from analysis of the 16 teacher-leaders' content and context interviews: Expected but Explainable and Controllable (EEC), Noise in Signal and Noise (NSN), and Expectation and Deviation from Expectation (EDE). These three types of conceptions reveal identifiably unique views of variation, but they do not form a hierarchy with regard to understanding of variation. At least one teacher with each conception exhibited reasoning consistent with robust understanding of variation. In the next chapter, I describe what it means to have a robust understanding of variation and how robust understanding of variation arises for these three different types of conceptions of variation.

Chapter 6

Robust Understandings of Variation

A prerequisite to addressing the second research question of this study (For those secondary AP Statistics leaders who exhibit robust understandings of variation, what are the activities and actions that contributed to their current understandings of variation as reflected in their perceptions and recollections of experiences?) was the development of a description for what it means to have robust understandings of variation. For the purposes of this study, robust understandings of variation are defined to be integrated understandings of variation from the design, data-centric, and modeling perspectives. This chapter elaborates on this definition and provides evidence of reasoning that is indicative of robust understandings drawn from the responses of the five teachers who exemplified robust understanding. The chapter concludes with a description of a relationship between robust understandings of variation and the Expected but Explainable and Controllable (EEC), Noise in Signal and Noise (NSN), and Expectation and Deviation from Expectation (EDE) conceptions of variation.

The description of robust understandings of variation draws on a conceptual framework using the Structure of the Observed Learning Outcome (SOLO) Model (Biggs & Collis, 1982, 1991). Development of the conceptual framework began with descriptions of conceptions and understandings of variation and reasoning about variation in existing expository and research literature, much of which utilizes the SOLO Model. The framework evolved further from analysis of the 16 teachers' conceptions and understandings of variation.

Framework Based on SOLO Model

In Chapter 3, the SOLO Model (Biggs & Collis, 1982, 1991) was introduced to cast understandings of variation in terms of levels of response in the formal mode. To recap what was described in Chapter 3, the unistructural (U_1), multistructural (M_1), and relational (R_1) levels within the formal mode form a cycle of levels of reasoning about variation for each perspective: design, data-centric, and modeling. Figure 6-1 displays a graphical representation of the cycle of levels of response for understandings of variation within each of the three perspectives as used in this study. The unistructural level corresponds with responses focused on a single element from a given perspective, such as anticipating the effects of sample size on the variability of a sample or statistics used to characterize a sample when designing a study or critiquing a study design from the design perspective. (Table 6-1, which appears at the end of this discussion of the "Framework Based on SOLO Model" and the "Development of the Detailed Framework," contains a complete list of these elements and indicators of the elements for each perspective, which will be discussed in greater detail in succeeding sections of this chapter.) The multistructural level corresponds with responses that embody two or more disconnected elements from a given perspective. The relational level corresponds with responses that reveal integrated reasoning among elements from a given perspective, indicative of *relational reasoning within a perspective*.

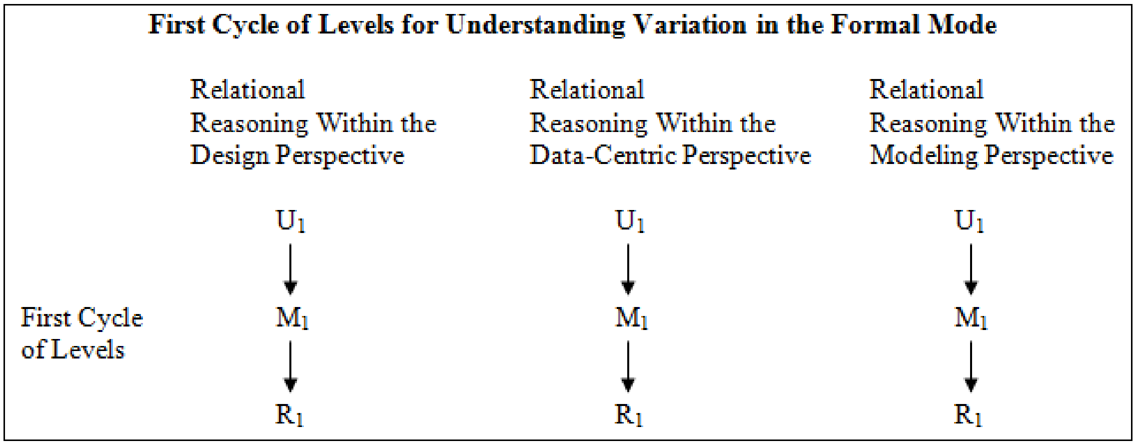

Figure **6-1**: SOLO and the Cycle of Levels for Each Perspective.

Figure 6-2 depicts two cycles of levels for understandings of variation. The second cycle of levels in the formal mode (the cycle in the bottom half of the figure) represents integrated reasoning about variation from the three perspectives. The levels subscripted with a "2" depict the second cycle and the arrows represent increasingly sophisticated reasoning. Reasoning indicative of relational reasoning within a perspective (R_1) becomes the unistructural level in the second cycle of levels of response (U_2). Individuals who reason at the multistructural level in this second cycle (M_2) exhibit relational reasoning for two or three perspectives. *Relational reasoning across perspectives*, indicative of *robust understandings of variation*, requires an integration of reasoning about variation across the three perspectives (R_2).

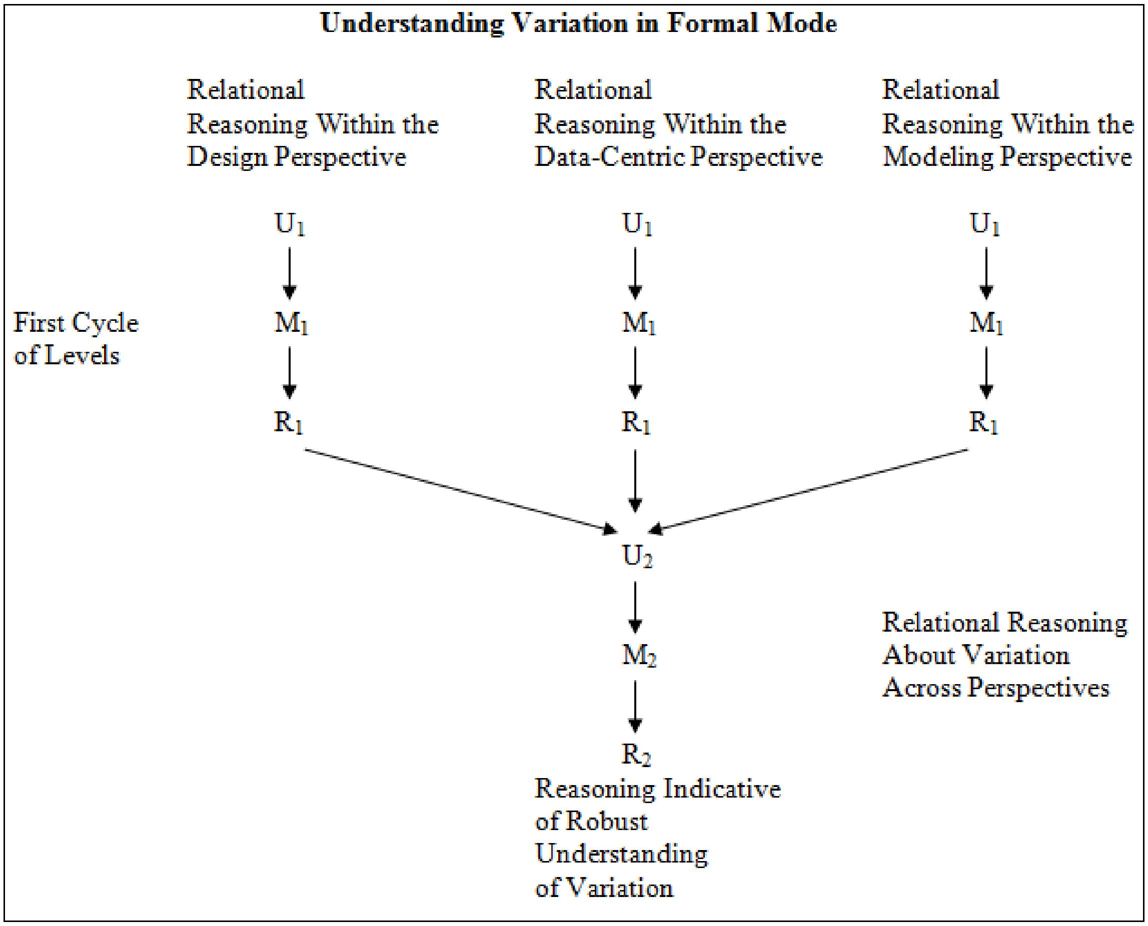

Figure **6-2**: The SOLO Model and Robust Understandings of Variation.

Development of Detailed Framework

Chapter 4 details the data analysis that guided development of the framework for robust understandings of variation. Through analysis of teachers' responses and considerations of statistics education literature, a total of four considerations of or aspects of variability that transcend perspectives, referred to as *elements*, emerged from the data, along with detailed characteristics of indicators for each element. The four elements that elicit reasoning about considerations of or aspects of variation across perspectives are: variational disposition, variability in data for contextual variables, variability and relationships among data and variables, and the effects of sample size on variability. The list of observable indicators was a major tool in the analysis and development of the framework (See Table 6-1, which appears at the end of this section.)

The four elements and reasoning indicative of the elements for each perspective are displayed in Figure 6-3 and discussed in greater detail in succeeding sections. Figure 6-3 also shows how the elements fit into the SOLO Model. For the first cycle of levels of response, attention is on elements from one perspective. The unistructural level (U_1) corresponds with responses focused on a single element, and the multistructural level (M_1) corresponds with responses focused on a number of disconnected elements. The relational level (R_1) corresponds with responses that exhibit integrated reasoning among elements from a given perspective and is indicative of relational reasoning about variation within that perspective. Interpretation of the second cycle of levels remains as initially described in discussion of the SOLO Model, with the presence of elements and integrated reasoning from multiple perspectives as one means to identify integrated reasoning across perspectives to determine robustness of understanding. Table 6-1 contains an elaboration of the table in Figure 6-3 and contains indicators indicative of each type of reasoning.

Elements and Reasoning Indicative of Robust Understanding of Variation			
Perspective / Element	**Design Perspective**	**Data-Centric Perspective**	**Modeling Perspective**
Variational disposition	DP1: Acknowledging the existence of variability and the need for study design	DCP1: Anticipating reasonable variability in data	MP1: Anticipating and allowing for reasonable variability in data when using models
Variability in data for contextual variables	DP2: Using context to consider sources and types of variability to inform study design or to critique study design	DCP2: Describing and measuring variability in data for contextual variables as part of exploratory data analysis	MP2: Identifying the pattern of variability in data or the expected pattern of variability for contextual variables by
Variability and relationships among data and variables	DP3: Controlling variability when designing studies or critiquing the extent to which variability was controlled in studies	DCP3: Exploring controlled and random variability to infer relationships among data and variables	MP3: Modeling controlled or random variability in data, transformed data, or sample statistics
Effects of sample size on variability	DP4: Anticipating the effects of sample size when designing a study or critiquing a study design	DCP4: Examining the effects of sample size through the creation, use, or interpretation of data-based graphical or numerical representations	MP4: Anticipating the effects of sample size on the variability of a sampling distribution

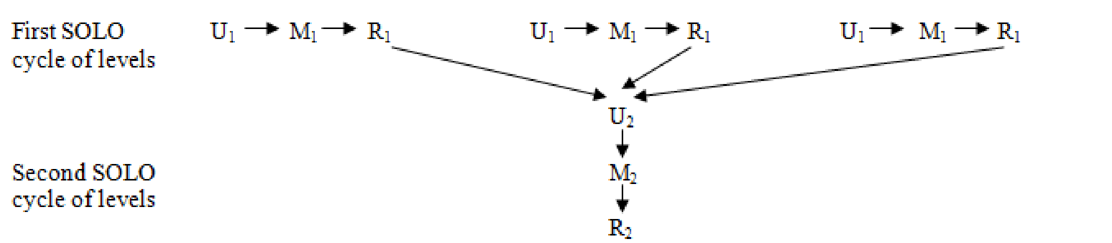

Figure **6-3**: SOLO and Elements and Reasoning Indicative of Robust Understanding of Variation.

Table **6-1**: Indicators of Robust Understandings of Variation.

	Design Perspective	**Data-Centric Perspective**	**Modeling Perspective**
Variational disposition	DP1: Acknowledging the existence of variability and the need for study design in • controlling the effects of variation from extraneous variable(s); • including considerations of variation for variable(s) of interest during data analysis; or • using sample statistics to infer population parameters for the variable(s) of interest	DCP1: Anticipating reasonable variability in data by • considering the context of data; • recognizing that data descriptions should include descriptions or measures of variability (and center); or • recognizing unreasonable variability in data (e.g., that which could result from a data entry error)	MP1: Anticipating and allowing for reasonable variability in data when using models for • making predictions from data or • making inferences from data
Variability in data for contextual variables	DP2: Using context to consider sources and types of variability to inform study design or to critique study design by • considering the nature of variability in data (e.g., measurement variability, natural variability, induced variability, and sampling variability) or • anticipating and identifying potential sources of variability	DCP2: Describing and measuring variability in data for contextual variables as part of exploratory data analysis by • creating, using, interpreting, or fluently moving among various data representations to highlight patterns in variability; • focusing on aggregate or holistic features of data to describe variability in data; or • calculating, using, or interpreting appropriate summary measures for variability in data (e.g., measures of variation such as range, interquartile range, standard deviation for univariate data sets; correlation and coefficient of determination for bivariate data sets)	MP2: Identifying the pattern of variability in data or the expected pattern of variability for contextual variables by • modeling data to explain variability in data or • considering contextual variables in the formulation of appropriate data models or in • modeling data to describe holistic features of data or • considering or creating distribution-free models or simulations to explore contextual variables
Variability and relationships among data and variables	DP3: Controlling variability when designing studies or critiquing the extent to which variability was controlled in studies by • using random assignment or random selection of experimental/observational units to (in theory) equally distribute the effects of uncontrollable or unidentified sources of variability or • using study design to control the effects of extraneous variables (e.g., by incorporating blocking in experimental design or stratifying in sampling designs) to isolate the characteristics of the variable(s) of interest or to isolate systematic variation from random variation	DCP3: Exploring controlled and random variability to infer relationships among data and variables by • using and interpreting patterns of variability in various representations of data; • focusing on aggregate or holistic features of variability in data to make comparisons; • using or interpreting appropriate summary measures of the variability in data to make comparisons (e.g., transformed versus untransformed data); or • examining the variability within and among groups	MP3: Modeling controlled or random variability in data, transformed data, or sample statistics for • making inferences from data (e.g., isolating the signal from the noise for univariate or bivariate sets of data or formally testing for homogeneity in variances) or • assessing the goodness of a model's fit by examining deviations from the model
Effects of sample size on variability	DP4: Anticipating the effects of sample size on the variability of • a sample or • statistics used to characterize a sample (e.g., mean, proportion, median) when designing a study or critiquing a study design	DCP4: Examining the effects of sample size on the variability of • a sample or • statistics used to characterize a sample (e.g., mean, proportion, median) through the creation, use, or interpretation of data-based graphical or numerical representations	MP4: Anticipating the effects of sample size on the variability of a sampling distribution to • model the sampling distribution or • consider significance, practical or statistical significance, of inferences

Elements and Reasoning Indicative of Robust Understandings of Variation

As illustrated, reasoning with a *variational disposition* can be implicit in reasoning from the design, data-centric, or modeling perspectives. Reasoning with a variational disposition from the design perspective occurs when expectation of variation accompanies recognition of a need to consider and implement design strategies for collecting data. When expectation exists outside design, a variational disposition arises in reasoning from the data-centric perspective or from the modeling perspective. Specific examples were discussed in a previous section of Chapter 4, "Pre-Pilot and Pilot Study Analysis" as part of "Data Analysis to Address Research Question One." General descriptions of reasoning about a variational disposition from the three perspectives and a similar set of descriptions of the other elements, which will be discussed below, appear in Figure 6-3.

Reasoning about *variability in data for contextual variables* is a second element of variability for which reasoning can transcend all three perspectives. Reasoning about variation across the three perspectives for this element would include anticipating potential sources and types of variability in study design, exploring contextual data by identifying characteristics of the data through representations and measurements, and fitting models to the data. A general description of reasoning about this element from the three perspectives is provided in Figure 6-3. One example or type of integrated reasoning about variability in data for contextual variables from both data-centric and modeling perspectives is distributional reasoning (e.g., Bakker & Gravemeijer, 2004; Ben-Zvi, Gil, & Apel, 2007). Distributional reasoning, or reasoning about patterns and trends in data from aggregate views based on data models and reasoning about individual values, such as outliers or influential observations, from pointwise views, is reasoning that transcends perspectives.

Reasoning about *variability in examining relationships among data and variables*, a third element that may be implicit in reasoning from any or all of the three perspectives, includes reasoning about strategies to control variability when designing studies or considering study designs, exploring controlled and random variability in data, and modeling controlled or random variability in data.

Reasoning about the *effects of sample size on variability* is reasoning associated with a fourth element. It includes reasoning about the effects of sample size on a sample and the effects of sample size on statistics used to characterize a sample—reasoning that can stem from design, data-centric, or modeling perspectives. General descriptions of reasoning about all four elements from the three perspectives appear in Figure 6-3.

For the purposes of this study, robust understandings of variation are indicated by relational reasoning across the design, data-centric, and modeling perspectives. Teachers' understandings are inferred through the structure of indicators evidenced in their reasoning about variation. The integration implicit in relational reasoning occurs within perspectives and across perspectives. Returning to Figure 6-3, table cells in the figure contain general descriptions of reasoning about an element of variation from a particular perspective. Relational reasoning about variation within a perspective is indicative of relational reasoning about the elements of variability for that perspective in the first cycle of levels, R_1. For example, relational reasoning of variation within the design perspective is indicated by integrated reasoning about variation for the four elements from the design perspective—reasoning that contains the indicators represented by cells DP1, DP2, DP3, and DP4 in integrated form. Relational reasoning of variation across perspectives that includes relational reasoning within all three perspectives, indicative of robust understandings of variation, is integrated reasoning about elements of variation across perspectives in the second level of cycles, R_2. The cells of the table contain descriptions of reasoning, each of which can be realized in multiples ways. Identification of and discussion about the detailed indicators of reasoning about variation are presented by perspective.

Design Perspective

Indicators illustrative of reasoning about variation for each element from the design perspective appear in Table 6-2. Cell headings indicate general reasoning about variation for the given element from the design perspective. For example, “acknowledging the existence of variability and the need for study design” exemplifies general reasoning about variation with a variational disposition from the design perspective. Cell headings in combination with lettered descriptors represent specific indicators of reasoning for each element. For example, indicator DP1(a) is “acknowledging the existence of variability and the need for study design in controlling the effects of variation from extraneous variable(s)” and represents one way to exhibit a variational disposition from the design perspective. This section contains descriptors for each indicator from the design perspective and illustrates each indicator with examples from the data corpus.

Table **6-2**: Indicators of Relational Reasoning About Variation Within the Design Perspective.

Element	Indicators for the Design Perspective
DP1: Variational disposition	Acknowledging the existence of variability and the need for study design in (a) controlling the effects of variation from extraneous variable(s); (b) including considerations of variation for variable(s) of interest during data analysis; or (c) using sample statistics to infer population parameters for the variable(s) of interest
DP2: Variability in data for contextual variables	Using context to consider sources and types of variability to (1) inform study design or to (2) critique study design by (a) considering the nature of variability in data (e.g., measurement variability, natural variability, induced variability, and sampling variability) or (b) anticipating and identifying potential sources of variability
DP3: Variability and relationships among data and variables	Controlling variability when (1) designing studies or (2) critiquing the extent to which variability was controlled in studies by (a) using random assignment or (b) random selection of experimental/observational units to (in theory) equally distribute the effects of uncontrollable or unidentified sources of variability or (c) using study design to control the effects of extraneous variables (e.g., by incorporating blocking in experimental design or stratifying in sampling designs) to isolate the characteristics of the variable(s) of interest or to isolate systematic variation from random variation
DP4: Effects of sample size on variability	Anticipating the effects of sample size on the variability of (a) a sample or (b) statistics used to characterize a sample (e.g., mean, proportion, median) when (1) designing a study or (2) critiquing a study design

Variational Disposition (DP1)

Relational reasoning about variation within the design perspective encompasses reasoning about variation with a *variational disposition*. Acknowledging the existence of variability and the need for study design to deal with variability exemplifies reasoning for a general indicator of reasoning with a variational disposition from the design perspective [DP1]. We see evidence of specific indicators for dealing with variability when acknowledgment accompanies articulation of the need for study design to control the effects of variation from extraneous variable(s) [DP1(a)], include considerations of variation for variable(s) of interest during data analysis [DP1(b)], and use sample statistics to infer population parameters for variable(s) of interest [DP1(c)]. As mentioned previously, general and specific indicators of a variational disposition [DP1] arise from statistics education literature (e.g., Cobb & Moore, 1997; Moore, 1990; Snee, 1990; Wild & Pfannkuch, 1999). One objective of statistics education is to develop students' statistical thinking, which has been described, in part, as dealing with the "omnipresence of variability; statistical problem solving and decision making depend on understanding, explaining, and quantifying the variability" in data (Franklin et al., 2007, p. 6). Acknowledging the existence of variability and articulating the role of study design in understanding, explaining, and quantifying variability are similar to the DP1(a), DP1(c), and DP1(b) indicators of a variational disposition from the design perspective, respectively.

Reasoning that evidences the general indicator of a variational disposition and specific indicators of a variational disposition from the design perspective appears in Dustin's articulation of what he thinks about when he hears the word variation. Key phrases are highlighted in bold font and labeled according to the general and specific indicators illustrated by the phrases.

> What comes to mind immediately is it is, um, the fact that **in any situation**, with a school or ten students, **not everybody acts the same, reads the same, is the same [DP1].** So **there is variation inherent in everything you look at [DP1]**…Um, **it's the reason you take samples larger than 1 [DP1(a)].** Because you want to meas – **you want some measure [DP1(c)]**, so if you have some characteristic, some treatment A and you give it to one person, great, you've got anecdotal evidence, but the question is you got a response, but over **a group of people, what's the average response and how much variation is there [DP1(b)]**?...So you always want a measure of how that variation is affecting, **because it affects everything you do [DP1(a)]**. Um, so, I mean **variation is there, and you've got to deal with it [DP1]**. (Dustin, Content, Lines 2075-2100, [italics and coding added])

Dustin acknowledges the existence of variability at several points while referring to sample size and treatments as elements of design [DP1]. He seems to acknowledge variation in conjunction with study design, and he mentions that in "*any* situation", variation is encountered through a lack of "sameness." He notes that variation is "inherent in *everything*" and something with which one needs to deal. The all-encompassing nature of his language indicates a broad sense of the omnipresence of variation in relation to study design [DP1].

The passage also provides evidence of specific indicators of a variational disposition from the design perspective. Dustin introduces the need for sample sizes larger than one based on differences among individuals and variation affecting "everything you do." His words suggest that he might expect multiple sources of variability (everything) and that he can control the effects of variability associated with individuals by using sample sizes greater than one [DP1(a)]. He alludes to desiring some measurement of a (population) characteristic or the effect of a treatment that results from examining groups larger than one. He seems to be describing a general characteristic or effect, which may indicate reference to a population characteristic or an effect inferred from a sample characteristic or effect [DP1(c)]. Dustin also mentions identifying an "average response" and variation for "a group of people." His allusion to variation for a "group" of people (without mentioning the entire population of people) may indicate that he considers variation in sample data for variable(s) of interest during data analysis [DP1(b)]. Through multiple acknowledgements of the existence of variation and links to design considerations, Dustin's passage provides an image of reasoning indicative of a variational disposition from the design perspective.

Variability in Data for Contextual Variables (DP2)

Context plays an important and unique role in statistics. Cobb and Moore (1997) observe that meaning from data analysis "depends on how the threads of those patterns [observed in data analysis] interweave with the complementary threads of the story line" (Cobb & Moore, 1997, p. 803). Cobb and Moore see design as a core element of statistical thinking and note, "the conclusions from a study depend crucially on how the data were produced" (Cobb & Moore, 1997, p. 807). Taking both of Cobb and Moore's observations into consideration, we see that data production methods need to take context into consideration for analysis of data to have meaning. Indicators for the element of *variability in data for contextual variables* represent some needed contextual considerations for data production. Using context to consider sources and types of variability when designing a statistical study describes one general indicator for reasoning about *variability in data for contextual variables* [DP2(1)]. A second general indicator is considering sources and types of variability in a given context when critiquing statistical studies designed by others [DP2(2)]. Specific indicators include using context to inform study design (or to critique study design) by considering the nature of variability in data [DP2(1a) and DP2(2a)] and by anticipating and identifying potential sources of variability [DP2(1b) and DP2(2b)].

For this study, tasks were couched in contexts assumed to be familiar to most teachers for the purposes of allowing context considerations to arise naturally and to avoid having the researcher cast in an expert role. In particular, the Consultant Task uses the context of standardized assessment. We see an example of considering the nature of variability for a study designed by others in Hudson's reactions to reading the Consultant Task description [DP2(2a)]. He indicates that the observed difference in means for consultants' scores could be "**due to differences in grading practices [or differences in] applying the rubric [DP2(2a)]**," or just "**due to the random selection of the 50 papers for each of the two consultants [DP2(2a)]**"

(Hudson, Content, Lines 68-71). Hudson considers two types of variability: chance variability as it relates to sampling variability and systematic variability from true differences in scoring [DP2(2a)]. He goes on to suggest that if consultants' scores were randomly selected, he would expect some variation and might conclude that any difference in means (grading practices) is due to chance variability. Alternatively, a large difference might indicate differences in grading practices or some type of systematic difference in the way consultants scored exams.

Continuing with Hudson's reactions to the task statement, we see evidence of the second specific indicator of using context by anticipating and identifying potential sources of variability. Hudson notes, "We know that, uh, **individual student papers vary in terms of their quality [DP2(2b)]** and **the ease with which the rubric would be applied [DP2(2b)]**" (Hudson, Content, Lines 119-121). Hudson identifies two potential sources of variability—variability in writing quality and consultants' ease of applying the rubric to each paper [DP2(2b)]—by considering the context of the consultant task. Hudson's reactions to the task description illustrate consideration of the nature of variability expected for the given context and anticipation and identification of potential sources of variability for the context as part of a critique of the study's design. The evidence presented here occurred in response to evaluating the administrators' design for the Consultant Task; parallel reasoning illustrative of considering the nature of and sources of variability, DP2(1a) and DP2(1b), occurs when designing a study.

Variability and Relationships Among Data and Variables (DP3)

The third design element of *variability and relationships among data and variables* refers to design strategies that control variability. General indicators of reasoning about this element appear from controlling variability while designing studies [DP3(1)] or critiquing the extent to which variability is controlled in studies designed by others [DP3(2)]. Wild and Pfannkuch

(1999) define control as "changing the pattern of variation to something more desirable" (p. 236), and specific indictors of reasoning about variability and the relationships among data and variables involve the use of strategies to achieve the goal of control. Two specific indicators involve controlling variability by using random assignment [DP3(1a) and DP3(2a)] or random selection [DP3(1b) and DP3(2b)] of experimental/observational units to (in theory) equally distribute the effects of uncontrollable or unidentified sources of variability among study groups. A third specific indicator includes controlling variability by using study design to control the effects of extraneous variables in order to isolate characteristics of the variable(s) of interest or to isolate systematic variation from random variation [DP3(1c) and DP3(2c)].

Data production methods need to be carefully considered and implemented in order for subsequent analysis of the produced data to have meaning (Cobb & Moore, 1997), as mentioned previously. Development of the articulated set of indicators for variability and relationships among data and variables incorporated related aspects from expository and research literature. For example, Cobb and Moore (1997) emphasize the importance of distinguishing between randomness in experiments and randomness in observational studies. In particular they note the use of random assignment for experimental designs [DP3(1a) and DP3(2a)] and random selection for observational studies [DP3(1b) and DP3(2b)]. Their emphasis on this distinction, and similar emphases expressed by teachers in their reasoning about variation, prompted inclusion of indicators that highlight the distinction.

Results of research also informed and confirmed the development of indicators for this element. For example, Groth (2003) investigated students' thinking with regard to study design and used the SOLO Model in the concrete-symbolic mode to describe students' reasoning. Reasoning was at the relational level if students strategized to produce representative samples and recommended more than one design method to "ensure" representativeness for observational studies. Groth did not focus explicitly on variability in his study, but the idea of a representative sample is related to random sampling that equally distributes the effects of uncontrollable or unidentified sources of variability [DP3(2b)]. This indicator, DP3(2b), also was informed by considering students' reasoning about representative and random samples in other research (Derry, Level, Osana, Jones, & Peterson, 2000; Rubin, Bruce, & Tenney, 1990; Watson & Moritz, 2000).

Someone who can reason relationally about variation within the design perspective recognizes the controlling effects of randomization and considers additional strategies for controlling variation in observational or experimental studies [DP3(1a), DP3(2a), DP3(1b), and DP3(2b)]. That person also recognizes the differences between random assignment and random selection, a confusion observed in prior research (Derry, Levin, Osana, & Jones, 2000). For example, when Blake responds to the Handwriting Task, he provides evidence of recognizing the controlling effect of random assignment [DP3(1a)] and incorporating design strategies for the purposes of controlling variation from the effects of extraneous variables [DP3(1c)].

> The thing I want to measure is simply does the quality of the handwriting affect the score given? Okay? And I guess what **I want to look at is any other, any other thing that might affect the score given, I want to control for [DP3(1c)]** … I guess [what] I was trying to do is create a homogeneous group. Usually you call it **blocking of some sort, um, and so that all of the graders are very similar [DP3(1c)]**…**you can just randomly assign [DP3(1a)]**, and you just kind of hope all these variables that you're thinking about get smoothed over all the groups. (Blake, Content, Lines 1714-1753)

Although Blake never mentions the word variation, when coupled with his earlier talk about using sample size to control variability and his later identification of variability as an underlying theme across interview tasks, it seems reasonable to believe that when Blake talks about control, he means control of variation. In this passage, control seems to be central to his reasoning about study design. Blake associates strategies such as blocking [DP3(1c)] and randomization [DP3(1a)] with control, and he sees control as something that enables him to isolate the systematic effects of handwriting on scores [DP3(1c)]. Blake introduces blocking to control identified sources of variation different from handwriting [DP3(1c)], and he uses randomization to "smooth over," or equally distribute, the effects of extraneous factors [DP3(1a)]. This passage contains evidence of the indicator of using randomization for the purposes of control [DP3(1a)] and the indicator of using additional controlling strategies in the design of an experiment [DP3(1c)] to investigate the relationship between two variables.

Parallel reasoning in reaction to a study designed by others appears in Hudson's reasoning about the Consultant Task. He mentions that he would like to know whether the "50 exam papers that each consultant scored were a **random sample [DP3(2b)]** from all the exam papers that they, that they had available" (Hudson, Content, Lines 55-57). He describes a need for a random sample in this observational study to avoid bias, such as a case in which "**consultant may have systematically scored the exams [initially], um, in a particular way based on their understanding of the rubric**" before fully understanding the rubric [**DP3(2b)**, Hudson, Content, Lines 39-41]. Hudson connects the idea of a representative sample to an unbiased sample, which is reasoning about sample that is more sophisticated than the relational level in the concrete-symbolic mode identified by Watson and Moritz (2000). Further, Hudson suggests that a matched-pairs design would be one way to avoid issues such as "individual student papers [that] vary in terms of their quality and the ease with which the rubric would be applied" (Hudson, Content, Lines 119-121). His matched pairs design effectively controls variation from sources such as differences in student papers and differences in ease of scoring [DP3(2c)]. Through the excerpts from Hudson's interview and from Blake's interview, a picture of reasoning about variability and the relationship among data and variables from the design perspective is conveyed.

Effects of Sample Size on Variability (DP4)

There are two effects of sample size on variability that should be considered in concert with design—effects of sample size on the variability of a sample [DP4(1a) and DP4(2a)] and effects of sample size on the variability of a sampling distribution [DP4(1b) and DP4(2b)]—both when designing studies [DP4(1)] or critiquing the design of studies conducted by others [DP4(2)]. Research reports describe students' difficulties in recognizing the effects of sample size, particularly with regard to sampling distributions (e.g., Chance, delMas, & Garfield, 2004; Fischbein & Schnarch, 1997). Individuals who reason relationally about variation within the design perspective and reason relationally about variation across perspectives consider the effects of sample size on both samples and on sampling distributions.

Everett's response to comparing a sample of size 15 with a sample of size 30 in the Consultant Task illustrates reasoning of anticipating the effects of sample size on the variability of a sample [DP4(2a)]. Everett notes, "**a sample of only 15's not going to give you a great picture of anything [DP4(a)]**" (Everett, Content, Lines 900-901). He reasons that a larger sample gives him more confidence for hypothesizing population characteristics because larger samples more closely resemble the population distribution from which the sample is selected [DP4(a)]. He notes, "**the bigger your sample size, the closer it's going to resemble the population distribution [DP4(a)]**" (Everett, Content, Lines 920-922).

Indicators for anticipating the effects of sample size on the variability of a sample [DP4(a)] and on the variability of statistics used to characterize a sample [DP4(b)] can both be seen in Isaac's comparison of size-15 samples with size-50 samples. At one point, Isaac mentioned that a sample of size 50 is better than a sample of size 15 "in any sampling scheme" (Isaac, Content, Line 1084). Isaac compares samples with different sizes as he considers issues of design.

> Well generally speaking … I would pretty much expect that **I'd get distributions in the same place in the same shape**… **I'm just going to pick up a few in the tails [DP4(a)]** but the, uh, **I'm going to expect that the average – the mean, is going to fluctuate a great deal less in the sample of size 50, um, than in the sample of size 15 [DP4(b)]**. (Isaac, Content, Lines 1117-1147)

In this passage, Isaac associates large sample size with samples more representative of a population [DP4(a)] than small samples. He mentions some characteristics of "representative" samples, which suggests his awareness of the effects of sample size on the variability of a sample [DP4(a)]. Isaac also associates large sample size to the production of means that "fluctuate a great deal less" [DP4(b)]. Less fluctuation in means translates into less variability in statistics used to characterize a sample, indicative of anticipating the effects of sample size on the variability of a statistic [DP4(b)].

Relational Reasoning About Variation Within the Design Perspective

Using the SOLO Model (Biggs & Collis, 1982, 1991), relational reasoning about variation within the design perspective can be manifested as the combination of indicators for each element and integrated reasoning of ideas across the four elements. The examples above convey what reasoning within an element might be. Evidence of reasoning for only one element is evidence of unistructural reasoning and reasoning for multiple elements is evidence of multistructural reasoning. In Blake's reasoning in response to the Handwriting Task, we see an example of integrated reasoning, indicative of relational reasoning. Of the examples already presented, Blake's reasoning about control in response to the Handwriting Task exemplifies integrated reasoning related to the four elements when extended beyond the passage presented here. In his reasoning, Blake expresses an expectation of variation from scorers as he exhibits a variational disposition while describing steps he would take to control some of the effects of variation from scorers [DP1a]. He reasons about variability in data for contextual variables such as scorers [DP2(1a)] and the nature of variability as he considers some of the systematic variation that might be induced from giving different scorers essays with different handwriting qualities [DP2(1b)]. He recommends both blocking [DP3(1c)] and randomization [DP3(1b)] for the purposes of controlling variation. Prior to the passage presented earlier, Blake mentioned the controlling effects of sample size on variability and alluded to those effects in his design for the Handwriting Task [DP4]. Blake reasons about each element from the design perspective in close proximity and in a coordinated and cohesive manner, considering the combined effects of strategies, as he designs his study for the Handwriting Task. Blake's reasoning thus provides an image of relational reasoning about variation within the design perspective.

The examples presented from Dustin, Everett, Hudson, and Isaac do not provide evidence of all four elements of reasoning from the design perspective. The teachers' reasoning beyond those excerpts does, however, provide evidence of integrated reasoning among the four elements. Additionally, because their reasoning with and without prompting extends beyond the given data for a particular task, their responses are consistent with thinking within the formal mode of SOLO (Biggs & Collis, 1982, 1991; Pegg, 2003). Dustin, Everett, Hudson, and Isaac reasoned in ways consistent with relational reasoning about variation within the design perspective in the formal mode.

Data-Centric Perspective

To examine teachers' reasoning from the data-centric perspective, I again consider reasoning with or about variation with regard to the four elements described previously. Table 6-3 displays these four elements and indicators of reasoning illustrative of each element for the data-centric perspective. There are many indicators associated with reasoning about variation from the data-centric perspective, with a large number involving creation, use, or interpretation of data representations. To facilitate discussion of indicators for the data-centric perspective, the numbering system for creation DCPx(1); use DCPx(2); and interpretation DCPx(3), remains constant among elements. Due to the large number of indicators for the data-centric perspective, not every indicator is illustrated with an example from the data corpus, but each indicator is discussed. Relational reasoning within the data-centric perspective is discussed at the end of this section.

Table **6-3**: Indicators of Relational Reasoning About Variation Within the Data-Centric Perspective.

Element	Indicators for the Data-Centric Perspective
DCP1: Variational disposition	Anticipating reasonable variability in data by (a) considering the context of data; (b) recognizing that data descriptions should include descriptions or measures of variability (and center); or (c) recognizing unreasonable variability in data (e.g., that which could result from a data entry error)
DCP2: Variability in data for contextual variables	Describing and measuring variability in data for contextual variables as part of exploratory data analysis by (a) (1) creating, (2) using, (3) interpreting, or (4) fluently moving among various data representations to highlight patterns in variability; (b) focusing on aggregate or holistic features of data to describe variability in data; or (c) (1) calculating, (2) using, or (3) interpreting appropriate summary measures for variability in data (e.g., measures of variation such as range, interquartile range, standard deviation for univariate data sets; correlation and coefficient of determination for bivariate data sets)
DCP3: Variability and relationships among data and variables	Exploring controlled and random variability to infer relationships among data and variables by (a) (2) using and (3) interpreting patterns of variability in various representations of data; (b) focusing on aggregate or holistic features of variability in data to make comparisons; (c) (2) using or (3) interpreting appropriate summary measures of the variability in data to make comparisons (e.g., transformed versus untransformed data); or (d) examining the variability within and among groups
DCP4: Effects of sample size on variability	Examining the effects of sample size on the variability of (a) a sample or (b) statistics used to characterize a sample (e.g., mean, proportion, median) through the (1) creation, (2) use, or (3) interpretation of data-based graphical or numerical representations

Variational Disposition (DCP1)

Anticipating reasonable variability in data is indicative of reasoning with a variational disposition from the data-centric perspective [DCP1]. An awareness that data descriptions should include attention to variation in addition to center is one specific manifestation of a variational disposition [DCP1(b)]. Traditionally, instruction and curricula have focused on centers to the exclusion of spread (Shaughnessy, 1997). Many experienced teachers were educated during the time of this focus and may not have had sufficient opportunities to develop a variational disposition as students. Almost one fourth of the middle school teachers in a study by Canada (2008) alluded only to centers in their comparisons of two sets of data with equal means and unequal standard deviations—data specifically designed to elicit reasoning about variation and distribution. Those who understand variation recognize the insufficiency of reasoning about data from centers alone. They recognize a need not only to consider variation but also to use context to anticipate reasonable variability in data, a second specific indicator of a variational disposition [DPC1(a)]. Considerable research exists to suggest that anticipating reasonable variation is a nontrivial accomplishment. Middle school and high school students have difficulty anticipating reasonable variability in sampling settings (e.g., Reading & Shaughnessy, 2000; Shaughnessy, Canada & Ciancetta, 2003; Shaughnessy, Ciancetta, Best, & Canada, 2004; Shaughnessy, Ciancetta & Canada, 2004) and when estimating and comparing measures of variability in data (e.g., delMas & Liu, 2004; Lann & Falk, 2003; Loosen, Lioen, & Lacante, 1985). Related to the anticipation of reasonable variability is recognition of unreasonable variability in data, a third specific indicator of a variational disposition from the data-centric perspective [DCP1(c)].

For this study, evidence of a variational disposition from the data-centric perspective appears throughout teachers' reasoning about the Consultant Task. Every teacher articulated the importance of knowing something about variability in addition to center [DCP1(b)]. Consider

Blake's reaction to the Consultant Task. He noted that he could use a *t*-test to compare consultants' scores, "except for the fact that **I don't know their standard, you know, their standard deviations [DCP1(b)]**" (Blake, Content, Lines 215-217). Blake's variational disposition is noticeable in his articulated need for standard deviation values. Other teachers evidenced a variational disposition by asking for information about each consultant's distribution of scores, with requests for information about variation inherent in their requests. From these teachers, we see variational dispositions in their recognition that data descriptions should include information about variation to make comparisons or to come to statistics-based conclusions [DCP1(b)].

A second data-based indicator of a variational distribution is recognition of unreasonable variation [DCP1(c)]. Blake's response to the Consultant Task exemplifies this recognition. After being given the standard deviation values for both consultants' scores, Blake does not conduct the *t*-test to which he alluded but instead focuses on the standard deviation value of 20.2. He immediately expresses concern and notes why the value is unreasonable, stating, "a **standard deviation of 20 almost seems absurd [DCP1(c)]**…If you have a 10 as your mean, and you're saying on average…an individual scores 20 away from that, **that's saying that it's not unusual…to see a score of 30**…**out of a possible 15 [DCP1(c)]**" (Blake, Content, Lines 299-323). We see recognition of unreasonable variation in Blake's statement about the absurdity of a standard deviation value of 20.2. We also see justification to support an assertion of absurdity. Blake argues against the value of 20.2 based on the context of having tests scores between 0 and 15, inclusive, in combination with his view of standard deviation as an approximate average deviation. In his invocation of context, we also see evidence of considering context in reasoning about reasonable variability [DCP1(a)].

A variational disposition also can be seen in reasoning that relies on context to consider a reasonable realm of possibilities for data [DCP1(a)]. Consider Isaac's reasoning in response to evaluating the worth of calipers using the full scatterplot of points in the Caliper Task. He begins by considering the standard deviation of the residuals and estimates the value to be 0.08. He inquires about the purpose of the calipers and concludes that the standard residual error of 0.08 is "perfectly fine" for calipers used by middle school students (Isaac, Content, Line 1560). His justification relies on context.

> If you're **selling this to the Charles River**… **measurement company**…an estimated standard error measure of, um, **point oh, eight would not be good enough [DCP1(a)].** But **for middle schools**… presumably there's an increased cost with the greater precision of the measurement…I think what you're primarily concerned with, um, well safety for one thing. Uh, and secondly that kids have an idea what measurement is about and so on. And you're not really concerned that they get it really right…**that's good enough for the kids [DCP1(a)]**. (Isaac, Content, Lines 1563-1578*)*

In this passage, we see that context alone determined reasonableness for the measure. For middle school students, the amount of variation is reasonable given the constraints that are likely to exist in tools used by young adolescents. For a company that needs precise measurements, the amount of variation is unreasonable. Isaac's reasoning gives clear indication of an example of using context to determine whether a measure of variation is reasonable.

Variability in Data for Contextual Variables (DCP2)

Reasoning during exploratory data analysis is indicative of consideration of data for contextual variables. It follows that reasoning specifically about variation in contextual settings indicates reasoning about the element of variability in data for contextual variables from the data-centric perspective [DCP2]. Specific indicators of reasoning consistent with robust understanding for this element include a focus on aggregate or holistic features of data [DCP2(b)] and calculation [DCP2(c1)], use [DCP2(c2)], or interpretation [DCP2(c3)] of appropriate summary measures to reason about data variability. Considerable research has focused on this element of reasoning about variation and informed the identification of indicators for reasoning about this element. Results reported in these studies reveal the difficulties that students (and teachers) experience in transitioning from intuitive and informal reasoning about variation to reasoning with formal measures of variation (e.g., Garfield, delMas, & Chance, 2007; Makar & Confrey, 2005), reasoning about aggregate aspects of and measures of data (e.g., Hammerman & Rubin, 2004; Makar & Confrey, 2005), and reasoning about conceptual rather than procedural aspects of variation (e.g., Silva & Coutinho, 2006, 2008; Sorto, 2004).

Numerous examples of reasoning indicative of this element from the data-centric perspective have been discussed already, particularly with regard to the reasoning of individuals with NSN conceptions of variation. Transnumeration (Wild & Pfannkuch, 1999), in particular, evidences the indicators of creation [DCP2(a1)], use [DCP2(a2)], and interpretation [DCP2(a3)] of various representational forms and fluent movement among representations [DCP2(a4)] to explore variability in data and to select appropriate summary measures for variability [DCP2(c)]. Additionally, upward and downward views of data and distribution (Bakker & Gravemeijer, 2004) are indicative of reasoning about pointwise and, more importantly, aggregate aspects of data [DCP2(b)]. Transnumeration and reasoning about the aggregate of data are not limited to individuals with NSN conceptions. More generally, individuals transnumerate to select the best measures and representations to capture important characteristics of data and variation [DCP2(a) and DCP2(c)], and they use the measures and representations to reason about the aggregate of data [DCP2(b)].

Illustrative of reasoning that touches upon multiple indicators of data-based reasoning about variability from the data-centric perspective are Hudson’s reactions to the size-15 samples from the Consultant Task. He immediately creates representations in the form of boxplots from the lists of scores given to him [DCP2(a1)], and he reasons from both forms of data [DCP2(a4)]. His decision to construct boxplots changes the focus on data from individual values to an aggregate form, which allows reasoning, including reasoning about variation, to focus on the aggregate of data [DCP2(b)]. He reasons that boxplots allow him to identify outliers. He indicates that outliers affect values for the mean and standard deviation more so than for the median and interquartile range, making the median and interquartile range more appropriate for describing consultants' data [DCP2(c1) and DCP2(c3)]. Modified boxplots retain a pointwise nature in their treatment of outlying values, which then allows further interpretation of aggregate features of a majority of the data [DCP2(b)]. Hudson describes the variation in scores by using boxplots to focus on the variation in the upper and lower halves of the scores for both consultants [DCP2(a3)]. He uses representations and measures to describe features and patterns of data and variability for each distribution [DCP2(a2) and DCP2(c2)]. Hudson's reasoning exemplifies describing and measuring variability in data for the contextual variable of scores [DCP2] through his transnumerating among measures, representations, and data to describe aggregate features of the scores.

Variability and Relationships Among Data and Variables (DCP3)

Exploring variability in data for contextual variables often is done as part of or in preparation for investigating relationships among data and variables. Comparisons between data sets, between variables, and between samples and hypothesized populations are common in introductory statistics and are part of reasoning about the element of variability and relationships among data and variables from the data-centric perspective. A general indicator of this element is exploring controlled and random variability to infer relationships among data and variables [DCP3]. Specific indicators include using [DCP3(a1)] and interpreting [DCP3(a2)] patterns of variability from representations; using aggregate features of variability [DCP3(b)]; using [DCP3(c1)] or interpreting [DCP3(c2)] appropriate summary measures; and examining the variability among groups in addition to within groups [DCP3(d)] to make comparisons among data and variables. Group comparisons have provided the setting for several studies to investigate students' and teachers' reasoning about variation (e.g., Makar & Confrey, 2004, 2005; Watson, Callingham & Kelly, 2007). Results from these studies informed the list of initial indicators for reasoning about variability and relationships—indicators that through data analysis became those displayed in Table 6-3. In one of these research studies, Makar and Confrey (2004) note similarities in teachers' reasoning about within-group variability but differences in their reasoning about between-group variability. Although all of their teachers were able to reason about within-group variability, some struggled to make between-group comparisons. Others were able to articulate only intuitions about between-group variation in their comparisons of two distributions. Reasoning about both within-group and between-group variability from data using representations and characteristics of the data is needed for relational reasoning within the data-centric perspective.

Reasoning about the variability between distributions of consultants' scores exemplifies reasoning for the element of variability and relationships among data and variables. In one illustration of this element, Everett responds to the question of whether there is a difference in consultants' scoring by acknowledging two main differences.

> The **center, or the average score for Consultant One definitely seems higher**, uh, **than Consultant Two**. Also, **Consultant One has quite a bit more variability in his scores, uh, than Consultant Two [DCP3(d)]**…The standard deviation's larger, um, and the, uh, the range is obviously larger. (Everett, Content, Lines 306-313)

In this passage, we see comparisons between distributions in terms of the aggregate measures of average, presumably mean; standard deviation; and range [DCP3(c2) and DCP3(d)]. Everett focuses on the aggregate [DCP3(b)] to compare consultants' scores and evidences informal inferential reasoning in suggesting differences in scores.

Reasoning about relationships is not confined to comparisons of univariate distributions. Variability also needs to be considered in reasoning about the relationship between variables in, for example, bivariate data. The same indicators apply to reasoning about bivariate data. Some research exists to suggest that students and teachers tend not to reason about aggregate features of bivariate data (e.g., Hammerman & Rubin, 2004). For example, Brasell and Rowe (1993) found that the students in their study tended to construct and then interpret bivariate data by focusing on specific data points rather than general trends in the data. Students' difficulties in viewing data on a global level were also observed by Ben-Zvi and Arcavi (Ben-Zvi, 2004; Ben-Zvi & Arcavi, 2001). They characterize two understandings of data: local understanding as a focus on an individual, or relatively few, data values within a larger group of data and global understanding that entails recognizing and describing general patterns of data. Ben-Zvi and Arcavi's description of global understanding of data aligns with aggregate reasoning about data. Individuals with robust understandings of variation are able to reason globally about data for both univariate and bivariate distributions and use multiple representations to interpret and use summary measures of variation, particularly for making comparisons and exploring relationships.

Reasoning about relationships between variables can be seen in teachers' reasoning about the Caliper Task. Dustin provides evidence that he considers variation while reasoning about the relationship between variables in the Caliper Task. He interprets the pattern of variability in the data to possibly be quadratic and suggests that a transformation of the data may be needed [DCP3(a3)]. He also suggests creating a residual plot of the data and using an interpretation of the pattern of variability in the residual plot [DCP(a3)] in conjunction with values for the correlation coefficient and the coefficient of determination [DCP(c2) and DCP(c3)] to reason about aggregate features of the relationship between variables [DCP3(b)]. He notes that he would like to "have **a reasonably high correlation, *r* value, and *r* squared [DCP(c2) and DCP(c3)]. Um, and…the residuals are not showing a pattern [DPC3(b)]**" (Dustin, Content, Lines 1428-1447).

Through his reasoning for the Caliper Task, Dustin exemplifies the indicators of reasoning about variability for the relationship between two variables. He focuses on aggregate measures—the correlation coefficient and the coefficient of determination—and on a holistic examination of the residual plot to consider whether there is any pattern of variability in residuals.

Effects of Sample Size on Variability (DCP4)

Considering the effects of sample size on the variability of a sample [DCP4(a)] or sample statistics [DCP4(b)] from the data-centric perspective entails examining the effects of sample size through the creation [DCP4(a1) and DCP4(b1)], use [DCP4(a2) and DCP4(b2)], or interpretation [DCP4(a3) and DCP4(b3)] of data-based graphical or numerical representations of data. A considerable amount of research has investigated students' reasoning about and understanding of sampling distribution, with common difficulties identified as confusion between a sample distribution and a sampling distribution (Saldanha & Thompson, 2002) and confusion between the variation of a sample and the variation of a sampling distribution (Garfield, delMas, & Chance, 2007; Meletiou-Mavrotheris & Lee, 2003). For someone to exhibit relational reasoning within the data-centric perspective, the person must articulate a clear distinction between the two types of distributions and be able to reason about the differences in variation between the two.

Everett illustrates data-based reasoning about sample size in his reasoning to compare a size-15 sample from the Consultant Task with a larger sample.

> **The bigger the sample size you have…the less likely it is that you're going to get one or two unusual values that's going to throw off an average or something like that. You might get a couple extra low values, or a couple extra high values, but the more you have, the better chance you have of those things balancing each other out. [DCP4(a3)]** And looking more like the population. (Everett, Content, Lines 911-918)

In this passage, we see a data-based description of the effects of sample size on samples from an expectation that larger samples are more likely to exhibit distributional characteristics similar to the population [DCP4(a3)]. Everett also demonstrates reasoning indicative of considering the effects of sample size on the variability of statistics used to characterize samples when he alludes to the lesser effect of outliers on averages [DCP4(b)]. After he has the corrected summary measures for the size-50 samples of scores in the Consultant Task, Everett begins to reason about standard error to determine whether there is a significant difference in consultants' scoring. He notes the following.

> With a sample size of 50, **I would be dividing by square root 50, which** is a little over 7, so that **would cut the standard error [DCP4(b1)]** to a little bit less than half and like a quarter… these [sample means] are at least two standard errors, maybe two and a half standard errors apart. (Everett, Content, Lines 661-672)

We see that the actual characteristics of the samples are no longer considered to address the issue of whether a difference exists. Instead, characteristics of the samples, namely the mean and standard deviation, are used to reason about the difference in means in relation to the standard error. Reasoning shifts to consideration of the sampling distribution of differences in means and the effects of sample size on the variability of differences in means [DCP4(b1)].

Relational Reasoning About Variation Within the Data-Centric Perspective

Using SOLO (Biggs & Collis, 1982, 1991), relational reasoning about variation within the data-centric perspective is evidenced by reasoning that includes indicators within each of the four elements and integrated reasoning across the four elements. Of the examples used to illustrate data-based reasoning, Everett's reasoning about size-50 samples in the Consultant Task shows integrated reasoning across the four elements. He provides evidence that he views a standard deviation of 20.2 as unreasonable for scores on an interval from zero to 15, evidence of a variational disposition [DCP1(a) and DCP1(c)]. From the dotplot, he confirms that 20.2 cannot be the standard deviation for Consultant Two's scores. He focuses on holistic features of the data [DCP2(a2) and DCP2(b)] to estimate the actual value. He reasons about variation within each distribution to display evidence for reasoning about data for contextual variables [DCP2(c1)], and he reasons about the variability between the two distributions, which is indicative of reasoning about variability for the third element [DCP3(d)]. Lastly, he reasons about the difference in means in relation to the standard error to provide evidence that he considers the effects of sample size [DCP4]. Everett reasons about each element in close succession and in combination to form a conclusion about differences in scoring. Through his reasoning to determine whether a difference in scoring exists, Everett provides an example of relational reasoning about variation within the data-centric perspective. Examples indicative of relational reasoning exist in the reasoning of the four other teachers, Blake, Dustin, Hudson, and Isaac, whose reasoning was used to exemplify reasoning for a particular indicator.

Modeling Perspective

Relational reasoning about variation within the modeling perspective is evidenced by integrated reasoning among the indicators listed in Table 6-4 for the four elements.

Table **6-4**: Indicators of Relational Reasoning About Variation Within the Modeling Perspective.

Element	Indicators for the Modeling Perspective
MP1: Variational disposition	Anticipating and allowing for reasonable variability in data when using models for (a) making predictions from data or (b) making inferences from data
MP2: Variability in data for contextual variables	Identifying the pattern of variability in data or the expected pattern of variability for contextual variables by (a) modeling univariate data to explain variability in data or (b) considering contextual variables in the formulation of appropriate data models or in (c) modeling data to describe holistic features of data or (d) considering or creating distribution-free models or simulations to explore contextual variables
MP3: Variability and relationships among data and variables	Modeling controlled or random variability in data, transformed data, or sample statistics for (a) making inferences from data (e.g., isolating the signal from the noise for univariate or bivariate sets of data or formally testing for homogeneity in variances) or (b) assessing the goodness of a model's fit by examining deviations from the model
MP4: Effects of sample size on variability	Anticipating the effects of sample size on the variability of a sampling distribution to (a) model the sampling distribution or (b) consider significance, practical or statistical significance, of inferences

Variational Disposition (MP1)

Indicators of a variational disposition from the modeling perspective center on anticipating and allowing for reasonable variability when reasoning from models [MP1]. One specific indicator is realized through allowing for variability when making predictions from data [MP1(a)], such as using interval estimates rather than point estimates for predictions. A second indicator arises from anticipation for and allowance for variability when making inferences from data [MP1(b)]. Both indicators equate with reasoning about the probabilistic rather than deterministic nature of estimating population characteristics from sample characteristics. One description of deterministic thinking is that for which "every result must have an explainable cause;" probabilistic thinking entails thinking about results as being "due to many unexplainable factors coming together, the resulting effect of which is called *chance*" (Scheaffer, 2006, p. 310). Mathematical thinking is often deterministic, whereas statistical thinking is often probabilistic (e.g., Meletiou-Mavrotheris & Stylianou, 2003; Scheaffer, 2006). Researchers have noted the propensity of teachers trained to think mathematically to "handle variability by [arranging data and] finding subsets of the data about which they can make more deterministic claims" (Hammerman & Rubin, 2004, p. 35). These researchers illuminate the difficulties some mathematics teachers have in thinking statistically as they apply their deterministic beliefs about the nature of mathematics to statistics (Meletiou-Mavrotheris & Stylianou, 2003). Inference—particularly standard parametric methods—relies on probability models to make decisions from data. A variational disposition from the modeling perspective seems necessary for invoking inferential methods to make meaningful, probabilistically conditioned conclusions. Individuals with robust understandings of variation reason probabilistically to anticipate and acknowledge variability and exhibit a variational disposition from the modeling perspective.

Examples of reasoning with a variational disposition from the modeling perspective exist for every teacher in this study. The Consultant Task description was designed to reveal whether teachers would display a variational disposition. It was anticipated that individuals with deterministic dispositions would suggest that a difference in consultants' scoring exists based strictly on the observed difference in sample means, whereas those with variational dispositions would not draw definitive conclusions. Dustin's reasoning in response to the administrators' question is typical of reasoning that exemplifies a variational disposition.

> Well, they can't conclude much of anything at this particular point. I mean, **you've got a ten point three versus nine point seven**. **The question is whether or not there's really a difference [MP1]— or, just in the random selection of the exams, um, that difference could occur on a fairly—would occur fairly often [MP1(b)].** (Dustin, Content, Lines 26-37)

In this passage, Dustin draws no conclusion from different mean scores. He notes that the question is really whether the difference in means is one that random sampling could produce with reasonably high probability [MP1(b)], showing that he allows for variation in sample characteristics. We see evidence of a variational disposition through Dustin's probabilistic reasoning in suggesting that the observed difference be compared against differences that "would occur fairly often."

A variational disposition from the modeling perspective can also be seen when reasonable variability is allowed in predictions from data. The Caliper Task was designed to reveal this aspect of a variational disposition by asking for a prediction of caliper measurement for objects with a known length in centimeters. We see evidence of this indicator in Dustin's response to making a prediction from the full scatterplot of points in the Caliper Task. He notes that he would want to project "a reasonable estimate for what it [caliper measurement] would be for 4…**We would estimate**, you know, an **interval within which the measurements should show up [MP1(a)]**, uh, based on that" (Dustin, Content, Lines 1627-1640). In this passage, Dustin identifies a point estimate for the caliper measure that utilizes the best-fit line. We see allowance for variability in the prediction interval he draws around the point estimate [MP1(a)]. (See Figure 6-4.) Dustin's construction of a prediction interval and probabilistic anticipation of variability exemplifies reasoning indicative of a variational disposition from the modeling perspective.

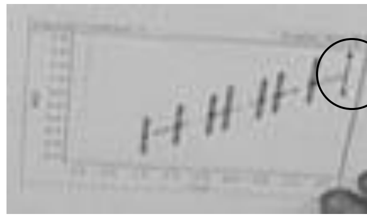

Figure **6-4**: Dustin's Prediction Interval for the Caliper Measure of a Four-Centimeter Object.

Variability in Data for Contextual Variables (MP2)

From the modeling perspective, reasoning about variability in data for contextual variables can be evidenced in multiple ways. Two indicators are identifying a pattern of variability in data or an expected pattern of variability by modeling data to explain variability in univariate distributions [MP2(a)] or by considering contextual variables in the formulation of models for data [MP2(b)]. Research suggests that students in introductory course settings tend to be drawn to normal distributions when reasoning about univariate distributions. At times students reason about data using characteristics of normality even when doing so makes little sense in a given context (delMas, Garfield, Ooms, & Chance, 2007). In part, students' difficulties may stem from solving standard textbook problems using characteristics of normal distributions with little understanding for why normal distributions are appropriate, a difficulty documented by Wilensky (1995, 1997) in work with students experienced in statistics. In their description of building blocks for understanding variability, Garfied and Ben-Zvi (2005) list knowing that the mean and standard deviation "provide useful and specific information about variability" (p. 94) in normal distributions as necessary for deep understanding of variation. When considered in tandem with Wilensky's work, it seems that individuals who exhibit relational reasoning within the modeling perspective should be able to model data with distributions appropriate to a given context and with understanding for why their model is appropriate.

Two additional indicators of a modeling perspective in reasoning about variability for contextual variables are using models to identify the pattern of variability in data in order to describe holistic features of data [MP2(c)] or to consider or create distribution-free models or simulations to explore contextual variables [MP2(d)]. Individuals who have relational reasoning within the modeling perspective at the introductory statistics level should be able to provide evidence of the first three indicators [MP2(a-c)]. Depending upon their learning experiences, they may also provide evidence of the fourth indicator—that of employing the use of simulations or distribution-free models [MP2(d)]. Students and teachers rarely have experiences with nonparametric methods in the setting of introductory statistics (e.g., The College Board, 2004), although recommendations for introducing nonparametric methods in introductory courses are becoming more common (e.g., Cobb, 2005). Additionally, teachers may not have experience with constructing simulated models to aid in inferential reasoning. As a result, it would not be surprising if this fourth indicator were not evidenced by AP Statistics teachers. The indicator was included because several teachers in this study alluded to, described, or recommended nonparametric methods and simulations to establish whether a difference exists in the samples for the Consultant Task.

An example of reasoning with distribution-free methods can be seen in Everett's reasoning for the Consultant Task. Everett introduces the idea of a randomization test in his comparison of size-50 samples, and he describes the process of conducting the test.

> **We could do a sort of a randomization test [MP2(d)]**…We could throw all 100 of these scores into, you know, one set and then split them up into groups of 50, um, find the averages for both groups. See what the difference is. And then do that a bunch of times, over and over and over and over again. And see if a difference of point six is likely to come up just due to the random separation of the scores into two groups. (Everett, Content, Lines 357-377)

Although Everett does not actually create the model for the randomization test he describes, his suggestion provides evidence of exploring variability for the contextual variable of a difference of means through consideration of a distribution-free model [MP2(d)]. Creation and use of an empirical model lies at the heart of his reasoning about this indicator.

We see examples of using models to reason holistically about data in teachers' reasoning about consultants' scores [MP2(c)]. Modeling data with a distribution that has known characteristics facilitates holistic reasoning about data. As an example, consider Everett's description of a possible distribution for Consultant Two's scores. Everett appeals to a normal distribution model to describe what a standard deviation of one would mean for data centered at 10.3 [MP2(c)]. He notes, "**the person who averaged 10 point 3 would be, you know, giving scores mostly between like 8 and 12 [MP2(c)]**" (Everett, Content, Lines 149-151). Everett clarifies that from his experiences with test data, he expects test scores to be normally distributed, which suggests to him that than an appropriate model for test scores is a normal distribution model [MP2(b)]. He notes that, "in a distribution that's approximately normal, um, or at least symmetric like that, **most of the values are going to be within two standard deviations [MP2(c)]**" (Everett, Content, Lines 162-166). We see that Everett introduces a normal probability model to reason about probable scores for Consultant Two's tests. He exhibits holistic reasoning in his reasoning about standard deviation in general terms and in his consideration of the interval over which a majority of scores would fall. He reasons about data by using a model for the data and known characteristics of the model. In this case, the data are hypothetical rather than empirical. We see justification for the claim that a standard deviation of one would result in scores between 8 and 12 from holistic reasoning about data with the described characteristics.

Everett's use of a normal distribution to model test scores exemplifies consideration of contextual variables in the formulation of appropriate models to fit data [MP2(a)]. Beyond univariate data, multiple regression provides a setting for demonstrating this indicator. As was true for distribution-free methods, many introductory courses do not include a focus on multiple regression but instead focus on linear regression (The College Board, 2004). As a result, it was not expected that teachers participating in this study would consider multiple regression. Another setting in which this indicator arises naturally even in introductory settings is reasoning about the coefficient of determination for bivariate data and speculating about additional factors that might explain more variability. In a sense, this reasoning occurs as a prelude to multiple regression. Dustin illustrates this indicator in his reasoning about the coefficient of determination by introducing an example to clarify his description of the coefficient of determination. He describes a set of data consisting of ordered pairs of values for "mileage on a car and the trade-in value of that car" (Dustin, Content, Lines 1500-1501). He sketches a scatterplot of hypothetical data along with a model for the relationship between variables, as shown in Figure 6-5, and then describes how he would think about a car with 60,000 miles that sells for $4500.

> **This linear model pretty much can explain a sixty thousand car**, uh, and let's say I've got all the same make and model here so that I'm comparing apples to apples. That the trade-in value—that **this model will explain thirty-five hundred dollars**…That's what cars that are about sixty thousand miles of that make and model tend to be getting. But **what I'm unable to explain is why is this car getting like…forty-five hundred dollars**. Um, and **it may be that it's like in mint condition [MP2(b)**]. Um, **it's been incredibly well taken-care of. It hasn't been ridden all over the country [MP2(b)]**…**so the model explains why the price should be here** [*$3500*], **but it doesn't explain why the price is at forty-five hundred dollars**. (Dustin, Content, Lines 1516-1534)

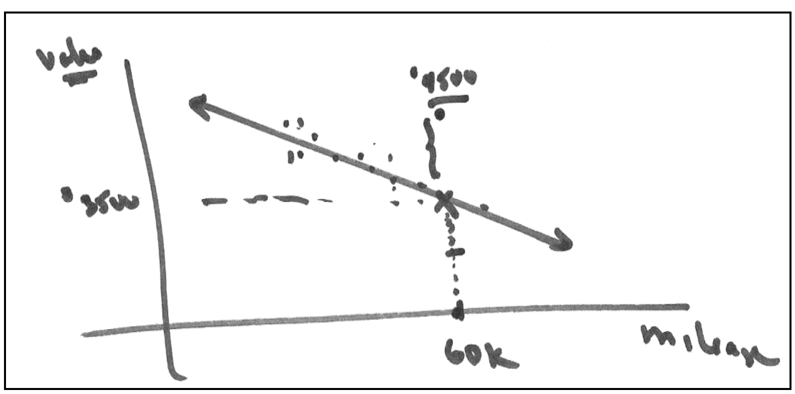

Figure **6-5**: Dustin's Graph to Explain Coefficient of Determination and Sources of Variation.

In this passage, we see clear evidence of considering contextual variables in the formulation of a model for value versus mileage of a car [MP2(b)]. We also see consideration of additional factors that may contribute variability to the value of a car, such as the condition of the car, through focus on a point that deviates from the linear model. In this way, although Dustin does not suggest adding variables to the model, we see reasoning about contextual variables in relation to explained variability and the coefficient of determination and in relation to the model used for this set of data.

Isaac exemplifies the indicator of considering contextual variables in the formulation of appropriate data models as he describes a view of statistics that is consistent with that of many statisticians (Wild & Pfannkuch, 1999). He speaks of statistics and the relationship of models and variability to statistics as he describes his evolving understanding of mathematical models.

> I think that generalized to more and more of statistics as a, as a kind of a modeling activity. Where you—**somehow or other you're constructing a, um, a model of the behavior of the real world** and trying to account for, um, the, uh, for the variability, uh, that you're observing **by appealing to, uh, other variables that, that um, um, that I guess I would call, are being modeled. [MP2(b)]** (Isaac, Context II, Lines 484-491)

Although Isaac does not mention a specific model or specific variables, presumably context is taken into consideration when modeling real world behavior. Isaac notes that variables are considered in the construction of models that explain variability.

A fourth indicator of reasoning about variability in data for contextual variables is modeling univariate data to explain variability [MP2(a)]. As was true for the indicator of modeling data to describe holistic features of data, teachers' reasoning about the Consultant Task provided a large number of examples for this indicator. Dustin's process of estimating the standard deviation of Consultant Two's scores from the dotplot of 49 scores exemplifies modeling data to explain variability. He uses a normal distribution to model Consultant Two's scores to explain the variability in scores. From the model, he estimates a standard deviation value of two by dividing the range of six by four and noting that most of the data would fall within two standard deviations of the mean. We see evidence of using a normal model to estimate and describe aggregate measures of data. This example from Dustin provides evidence of modeling data to explain variability [MP2(a)] and to numerically describe holistic features of data [MP2(c)].

Variability and Relationships Among Data and Variables (MP3)

Reasoning about variability with regard to relationships among data and variables often is indicated by reasoning that includes formal parametric methods, such as performing linear regression to model a bivariate set of data or employing the use of z procedures to model the sampling distribution of sample proportions. Evidence of reasoning about variability and relationships among data and variables stems from using theoretical models to make inferences from data [MP3(a)]. Saldanha and Thompson (2003) suggest that meaningful inferential analysis stems from a multiplicative conception of sample, in which a sample is viewed "as a quasi-proportional mini version of the sampled population, where the 'quasi-proportionality' image emerges in anticipating a bounded variety of outcomes, were one to repeat the sampling process" (p. 266). In reasoning about the mean of a population from the mean of a sample, someone with a multiplicative conception of sample would anticipate the relative unusualness of the sample mean in relation to a distribution of sample means from repeated samplings from the population. In their work with eight inservice secondary mathematics and statistics teachers, Liu and Thompson (2009) found that only one teacher exhibited this multiplicative conception of sample. For someone to reason relationally within the modeling perspective, they need to be able to reason about the variability of a distribution of sample statistics to consider the relative unusualness of a sample.

Reasoning about relationships among data and variables from the modeling perspective is exemplified in many teachers' inferential considerations for the Consultant Task. Consider Blake's description for using a *t*-test to determine differences in the way consultants score exams. He describes modeling the average difference in scores from matched pairs of student exams with a *t* distribution centered at zero and with variability approximated from the sample standard deviation and sample size [MP3(a)]. He describes how the model is used to determine the probability of getting a result like that from the given sample difference. In this way, he reasons about a characteristic from a single mean difference in relation to a distribution of average differences, which allows him to make inferences about the true average difference in scores from the average of the sample differences [MP3(a)]. Blake's reasoning exemplifies using a model for sample statistics to make inferences from data.

A second indicator of variability and relationships among data and variables arises in situations in which the goodness of a model's fit is determined by examining residual plots and evaluating the pattern of variability displayed in the plots. An example of this indicator can be seen in Dustin's reasoning about residual plots during the course of reasoning about the Caliper Task.

> What you want is each difference to essentially be independent of the *x*-value to which it is associated. Because if you end up with something like this [*Dustin sketches a residual plot. See Figure 6-6.*], then what we're saying is **the prediction works pretty well at the beginning, then gets worse and worse and worse, so the value of *x*, I mean the residual is dependent upon the value of *x*. [MP3(b)]** (Dustin, Content, Lines 1380-1387)

From Dustin's words and diagram, we see that Dustin creates a residual plot with what is commonly known as a fanning effect in the residuals. He identifies this pattern as undesirable for a good model fit to data [MP3(b)]. In this way, he presents evidence of assessing goodness-of-fit for the model associated with this residual plot.

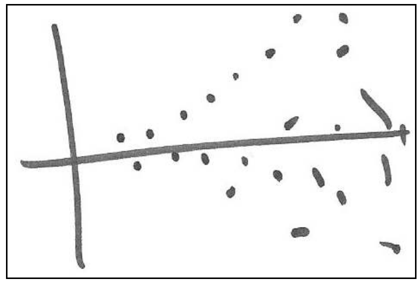

Figure **6-6**: Dustin's Example of a Nonrandom Residual Plot.

Effects of Sample Size on Variability (MP4)

The final element of reasoning from the modeling perspective is reasoning about the effects of sample size on variability [MP4]. One indicator is anticipating the effects of sample size on the variability of a sampling distribution [MP4(a)] and the second is considering the significance, practical or statistical, of inferences based on the variability of a sampling distribution [MP4(b)]. As noted in the description of reasoning from the data-centric perspective, results of research have suggested a commonplace confusion between a sample distribution and a sampling distribution (Saldanha & Thompson, 2002) and between the variability of a sample and the variability of a sampling distribution (Garfield, delMas, & Chance, 2007; Meletiou-Mavrotheris & Lee, 2003)—confusions that inhibit success in reasoning about the element of the effects of sample size on variability. Individuals with robust understandings of variation do not exhibit this confusion between samples and sampling distributions.

We see anticipation of the effects of sample size on the variability of a sampling distribution in Hudson's analysis of the regression output for the Caliper Task. When asked whether any of the values in the output were surprising, Hudson focuses on the test statistic (t = 1586.27) and standard error for the slope (SE = 0.000248). He explains standard error as a way of "accounting for the fact that in repeated runs of this measuring activity, the slopes would themselves vary" (Hudson, Content, Lines 1799-1800). He notes that the standard error is surprisingly small, indicating, "**I'm surprised it couldn't have varied more, but, um, there are a lot of points here [MP4(a)].** You said there were a number stacked on top of each other" (Hudson, Content, Lines 1813-1816). We see that Hudson is surprised by the low value produced for the standard error of the slope coefficient. He assuages his concerns by focusing on the sample size, implying that a large sample size can produce low variation in the standard error [MP4(a)]. In this way, Hudson's illustrates reasoning that considers the effects of sample size on sampling distributions.

The second indicator of anticipating the effects of sample size on a sampling distribution to consider significance is exemplified by Blake when he responds to whether he would use a sample size of 50 in his design for the Consultant Task.

> Let's say that we had **a smaller sample size**, uh, because of, of practical constraints, **that it isn't troubling statistically** – the, the, the mathematics is still gonna be solid.,,it's—**it's harder to find a significant difference, uh, with a smaller sample size**…But **you still need to see a bigger raw difference** because we know that **there's going to be more variability in the smaller sample size [MP4(b)]**. (Blake, Content, Lines 138-152)

From Blake's words, we can see that he considers the effects of sample size on a sampling distribution, noting that there is more variability for small samples [MP4] than for large samples. He evidences considering the effect of this greater variability through acknowledging that a bigger "raw difference" is needed to obtain significant results [MP4(b)].

Relational Reasoning About Variation Within the Modeling Perspective

Indicators of understanding for each of the four elements of reasoning about variation from the modeling perspective are displayed in Table 6-4. Relational reasoning about variation within the modeling perspective is demonstrated by reasoning that includes indicators within each of the four elements and integrated reasoning across the four elements. Dustin provides an example of relational reasoning. In his reaction to the Consultant Task, Dustin reveals a variational disposition by suggesting that the administrators' underlying question requires probabilistic consideration, namely whether the variation between consultants' scores is greater than chance would predict. He goes on to use characteristics of known models to numerically describe both distributions of consultants' scores, evidencing the second element. To respond to the administrators' question, Dustin describes how he would model the sampling distribution of differences in means to conduct a test of significance for inferring the relationship between consultants' scores, and through the formula he describes for calculating a test statistic, he considers the effects of sample size on the variability of a sampling distribution. The data corpus for Blake, Everett, Hudson, and Isaac—the other four teachers discussed in this section—provides evidence of relational reasoning among the four elements from the modeling perspective.

Relational Reasoning in the Second SOLO Level

The preceding sections use a framework based on the SOLO Model to describe relational reasoning about variation within each of the design, data-centric, and modeling perspectives. As shown at the bottom of Figure 6-3, relational reasoning within each perspective (R_1) represents reasoning in the first cycle of levels of the formal mode, which becomes the unistructural level in the second cycle of levels of response (U_2). Individuals who reason at the multistructural level in this second cycle (M_2) exhibit relational reasoning within two or three perspectives without exhibiting evidence of integrated reasoning across all three perspectives. *Robust understanding of variation* is evidenced by relational reasoning about variation across the three perspectives (R_2) in addition to relational reasoning within each of the three perspectives. Relational reasoning across perspectives is demonstrated through integrated reasoning across the three perspectives for one or more elements.

Integrated Reasoning Across Multiple Perspectives for a Single Element

Integrated reasoning across multiple perspectives is first illustrated for a single element using examples from individuals with different conceptions. As an example of integrated reasoning indicative of the element of a variational disposition, consider Hudson's (EDE) reasoning for the Consultant Task. He reasons from the design perspective when he acknowledges the need for proper study design to infer population parameters from sample statistics [DP1(c)] by confirming that consultants' scores were properly selected. Hudson displays a variational disposition from the data-centric perspective when he describes why he needs to know something about variation in addition to center to statistically address the administrators' question [DCP1(b)]. His justification transitions into the modeling perspective when he indicates that the difference in means seems to be small and that a measure of variability would help to determine whether the difference is large or small. Hudson exemplifies consideration of the difference in mean scores in relation to the larger population of scores for each consultant [MP1(a)]. In Hudson's reasoning, we see integrated reasoning through his consideration of the assumptions he needs to make with regard to data collection [DP1(c)] and the data characteristics he needs [DCP1(b)] to determine the significance of the difference in means, evidencing a variational disposition from all three perspectives in a coherent and connected manner.

Integrated reasoning for a single element in a longer temporal period can be seen in Everett's (NSN) reasoning about the effects of sample size in consideration of the Consultant Task. As described in the section on the "Effects of Sample Size on Variability (DCP4)", Everett suggests that larger samples are more likely to exhibit distributional characteristics similar to those of the population than small samples. He includes mention of variability as a distributional characteristic. He provides a data-based description of the effects of sample size on samples [DCP4(a)] after he reasons from the design perspective to anticipate the benefits of larger samples [DP4(a)]. He confirms details about the methods used to select consultants' exams and establishes that a data entry error had been made before he returns to his consideration of sample size. Everett takes the effects of sample size on the variability of statistics used to characterize samples into account [DCP4(b)] in his interpretations of summary measures for the complete size-50 samples. He identifies the standard deviations for each sample but then begins to reason about the standard error [MP4(a)] to contemplate whether there is a significant difference in consultants' scoring [MP4(b)], displaying reasoning from modeling perspective. Everett's reasoning about sample size spans across more than 30 minutes, but throughout, he remains focused on considering the effects of sample size through his anticipation related to study design and his anticipation in establishing significance of differences. Everett illustrates integrated reasoning across the three perspectives for the element of the effects of sample size.

Integrated reasoning across multiple perspectives can be seen not only when considering variability within univariate sets or between univariate sets but also when considering variability in reasoning about the relationship between variables. An example of reasoning about variables in a bivariate setting is found in Isaac's (EEC) reaction to receiving the names of the variables in the Caliper Task. Isaac uses context to identify potential sources of variability [DP2(2b)] for the rightmost two points when he suggests the measurements exceeded the limits of the caliper or that students' hands were too small to measure beyond a certain length. He focuses on aggregate features of the data to describe the variability of the data displayed in the scatterplot [DCP2(b)], and he considers the context and the variables under study to suggest models to fit the data [MP2(b)]. Isaac exemplifies integrated reasoning about variability in data for contextual variables from all three perspectives.

Integrated Reasoning Across Multiple Perspectives for Multiple Elements

Most often, reasoning across perspectives does not occur within a single element but rather spans multiple elements for one or more perspectives. Consider Dustin's (EDE) early considerations for the Consultant Task. He states a need for randomly selected exams and considers the potential effects of randomization on the variability of collected data. His acknowledgement of variation and focus on randomization suggests that he attributes importance to design, indicative of a variational disposition [DP1(a)]. Additionally, he considers the nature of variability within the Consultant Task context [DP2(2a)] and expresses concern about controlling variation from extraneous sources [DP3(2a)]. While continuing to familiarize himself with the problem setting, he indicates that he can only make inferences about consultants' scoring if in addition to data collection methods, he knows something about the standard deviations of each consultants' scores. He reasons from the modeling perspective when he acknowledges variability inherent to making inferences from data [MP1(b)] and a variational disposition from the data-centric perspective when he acknowledges a need for measures of variability [DCP1(b)]. In his reasoning to establish the task setting, Dustin exemplifies integrated reasoning within the design perspective and across perspectives for the element of a variational disposition. It makes sense that in considering a study designed by others, one considers the data collection methods used by researchers and considers variability from multiple perspectives in anticipation of the data analysis needed to respond to researchers' questions.

Integrated reasoning across multiple perspectives for multiple elements can also be seen in reasoning that takes place during analysis of data. Continuing to follow Dustin, we see that when he is given the standard deviations for consultants' scores (and prior to being given the dotplots of consultants' scores), Dustin interprets the variability in both sets of consultants' scores individually [DCP2(a)] and comparatively [DCP3(c)] before examining the variation between distributions [DCP3(d)]. To do so, he constructs a model to represent the interval of scores within a multiple of the standard deviation from the mean for each consultants' sample scores [MP2(b)]. He uses summary measures [DCP3(b)] and his model [MP2(c)] to reason holistically about both distributions. (See Figure 6-7 for Dustin's sketch.) He uses his model in conjunction with the summary values to suggest that the amount of variability between distributions and within the second consultant's scores would prohibit a conclusion of significant differences in terms of means but not in terms of variances [MP3(a)]. In his reasoning, Dustin exemplifies integrated reasoning for two elements (variability in data for contextual variables and variability and relationships among data and variables) from two perspectives (data-centric and modeling).

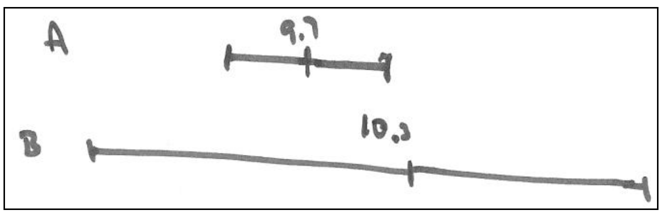

Figure **6-7**: Dustin's Model to Compare Consultants' Scores From Means and Standard Deviations.

Robust Understandings of Variation

As the preceding discussion might suggest, integrated reasoning across perspectives can occur through different combinations of indicators, elements, and perspectives. For the purposes of this study, robust understandings of variation are indicated from evidence of relational reasoning about variation within each perspective in the first cycle of levels and relational reasoning about variation across all three perspectives in the second cycle of levels. Evidence of robust understandings has four required components: (1) evidence of reasoning matching each general indicator for each element of each perspective, or evidence that falls into every cell of

Table 6-1; (2) evidence of relational reasoning within each of the three perspectives; (3) evidence of integrated reasoning across the three perspectives for at least one element and integrated reasoning across at least two perspectives for each of the four elements; and (4) the absence of consistently faulty reasoning, unsubstantiated claims, and missing claims appropriate for the task under consideration. For general indicators with multiple specific indicators, there needs to be evidence of at least one specific indicator. For example, with the indicators of creating [DCP2(a1)], using [DCP2(a2)], interpreting [DCP2(a3)], or fluently moving among [DCP2(a4)] various data representations to highlight patterns in variability for reasoning about variability in data for contextual variables [DCP2(a)], evidence is interpreted to be at least one specific indicator from DCP2(a1) to DCP2(a4) appropriate to the task under consideration. For those teachers identified as having robust understandings of variation, the teacher's data showed no evidence of missing indicators appropriate to a task or faulty reasoning about variation.

Conceptions and Robust Understanding

Relationship Between Conceptions and Robust Understanding

Achieving robust understanding of variation is not dependent upon viewing variation as Expected but Explainable and Controllable (EEC), Noise in Signal and Noise (NSN), or Expectation and Deviation from Expectation (EDE). At least one teacher with each type of conception exhibited integrated reasoning of indicators for each element of each perspective and reasoned in ways consistent with robust understandings of variation. Although these teachers exhibited relational reasoning across all three perspectives in the second level of SOLO, the emphases of their reasoning differed across conceptions. For example, an individual with robust understandings of variation who views variation as EEC exhibits relational reasoning about variation across all three perspectives, with the design perspective more dominant in his or her reasoning than the data-centric or modeling perspectives. Figure 6-8 shows a modified diagram from Figure 6-2, which highlights more prominent reasoning from the design perspective. What this diagram fails to show, however, is how that dominance looks in the second cycle of levels for which reasoning is integrated across perspectives.

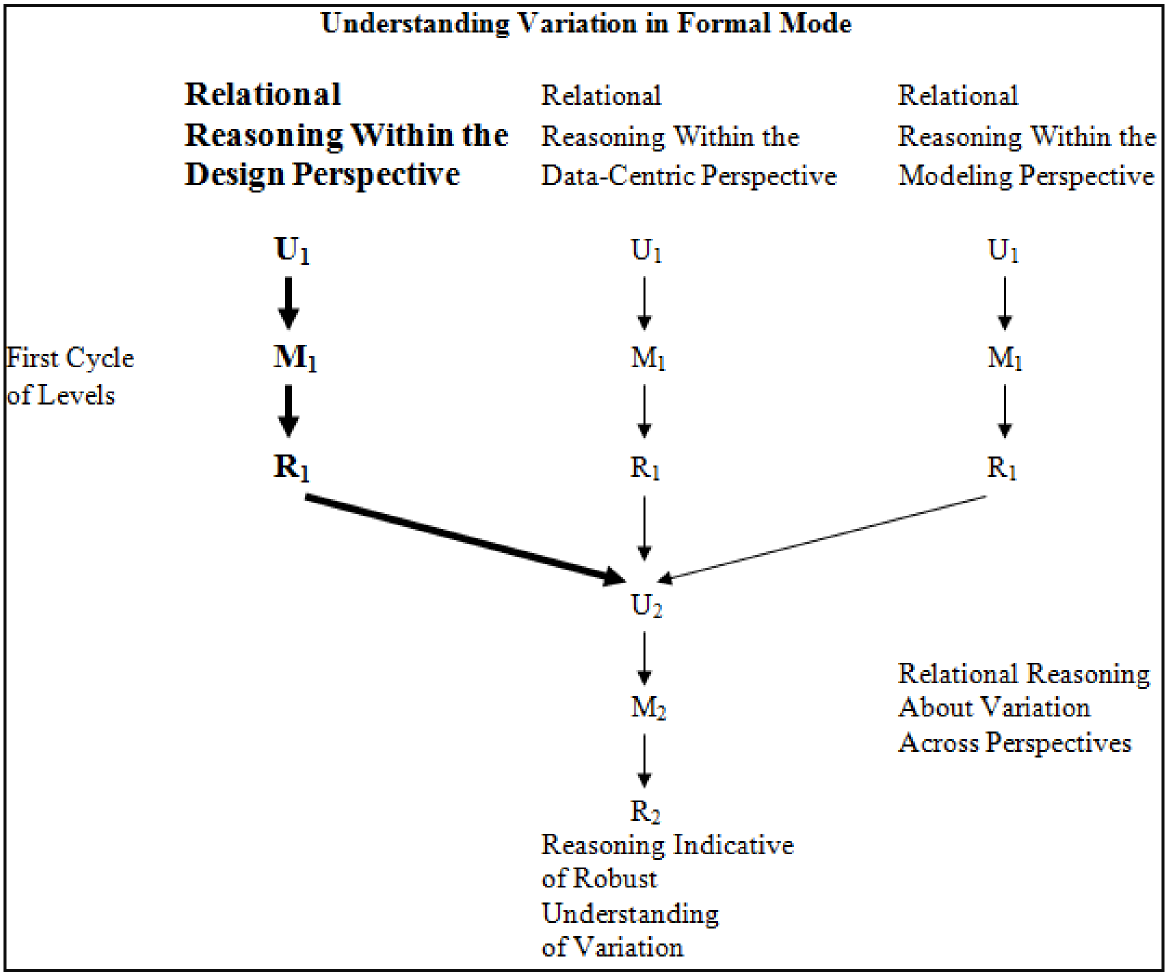

Figure **6-8**: The SOLO Model and Robust Understandings of Variation for EEC Conceptions.

Robust Understanding, Conceptions, and Teachers in This Study

The teachers participating in this study exhibited sophisticated reasoning about variation. With a few minor exceptions, all of the teachers in this study reasoned in the formal mode. They also reasoned at the highest level of the hierarchies from previous research that focused on students' descriptions and considerations of variation in particular contexts or in a repeated sampling environment (e.g., Reading, 2004; Reading & Reid, 2004; Reading & Shaughnessy, 2004; Reid & Reading, 2006; Watson & Kelly, 2004a), and a majority of the teachers reasoned about variation in the formal mode in ways consistent with the highest levels of hierarchies focused on students' considerations of variation in multiple contexts and environments (Reid & Reading, 2008; Watson, Kelly, Callingham, & Shaughnessy, 2003). As use of the word "majority" suggests, however, not every teacher who participated in this study exhibited reasoning indicative of the highest levels identified by those researchers. A majority of teachers in this study did not exhibit reasoning indicative of robust understanding of variation as defined in this study. In part, time limitations of the content interviews did not allow enough time for every teacher to respond to every aspect of every task. As a collection, the tasks were intended to elicit reasoning about different aspects of variation and some tasks focused more on eliciting reasoning from a particular perspective. Because the interview tasks were designed to elicit reasoning about different aspects of variation, some teachers were deprived of opportunities to exhibit reasoning for some of the large number of indicators shown in Table 6-1. There also were teachers who did not display indicators appropriate to the task under consideration or displayed faulty reasoning with regard to some elements and indicators. Of the 16 teachers participating in this study, five teachers exhibited evidence of robust understandings of variation. For these teachers, the data showed no evidence of faulty reasoning about variation.

For examples of relational reasoning about variation across all three perspectives for the five teachers identified with robust understandings, consider the examples of their reasoning presented in earlier sections. Table 6-5 displays the sections in which evidence was presented for each of the five teachers, with the lists within each cell appearing in the order in which the evidence was presented. Although reasoning about multiple elements or from multiple perspectives was not part of the discussions of the examples as they were presented in Chapter 5, many of the discussions implicitly present evidence of integrated reasoning. Additionally, at least one example of integrated reasoning for each of the five teachers was described in this chapter. Other examples of integrated reasoning occurred throughout their content interviews.

Table **6-5**: Illustration of Robust Understanding Categories Using Only Examples From Teachers With Robust Understandings.

	Design Perspective	Data-Centric Perspective	Modeling Perspective	Integrated Reasoning
Variational disposition	• Isaac, EEC data-centric • Hudson, EDE data-centric • Dustin, DP1 • Hudson, DP2	• Isaac, EEC data-centric • Everett, NSN center as signal • Blake, EDE dev of statistics • Dustin, EDE data-centric • Hudson, EDE data-centric • Blake, DCP1 • Isaac, DCP1	• Isaac, EEC data-centric • Dustin, MP1	• Hudson, R_2
Variability in data for contextual variables	• Isaac, EEC conception • Isaac, EEC explainable • Isaac, EEC data-centric • Everett, NSN design • Blake, EDE design • Hudson, DP2	• Isaac, EEC controllable • Isaac, EEC data-centric • Everett, NSN center as signal • Dustin, EDE dev from patterns • Hudson, DCP2	• Isaac, EEC explainable • Everett, NSN center as signal • Everett, NSN patterns as signal • Everett, NSN modeling • Dustin, EDE dev from patterns • Everett, MP2 • Dustin, MP2 • Isaac, MP2	• Isaac, R_2
Variability and relationships among data and variables	• Isaac, EEC explainable • Isaac, EEC controllable • Everett, NSN design • Hudson, EDE dev from relationships • Blake, EDE dev from relationships • Blake EDE design • Dustin EDE design • Blake, DP3 • Hudson, DP3	• Isaac, EEC explainable • Isaac, EEC modeling • Everett, NSN patterns as signal • Everett, NSN rel between variables • Everett, NSN modeling • Blake, EDE dev of statistics • Blake, EDE dev from relationships • Hudson, EDE dev from relationships • Everett, DCP3 • Dustin, DCP3	• Isaac, EEC explainable • Isaac, EEC data-centric • Isaac, EEC modeling • Everett, NSN patterns as signal • Everett, NSN rel between variables • Everett, NSN modeling • Blake, EDE dev of statistics • Hudson, EDE dev of statistics • Blake, EDE dev from relationships • Hudson, EDE dev from relationships • Blake, MP3 • Dustin, MP3	• Dustin, R_2
Effects of sample size on variability	• Isaac, EEC controllable • Everett, NSN design • Blake, EDE design • Everett, DP4 • Isaac, DP4	• Everett, DCP4	• Isaac, EEC controllable • Blake EDE design • Hudson, MP4 • Blake, MP4	• Everett, R_2
Integrated reasoning	• Blake, DP	• Everett, DCP	• Dustin, MP	

From Table 6-5 we see that the elements of variability in data for contextual variables and variability and relationships among data and variables are most common in the examples presented from the data corpus. This distribution of examples seems to be reasonable in light of the fact that these elements have more indicators than the remaining two elements and thus require more examples for illustration. Evidence among perspectives is distributed similarly. From the criteria established in this study, Blake, Dustin, Everett, Hudson, and Isaac provide evidence of robust understandings of variation. Their learning experiences are analyzed in Chapter 7.

Chapter 7

Influential Factors for Learning

In answer to the second research question, this chapter focuses on factors that may have influenced the statistical learning of the five teachers in this study who exhibited robust understandings of variation. In particular, I examine the nature of activities and actions that, in the teachers' perceptions, contributed to their current understandings of statistical variation. I use transformation theory to frame how these experiences may have contributed to learning that deepened their statistical knowledge. To begin, I examine the experiences that prompted dilemmas for these five teachers, the resolution of which allowed them to construct statistical knowledge. In subsequent sections, I expound on personal and environmental influences for their learning in addition to characteristics of their learning methods. Throughout the chapter, I present representative examples of the larger body of examples for each claim.

Triggers

Through events related to their teaching of statistics, each of the five teachers experienced one or more "triggers" that prompted self-awareness of limitations in their knowledge of statistics. By *trigger* I mean what Marsick and Watkins (2001) refer to as "an internal or external stimulus that signals dissatisfaction with current ways of thinking or being" (p. 29). The trigger stimulates one of two types of dilemmas. An *epochal dilemma* is resolved through transformation of a meaning perspective (Mezirow, 2000), or changes in the interwoven assumptions and expectations through which the world is viewed to make sense of current experiences (Cranton, 2006). An *incremental dilemma* is resolved through the creation, enhancement, or transformation of a meaning scheme (Mezirow, 2000) that consists of a specific expectation, knowledge, belief, attitude, or feeling (Mezirow, 1991) used to interpret everyday experiences (Cranton, 2006).

An epochal dilemma can be a *disorienting dilemma* (Mezirow, 1990), terminology which has become synonymous with a dramatic crisis that provokes strong, painful emotions such as anger, shame, intimidation, or fear (e.g., Baumgarner, 2001; Erickson, 2000; King, 2000; Taylor, 1997). Epochal dilemmas also may induce positive emotions through perceived affordances for learning (Erickson, 2007). To accommodate dilemmas that are accompanied by positive or negative emotion, researchers have either broadened their definition of "disorienting" (Taylor, 1987) or applied different adjectives that better capture the nature of dilemmas (Erickson, 2007). To describe other types of epochal dilemmas, I will use the terminology of *opportunity dilemma* for an epochal dilemma that is viewed as an opportunity for learning (Erickson, 2007) and *touchstone dilemma* for an epochal dilemma that is revisited often for meaningful resolution (Erickson, 2007). These different types of dilemmas are not necessarily disjoint. Although none of the teachers recounted experiences indicative of a disorienting dilemma, Isaac experienced a dilemma characteristic of both opportunistic and touchstone dilemmas.

In contrast to epochal dilemmas, incremental dilemmas are less intense emotionally and may induce questioning of assumptions related to meaning schemes. Resolution to this second type of dilemma produces a new, enhanced, or transformed meaning scheme. A series of resolutions to incremental dilemmas for meaning schemes within a meaning perspective can result in a transformed meaning perspective (Taylor, 2000). Because resolution of this second type of dilemma contributes to a perspective change that is incremental and cumulative, I refer to this second type of dilemma as an *incremental dilemma*.

The five teachers in this study experienced transformational learning in the form of a transformed meaning perspective for statistics, of which variation is a part. As a result, variation may not be central to their dilemmas, but resolution of the dilemmas may result in changed understandings of variation. Table 7-1 contains a brief description of each type of dilemma, along with the triggers experienced by teachers in this study. The nature of these types of triggers is discussed in greater detail in succeeding sections.

Table **7-1**: Types of Dilemmas and Triggers.

Dilemma Type		Characteristics	Triggers
Epochal dilemma	Disorienting dilemma	Dramatic crisis that provokes strong, painful emotion while reflecting critically on assumptions and beliefs	Learning experience or event that creates dissonance with assumptions and beliefs about the field of statistics Resolution results in a transformed meaning perspective
Epochal dilemma	Opportunity dilemma	Dramatic event viewed as opportunity for learning, with the individual's search for answers satisfying intellectual needs	Learning experience or event that stimulates intellectual needs which are viewed as opportunities for learning about the field of statistics Resolution results in a transformed meaning perspective
Epochal dilemma	Touchstone dilemma	Dramatic event that is revisited often and reformed to reflect new experiences and learning for resolution	Learning experience or event that stimulated thought about the field of statistics and that is revisited often during resolution Resolution results in a transformed meaning perspective
Incremental dilemma		Event that provokes discomfort in relation to a particular assumption or belief Resolution may result in a transformed meaning scheme. A series of resolutions to incremental dilemmas may result in a transformed meaning perspective	Learning experience that suggests misunderstanding with respect to a particular statistical topic or concept Exposure to previously unknown subtleties in content assumed to be known Attempts to make connections across content topics, particularly during lesson planning Active engagement in activities designed to elicit particular understandings Encountering new content/ terminology that conflicts with current understandings Considering subtleties in students' responses or in response to student inquiries

Triggers of Epochal Dilemmas

Of the five teachers who exhibited robust understandings of variation, only Isaac describes a trigger consistent with producing an epochal dilemma. This trigger—the statistical ideas encountered during a four-week institute focused on the learning and teaching of statistics—precipitated actions resulting in increased statistical knowledge for Isaac. In particular, the content of the institute triggered a dilemma that was both touchstone and opportunistic.

Prior to the institute, Isaac taught a statistics course that was "basically…the college course that I had had" (Isaac, Context I, Line 406). The institute stimulated changes in the way Isaac thought about statistics, which resulted in Isaac making changes to the content and pedagogy of his statistics course reflective of his changing perspective. Isaac notes, "content-wise that was where I learned about EDA [Exploratory Data Analysis]" (Isaac, Positive CI). "We spent a week doing EDA stuff. And so I went and I—I went back and said well I'm going to put this in my class. So I put that in, [and] learned, uh, a lot more about it" (Isaac Context I, Lines 418-421). Isaac's experience at the institute triggered an opportunistic dilemma in the sense that he saw possibilities in the ideas he encountered at the institute for his learning (and teaching).

Isaac resolved the dilemma by incorporating EDA in his statistics class, including elements related to statistical variation, and by learning additional content related to EDA. Professionally, the institute served as, "the genesis of all my subsequent work in the area of statistics education, and AP Statistics in particular" (Isaac, Positive CI). As Isaac continued to learn statistics through his experiences, he often revisited ideas that originated from this institute. The institute triggered a touchstone dilemma in that it prompted Isaac to think about statistics as the science of exploring data to answer questions of interest, an image of the subject that he revisited often throughout his varied experiences, as opposed to a set of procedures performed on data in an established and predetermined manner. In addition to experiencing the trigger of an epochal dilemma, Isaac encountered triggers of incremental dilemmas throughout his continued learning experiences in statistics.

Triggers of Incremental Dilemmas

Unlike triggers for the relatively rare epochal dilemma seen in this study, triggers for incremental dilemmas were commonly encountered by Blake, Dustin, Everett, Hudson, and Isaac. Most of the dilemmas articulated by the five teachers were triggered while engaging in workshop or conference activities, engaging in dialogue about the learning and teaching of statistics, planning to teach AP Statistics, or teaching statistics.

Triggers From Workshop or Conference Activities

A major purpose of professional development is promotion of teacher learning, including learning of subject-matter content. In alignment with this purpose, all five teachers in this study encountered dilemmas that were triggered from engagement in workshop or conference activities. The dilemmas arose when teachers were active listeners to presentations or participants in statistical activities.

The Conference Board of the Mathematical Sciences (2001) recommends that prospective secondary teachers take two courses in probability and statistics: “a calculus-based survey course in probability and statistics and a course in data analysis” (p. 137). Of the five teachers, only Hudson completed an exploratory statistics course focused on exploring data in context. As a result, teachers’ experiences with data analysis were largely limited to the introductory topics they teach. Data analysis content beyond that encountered in AP Statistics tended to trigger incremental dilemmas. Consider Everett's reaction to a conference presentation he attended.

> The topic of [the] talk was Linear Models and the connections between Multiple Regression and ANOVA…Based on this experience, I realized that I needed to fill in lots of missing steps if I was to ever understand the material. (Everett, Negative CI)

Everett left this session with a dilemma triggered from an awareness of limitations in his meaning scheme for regression. For Everett, this session, and others like it, stimulated action toward filling in the "missing steps" to resolve the dilemma and form deepened understandings of regression and the connections between variation and regression. These changes to his meaning schemes contributed incrementally to his transformed view of statistics.

Some incremental dilemmas were triggered from workshops or conference sessions that incorporated participants' engagement with exploratory activities. As an example, consider Hudson's engagement in a simulation activity that resulted in greater insight into how randomization allows him to determine whether an experimental result deviates from expectation by more than chance would predict. The activity incorporated two types of simulations—physical simulation with cards and computer simulation with a java applet. Hudson describes the effects of his engagement in the activity.

> By combining results from all those with decks of cards, [Presenter] produced a preliminary estimate of the likelihood that the experimental result could have been "just due to chance." He then presented a computer simulation [java applet] that was designed to mimic the physical card shuffling and dealing simulation. Using the computer simulation … [Presenter] developed another estimate for the likelihood of obtaining results as "unusual" as those that occurred in the actual experiment "just by chance."…To this point in my career, I don't think I fully grasped the meaning of "just by chance" in an experimental setting…This revelation on that evening in June really excited me! (Hudson, Positive CI)

Hudson suggests that he gained insight into affordances that arise from random assignment in experimental design—affordances that he did not "fully grasp" prior to this activity. Engagement with the activity triggered a dilemma through awareness of his gap in understanding. Resolution of the dilemma altered Hudson's meaning scheme for random assignment by illustrating for him how random assignment allows him to determine chance probability under the conditions of a null hypothesis.

Triggers From Dialogue With Colleagues and Statisticians

Learning occurred in response not only to incremental dilemmas triggered from experiences designed to facilitate learning but also from dilemmas triggered from statistical dialogue with colleagues and statisticians. Blake describes how interactions at the AP Readings often trigger dilemmas based on shortcomings in his understandings, particularly when the interactions focus on scoring rubrics. Discussion of responses for which students receive or do not receive full credit surfaces new issues that become the object of attention and conversation. During the course of conversation, Blake describes how he "runs into conflict where… either my knowledge is wrong, and probably…hardly ever wrong per se, but incomplete, okay, not fine tuned. And so then as I hear the discussion I'm able to fine tune my own thinking" (Blake, Context I, Lines 1511-1581). These interactions and conversations trigger dilemmas that Blake seeks to resolve. Although he suggests quick resolution to dilemmas, which may be indicative of learning through meaning schemes or learning new meaning schemes, he recognizes that not all dilemmas meet with immediate resolution: "It creates an understanding or a question that has to be resolved, and it sometimes takes years" (Blake, Context II, Lines 878-897). Blake suggests the AP Reading is a venue that regularly triggers subsequent learning for him.

Triggers From Classroom Teaching

A venue not designed for the primary purpose of enhancing teachers' learning but that regularly triggers incremental dilemmas that lead to learning is teaching. Each teacher describes interactions with students that triggered incremental dilemmas and provided the impetus for learning through students' unique observations and questions. Dustin describes dilemmas triggered by student questions, his lack of confidence "that I totally understand the intricacies" (Dustin, Context I, Lines 1785-1786), and the steps he takes to resolve his dilemmas and to answer students' questions. Blake indicates that as his students began to ask questions of "why" in his class, he sought answers for his students and for himself—answers to questions such as why use a measure of variation calculated from squared deviations (standard deviation) in place of a more intuitive measure of variation calculated from absolute deviations (average absolute deviation). His resolutions to answer questions of "why" may come about through transformations of meaning schemes. He also notes how his maturing understandings may evoke deeper questions from his students—questions that produce what Blake calls a "conflict of thought," a dilemma. He notes, "maybe as I matured…then kids are satisfied…through level one, and now they're asking questions on level two, where they never asked questions on level two before because I wasn't even satisfying them on level one" (Blake, Context I, Lines 1074-1092). Blake suggests that as his knowledge of statistics deepens over the years, the knowledge he observes among students also seems to deepen, or perhaps he becomes more aware of the depth of students' knowledge. In response to students' triggering of "conflicts of thought," Blake seeks resolution.

Triggers From Planning for AP Statistics Instruction

Many triggers arose for these teachers during the course of planning for instruction, particularly when planning to teach the AP Statistics course for the first time. Even though several taught probability and data analysis courses before the advent of AP Statistics, the AP course outline (The College Board, 1996) contained content not previously encountered. In particular, each of the five teachers commented that study design was a required area but one for which they had insufficient previous experiences to teach comfortably. Planning to teach the design of observational studies and experiments triggered numerous incremental dilemmas for them. Dustin attributes his early struggles with the course to the inclusion of design and his planning to teach design. “I think that’s why the first couple years I struggled, because I really had not had any formal design. So I was… trying to figure out what I should be doing in AP stat, in the design area” (Dustin, Context I, Lines 661-670).

The main purpose of design is controlling variability to deal with the uncertainty that arises from the omnipresence of variability—a largely nonmathematical endeavor (e.g., Groth, 2007). Dustin indicates a difference in the way he needs to think about statistics from the way he thinks about mathematics, noting that statistics “was something of an art form” (Dustin, Context I Lines 146-147). He contrasts statistics with mathematics, noting the following.

> Most of my courses, being math courses, it was like well there’s always a formula lying around. Just apply the formula and you’re done. But here, you actually could step back and say well, if I look at it this way, this is what I see. If I look at it this way, this is what I see. (Dustin, Context I, Lines 124-130)

For these teachers who each took numerous mathematics courses, the artistic aspects of design may explain why the area of design is the only broad content area in which their learning and planning for instruction triggered disorientation. Blake describes design as “one of those places where…you may learn some facts and then you understand why the facts later” (Blake, Context I, Lines 1013-1073).

Some design triggers stemmed from issues related to language. In particular, Dustin describes how he and another teacher would consult textbooks to resolve issues, and how differences in terminology, particularly in the area of design, would trigger incremental dilemmas that required learning through meaning schemes or learning new meaning schemes to resolve. In large part, Dustin describes confusion stemming from different terms for what appeared to be the same concept, such as treatments and factors or lurking variables and extraneous variables.

> We would go find a textbook and try to figure out what's going on. And because of the lack of a codified language, if you will, treatments, factors, all these issues, experimental units, lurking versus extraneous…That was one of those points where, you know, like so what is this lurking versus—are they the same? And we'd go and look and, and so we spent a lot of time just trying to figure stuff out. (Dustin, Context I, Lines 746-757)

Dustin suggests that different authors' terminology for seemingly identical concepts triggered dilemmas that required "a lot of time" to resolve.

Beyond different terminology for the same concept, subtleties in students' solutions to open-ended problems also created triggers. Everett credits the AP Statistics free response questions for alerting him to some wording subtleties. He often "solved" the free response questions from each year's examination in preparation for post-examination lessons. He describes how incremental dilemmas were triggered at the AP Reading when he discovered that the "solution" he presented to his students would not have received full credit. Of the AP Reading, he also notes that distinguishing between students' "essentially correct" and "substantially correct" responses—two holistic designations given to student response at the AP Reading for responses that have at most a trivial error or omission and responses that may be missing one essential aspect, respectively—illuminates, "the subtleties of, of what words mean and the ideas. I hadn't gotten the big picture in all those things yet" (Everett, Context I, Lines 49-59). He notes that the free response questions, particularly those focused on design, trigger incremental dilemmas that required additional learning about design—learning through his meaning schemes related to design—to resolve.

Reactions to Triggers

When these five teachers encounter triggers that stimulate an awareness of an epochal or incremental dilemma, they act to resolve their dilemmas. Dustin, for example, notes how he consults "another source [to] see what they have to say" (Dustin, Context I, Lines 1766-1767) or pages through his notes, which consist of folders that contain articles and items related to particular concepts. When he is unable to resolve his dilemmas after consulting the references readily available to him, he is apt to search the archives from the electronic discussion group or to consult secondary statistics teachers and statisticians to ask for help in resolving his dilemmas. When Dustin and the other four teachers experience dilemmas, they devise plans of action to resolve their dilemmas in the most expeditious manner possible.

Recursive Nature of Learning

These five teachers are open in acknowledging that they continue to encounter triggers that reveal previously unrecognized holes in their knowledge—triggers that create mostly incremental dilemmas for them. For example, when Dustin started teaching AP Statistics, his triggers stemmed mainly from the area of design. As he began to learn more statistics through attending professional development, teaching statistics at the secondary level, and conducting professional development, he began "to realize how little I truly know about the basics" (Dustin, EHC). Triggers arose in areas beyond design. Like some of the adult educators in King's (2004) study, deepened understanding in one domain leads to recognition of additional areas for study.

As one of her teachers describes, “I realize that not only have I learned a lot, but there is a lot more to learn!” (King, 204, p. 163).

Personal Factors Related to Learning

Numerous individuals encounter experiences and circumstances that trigger epochal or incremental dilemmas without the same learning effects in resolution to the dilemmas (e.g., Lohman & Woolf, 2001), leaving what types of factors or conditions facilitate perspective transformation as an open question for transformation theory (Taylor, 1997, 1998, 2000). Although teachers’ statistical learning may not be attributable to specific factors, common to all five teachers are personal factors that may have had some influence on their transformative learning. Key among these factors are interest in the field of statistics, motivation to encounter and resolve dilemmas, confidence in their abilities to learn, reflection on content, commitment to their students and to teaching, and statistical knowledge, with prior knowledge determining readiness for learning and teachers’ perceived need for an overarching framework in which to organize their knowledge.

Interest in the Field of Statistics

Each of the five teachers has an interest in statistics that seems to spur his desires to learn statistics. For example, Blake cites a clear interest in statistics that stemmed from his senior-level college statistics course. Although he does not attribute much of his learning to the course, he does credit the course and the instructor with piquing his interest in statistics.

> And my experience…I don’t know if I learned a lot of statistics there,…but what I did see, uh, I was interested in probability and statistics…I did see in

> [Professor] the passion and the, and so forth, and it kind of sparked an interest. (Blake, Context I, Lines 12-20)

Blake's instructor would describe his statistical work and the analysis he was doing with data from the America's Cup and from Big-Ten athletics. According to Blake, his instructor would give daily updates in class. The combination of sheer enjoyment and passion apparent in the instructor's storytelling, the sports context, and the clear practicality of the subject matter served as "the triggering mechanism[s] [for an interest in statistics], that somehow the statistics that he knew…he was using and enjoying" (Blake, Context I, Lines 128-131). Blake is not alone in having his interests piqued by an influential instructor. Hudson suggests that his instructor's "enthusiasm for the subject was, uh, part of what got me excited about it" (Hudson, Context I, Lines 22-23).

As was true for Blake, Dustin had an instructor who initiated classroom discussion around real problems on which she was working. For Dustin, it was the problem-solving and open-ended nature of the instructor's examples from which his interests in statistics arose. As he indicates, "I thought it was an enjoyable thing to say…let's look at this…how you looked at the data, um, would make a difference...I felt, I think, less straight-jacketed in what my options were...I liked having options" (Dustin, Context I, Lines 135-153).

Although the examples from Blake, Dustin, and Hudson suggest that their interests arise from the actions and interests of others, these five teachers also have interests that may originate internally. Everett, for example, notes that he is interested in answering more statistical questions than time allows for him to answer. He observes, "there's always questions rattling around in my mind about if I ever have the time, I'd like to pursue this and try to figure this thing out. I even have a file with questions, like things to think about" (Everett, Context I, Lines 674-675). Similarly, Isaac describes how his inquisitive nature feeds into his interest in statistics in a way that parallels Everett's expressions of interesting questions and problems. He notes, "I saw interesting problems that I could look at statistically and so in order to solve those particular problems I had, I learned a lot of stuff" (Isaac, Context I, Lines 1182-1208). As Isaac's words suggest, however, interests only stimulate a desire to learn; in order to learn, interests need to be pursued.

Motivation to Encounter and to Resolve Dilemmas

Motivation, the set of reasons that underlie behavior, is needed to act in pursuing interests and resolving epochal and incremental dilemmas. The reasons for acting may be intrinsic, with no clear external incentives for acting, or extrinsic if engagement in the behaviors stems from external sources. All five teachers exhibit interests in and actions towards learning the intricacies of statistics in order to develop their conceptual understandings, indicative of intrinsic motivation, and to benefit the learning of their students, reflective of extrinsic motivation.

The five teachers' intrinsic and extrinsic motivation at times leads them to add, alter, or transform meaning schemes. Blake's motivation in part seems to stem from his desire to resolve what he describes as his "own conflicts," suggesting that his motivation is at least partially intrinsic. He states a desire to understand why statistical formulas and procedures work, noting that "I want to know why" (Blake, Context I, Lines 1588-1711). He seems to find the applications of procedures to be fairly routine, but he seeks to know why a particular procedure is useful in a particular situation and why it might be more useful than another procedure. For example, one of the "why" questions he pursued was why the interquartile range as a measure of variation was needed when there was the standard deviation to measure spread. Even though Blake is motivated to learn, he is perhaps even more motivated to learn on behalf of his students and his teaching—external sources of motivation. For instance, he uses his oversights or mistakes as motivation to learn more in order to "teach it better" (Blake, Context II, Lines 467-468).

As with Blake's motivation to resolve conflicts that he creates for himself, each teacher is motivated to resolve dilemmas that he experiences. Consider Isaac's incremental dilemma that was triggered from an activity that placed him in close contact with statisticians and with access to statistical conversations among statisticians. He indicates, "I was only dimly aware of what they were talking about… especially about probability [and] experimental design" (Isaac, Context II, Lines 515-539). Their discussions served to make him "realize, uh, uh, pretty much how shallow my knowledge was" (Isaac, Context II, Lines 720-722). In order to deepen his knowledge and build new meaning schemes, he describes how he would take notes during the discussions and subsequently consult reference sources and spend considerable time making sense of the discussions of which he had been part. He observes, "I went back home… got some books and studied" (Isaac, Context II, Lines 716-718), with the intention of "making sure that, that my idea was consonant with what they were saying" (Isaac, Context II, Lines 737-738). Because Isaac met with these statisticians on a regular basis, his motivation to resolve his incremental dilemmas may have been partially extrinsic in that he may have wanted to appear knowledgeable to these individuals.

As Isaac's example might suggest, he and the other four teachers are motivated not only to resolve dilemmas but are also motivated to put themselves in positions that are likely to trigger dilemmas. Each teacher serves one or more leadership roles in AP Statistics—positions they knew might challenge their still-developing understandings when they accepted the positions. For example, when Dustin accepted a leadership position that required interacting with other leaders in AP Statistics, including statisticians, Dustin describes feeling "somewhat intimidated by the people" and experiencing "quite a bit of trepidation" (Dustin, Positive CI) prior to submitting his first written work to the group. He describes how he needed to develop a "thick skin" for this experience that made him "wrestle with my understanding of the course, including being aware when there are holes or misunderstandings on my part" (Dustin, Positive CI)—holes that could only be filled by learning. Implicit in the strategies Isaac used as he sought understanding and in Dustin's continued participation in his leadership role is an expectation that their actions would lead to learning and understanding.

Confidence in Ability to Learn

All five teachers felt they needed to learn statistical content to teach the AP Statistics course, and they were not only interested in and motivated to learn the content but also confident in their abilities to do so. Confidence is seen by some as a component of learning (e.g., Broekmann, 1998; Graven, 2004) and a particularly important component for transformative learning (e.g., King, 2004; Merriam & Caffarella, 1999; Taylor, 2000). The process of transformative learning often is "an intensely threatening emotional experience" (Mezirow, 2000, p. 6) for individuals as they react to triggers and resolve dilemmas, which suggests that learners need confidence in confronting difficult challenges to be successful in transforming their meaning perspectives (Taylor, 2000). With little statistical training, the confidence and poise to believe in their abilities was critical for the five teachers to learn the statistical competences they needed to comfortably teach the AP Statistics course.

Prior to teaching the course, Everett, for example, attended a weeklong AP Statistics summer institute that was designed to familiarize participants with the content, technology, and types of data-based activities needed to institute the course in ways that would be consistent with College Board recommendations. Everett left the institute feeling confident in his abilities to learn the statistical content that was new to him.

> The second semester things, the hypothesis testing and all of that, chi square, brand new to me. First time I'd ever seen it…I felt like between the workshop and just what I planned to do—reading and stuff, I thought I'd be okay with that. (Everett, Context I, Lines 731-744)

Even though Everett acknowledges that he had to learn content to teach the course, he believed that his attendance at the institute, background reading, and working through textbook examples and problems would yield the needed learning.

Even though each teacher either explicitly stated their confidence for learning statistical content or implied confidence through their actions, all five teachers say that they still have statistical content to learn. For example, Blake describes his knowledge of design as a 2 on a scale from 1 to 10. Hudson acknowledges that he continues to learn as he prepares to conduct professional development workshops and as he conducts the workshops. He notes that "it's just that a teacher's question or idea might catch me off balance sometime during a session and I might think hmmm. Nifty question. Wasn't what I was thinking" (Hudson, Context II, Lines 990-993). He suggests that his consideration of alternatives shared by teacher-participants broadens his perspective of statistical ideas, thus broadening his corresponding meaning schemes associated with the ideas. Even though these five teachers exhibit evidence of robust understandings of variation, they exhibit confidence in describing how they question their understandings of statistics in general and most often their understandings in the area of design. Their myriad learning experiences and multiple occasions when their experiences triggered learning that resulted in greater depth of understanding in what they thought they already knew may contribute to their confidence in stating what still needs to be learned, a condition that has been argued as "a primary condition for ongoing learning in a profession like mathematics teaching" (Graven, 2004, p. 181).

Reflection on Content

Teachers' interests, motivations, and confidence may contribute to their serious and conscious consideration, or reflection, of statistical ideas at the core of their dilemmas. Each of the five teachers provides evidence of reflection in his accounts of statistical learning. Some of them openly attribute their learning to reflection, whereas others imply the role of reflection in their learning. Blake, for example, talks about “resynthesizing” his knowledge. When asked *how* he resynthesizes his knowledge, he indicates, “I think it’s just a reflective piece” (Blake, Context I, Line 581). Blake is reflective in describing his progression of thoughts and how his learning advances from reflection—what he describes as a constructive process of “refining [his] thoughts” (Blake, Context II, Line 709) about a statistical concept. Much of his reflective process centers on asking himself questions of *how* and *why* that then lead him to consider additional concepts. He describes the process as one that involves “cycling my way back” (Blake, Context II, Line 777), in which his resolution to one “conflict” triggers another conflict and creates a series of trigger-incremental dilemma-resolution cycles. Blake’s description of “cycling his way back” seems to capture the idea of the recursive nature of learning (Kilpatrick, 1985).

In one example, Blake describes his progression of thoughts in his developing conception of randomization. Much of his process centers on asking himself a series of questions related to randomization that lead him to consider blocking and the role of variation in both randomization and blocking.

> Well why do we randomize, you know? There was always the, the, um, the catchphrase well we randomize to reduce or eliminate bias. Okay. And so you have that catchphrase in your head, and uh, well, but why? And pretty soon I’m pinned down on why... how is it dealing with bias…the bigger problem is that we may not know about this extraneous variable…if we don’t randomize, it may assist itself over one of the treatment groups more than the other… But why do you have to randomize, can’t—isn’t there other processes that make sure it doesn’t hit them. Well maybe, maybe not. But we at least know that randomization does. Or, you know, and if it doesn’t, we know…what are the chances… so just kind of pulling those types of pieces together, uh, but can we do better than randomization?…If we know about a variable, we can isolate it or block for it…maybe that’s why we say when we block we reduce variability because we don’t have to worry about the variability of the thing being randomly bounced around the two, the treatment groups. So these are—so all these thoughts…each time you think it, it builds on itself. It’s like another building block in the process. (Blake, Context II, Lines 793-834)

Blake suggests that this progression of thoughts about randomization took place over "the last two or three years" (Blake, Context II, Line 792). In this long passage, he describes how he progresses beyond the mantra that randomization reduces the chance for bias by considering *how* and *why* the reduction occurs. He reveals his insight that randomization in theory divides the effects of an extraneous variable equally into treatment groups and allows for the measurement of chance variability. As he continues to reflect on the effects of randomization and considers alternative strategies, he transitions into discovering *how* and *why* blocking reduces variation through isolating the effects of a particular variable. Seemingly through his reflections and questioning of assumptions related to randomization, Blake enhances his understanding of randomization, blocking, and variation.

Blake's example provides evidence not only of reflection but also of critical reflection. Critical reflection is one of the main elements of transformative learning articulated by Mezirow (1989, 2000) and also validated by others (e.g., Cuddapah, 2005; Taylor, 1998). In general, critical reflection involves an examination of presuppositions for which current problem-solving processes do not provide resolution to the problem at hand (Merriam & Caffarella, 1999). In the example from Blake, we see evidence that he examines his assumptions and beliefs about a number of concepts, including randomization and blocking—assumptions for which his current knowledge did not provide sufficient resolution to his questions of *why* randomization reduces the probability of creating biased treatment groups or *why* blocking reduces variation. Blake reflects on more than content or process (Mezirow, 1991)—the content of experimental design and the process of designing experiments. Blake reflects on the premises that underlie content and process by questioning his previously unexamined assumptions and beliefs (Cranton, 2006; Mezirow, 1985) about *why* randomization reduces the chance for bias and *why* blocking reduces variation. He engages in critical reflection on these concepts.

Although Blake's journeys towards deepened understandings take considerable time to effect change as he introduces new conflicts and works towards resolution, not all significant learning takes years for deepened understandings to develop. Dustin, for example, describes how he went about consolidating aspects of his statistical knowledge in preparation to teach AP Statistics. He felt that he "had all the pieces" of statistics without the bigger "picture" (Dustin, Context II, Lines 938-954). To construct connections among the pieces of his statistical knowledge, Dustin indicates that he "spent a lot of time that summer [preparing to teach AP Statistics for the first time] just thinking about how does all of this fit together" (Dustin, Context II, Lines 938-954). His "thinking" is his serious and conscious consideration of what he knew about statistics and his comparison of concepts and topics to find the connections—his reflections on these topics that resulted in learning through meaning schemes, learning new meaning schemes, and learning by transforming meaning schemes. He notes that authors' presentations of topics were "piecemeal. A little bit here, a little bit there, but no real connection. And if there was a connection, no one ever bothered to, like tell you what it was…you just discovered these things on your own" (Dustin, Context II, Lines 948-952). Dustin indicates that the connections among statistical topics were not made apparent by the authors of the references he consulted, which necessitated that he "discover" the connections for himself. He describes the different actions that he took in his quests towards uncovering connections in the area of design, noting that he consulted a variety of statistics textbooks and spent "a lot of time that first summer thinking about what they were trying to tell me" (Dustin, Context I, Lines 678-679). He indicates that he reflected on the content of his reading to synthesize different authors' interpretations of the content as part of his process of consolidation.

As with Blake, Dustin provides evidence of critical reflection during the course of describing his preparation to teach AP Statistics. In his initial considerations of design, Dustin read sections from a variety of books and reflected on the content presented in those books to develop a unified view of authors' descriptions of the concepts. He describes how over time he began to reflect on the process of design. Specifically, he describes how he moved from viewing the process as setting up a design to then analyze results to a process that centers on the question to be answered and how the design is determined by the question and the analysis to be performed. He also began reflecting on the premises behind the various design types, questioning both *why* we block and *why* blocking reduces variation, for example.

Everett, Hudson, and Isaac provide similar evidence to support the centrality of reflection to their learning of statistics and their engagement in critical reflection. The role of reflection in their learning is consonant with empirical and theoretical literature that places reflection at the center of learning (e.g., Kilpatrick, 1985; Wheatley, 1992), including work focused on teacher learning (e.g., Clarke & Hollingsworth, 2002; Lohman & Woolf, 2001).

Commitment to Students and Teaching

The examples used to illustrate teachers' reflections provide more than just evidence of reflection; the time and thought these teachers put into their statistical activities reveals some of their commitment to their learning and to their students. These five teachers provide additional evidence of their commitment to their teaching and, in particular, to their students.

Like other teachers (e.g., Crawford, 2005), these five teachers seemingly reveal their commitment to their students and to teaching through their attendance in ongoing professional development, their reading of current professional literature, and their service on leadership committees and in conducting professional development. Their reasons for undertaking these experiences are primarily pedagogical, but they often learn content through their pedagogical consideration. Their commitment also drives them to plan carefully for their classes. Leikin and Zazkis (2007) suggest that through predicting students' difficulties, teachers have the opportunity to confront their own uncertainties and questions, potentially triggering incremental dilemmas. Through his consideration of how to teach statistical content in ways that would "make it understandable" for his students (Dustin, Context II, Lines 947-948), Dustin was alerted to many questions about statistics that he himself had to confront. Unlike what appears to be the case in some studies of teacher learning (e.g., Hoekstra, Beijaard, Brekelmans, & Korthagen, 2007), these five teachers allow themselves to make mistakes not only in their planning but also in their teaching. They view their mistakes as learning opportunities. Isaac, for example, describes an understanding he has with his students that if a question arises that he, albeit temporarily, cannot answer, "I always reserve the right to say I don't know. Let me see if I can kind of look it up or figure it out and I'll get back to you" (Isaac, Context I, Lines 1216-1218). From his commitment to his students, he makes sure that he does research and respond to their questions.

An aspect of commitment to students that arose across discussions with multiple teachers in this study was a commitment to learning focused on preparing students for the AP Statistics examination. Rarely, if ever, did the subject of preparing students for the exam arise during conversations with the five teachers identified with robust understandings of variation. The current political climate is fraught with high-stakes testing and demands of teacher accountability (e.g., Shepherd, 2000), and among teachers' reactions to these demands is at least the perception that teachers may "teach to the test" (Schorr, Firestone, & Monfils, 2003, p. 376). Although state assessments are typically the object of these speculations, AP examinations are another form of high-stakes test in that college credit is awarded dependent upon students' performance on an examination. Although some teachers may react to the pressure by studying rubrics and looking for patterns to learn how to better prepare students for the exam, their descriptions of learning activities differ from those offered by Blake, Dustin, Everett, Hudson, and Isaac. The focus of others may center on scoring, for example, rather than the underlying understandings that motivate the scoring decisions. In contrast, the five teachers identified with robust understandings react to similar dilemmas by initiating dialogue with others or gathering resources to read about concepts and statistically explore why the rubrics were developed in a given manner. Arguably, their focus is on learning content, and their commitment is to their students' success in learning the content more so than to their students' success on the AP exam.

Knowledge Base

Two major tenets of constructivism are that knowledge is actively constructed by the learner and that learning is a process of adaptation based on current ways of knowing (e.g., von Glasersfeld, 1990). This view of learning can help to explain why teachers may leave a professional development experience such as the AP Reading with different learning outcomes than other teachers; each teacher enters with different meaning schemes and perspectives that they constructed from their own experiences (e.g., Merriam & Caffarella, 1999). The five teachers who exhibited robust understandings all describe settings in which they developed deeper insights into statistical concepts—insights that they admittedly may not have formed had those same experiences occurred earlier in their careers.

Prior Knowledge

Each teacher describes settings in which he developed insights into statistical concepts based on sufficient familiarity with statistical content to extend his knowledge base. As an example, consider Hudson's positive critical incident. During his attendance at a professional presentation, he participated in simulations designed to develop the idea of chance variability in an experimental setting. As powerful as the experience was for him, he indicates, "timing was important. I may not have been ready to understand that talk had he [the speaker] spoken about it four or five years earlier" (Hudson, Context I, Lines 1012-1014). At the time of this presentation, Hudson was considering the difference in implications of randomization for sampling versus experimental design. The simulations helped him to visualize the meaning of "just by chance" for differences seen in two randomly assigned treatment groups, thus helping him to learn through his meaning scheme for random assignment.

In his positive critical incident, Everett describes an experience that was "a pivotal moment in my understanding of topics beyond AP Statistics" (Everett, Positive CI). He describes a desire to have had that experience earlier in his career, but similar to Hudson, he notes, "if it was much earlier, I don't think I would have had the perspective to appreciate it" (Everett, Positive CI). At the time of this particular presentation, Everett had only recently begun to teach AP Statistics, and his statistical knowledge was limited to topics at the introductory level. He attributes the positive effects of this experience to a skilled teacher who "met me where I was at and then was able to take me to the next spot" (Everett, Context I, Lines 179-182). At this same institute, there were other presentations of content beyond AP Statistics that did not have the same learning effect for Everett. Everett attributes the low level of effectiveness in these sessions to instructors that assumed more prior knowledge than what he or others in the audience had, including Dustin. The experiences described by Everett and Hudson, as well as experiences described by the other three teachers, support the importance of instructors formatively assessing students' understanding in order to allow students to build from their foundational knowledge (e.g., Pegg, 2003).

Overarching View of or Framework for Statistics

An important characteristic shared by the five teachers is what they describe as a desire to see the "big picture" of statistics. Each of the teachers credits either the organization of content in the AP Statistics course description (The College Board, 1996) or his efforts in planning to teach AP Statistics with his development of a "big picture" view. Although a certain amount of statistical knowledge may be prerequisite to developing this view, none of the teachers express such a belief and instead suggest that instruction focused on facilitating students' construction of an overarching view benefits their students. As an example of developing an overarching view, consider the experiences of Isaac. Even though he taught statistics before the advent of AP Statistics, Isaac credits his activities related to the course with helping him to develop a broadened "perspective" of statistics. He suggests that he now sees statistics as an entity comprised of four interrelated parts—the four areas of data analysis, design, probability, and inference identified in the AP course description—that provided "a kind of a…framework, uh, where, um, my existing knowledge and then new knowledge, um, uh, as it came, it sort of found a home" (Isaac, Context II, Lines 179-181). He credits this idea of a framework for his knowledge with helping him to identify "what I thought I knew. And fill in a lot of holes, uh, in my knowledge" (Isaac, Context II, Lines 177-178). Through his consideration and solving of problems that required attention to all four areas, Isaac suggests that he began to make connections among the areas.

Like Isaac, Dustin describes how he had all of the pieces of statistics prior to his summer of learning, "but there was no picture. There was no magic eye, um, until the course showed up" (Dustin, Context II, Lines 941-942). For Dustin, it was the area of design that helped him to make connections and ultimately helped to transform his meaning perspective for statistics. "For me, the question is why not start with design? It seems a natural thing. How would you get data? You know, what are the ways we get data? And then, you move on to okay, now we have some data" (Dustin, Context II, Lines 516-520). By viewing the course from the perspective of design first, Dustin saw what he labels as the "big picture," which seems to equate with the process of statistical problem solving articulated by the authors of the GAISE report (Franklin et al., 2007).

After he started teaching AP Statistics, Hudson also started to think about design first and the positioning of the four areas in the problem-solving process. He credits his reading of the GAISE report with solidifying his view.

> I realized that this model was very useful in almost every situation where you encountered statistics. What's the question? How's the data going to be produced? Or if it's already been produced, was the data produced in a way that's going to help us answer that question? What do you do with the data once you've got it? What interpretations and conclusions can be drawn and do you need probability or simulation to help you do that, or any kind of inference, uh, tool. (Hudson, Context I, Lines 1120-1128)

Hudson describes what he also calls the "big picture" of statistics in terms of the problem-solving process articulated in the report. Although he had been developing a similar integrated view prior to encountering the report, he states appreciation for the simplicity of the articulated process and the integration of the four AP course components throughout the process and recognizes it as something that "applies to almost everything I use with my students" (Hudson, Context I, Lines 1144-1145). Hudson also describes how the report's focus on variation profoundly impacted the way in which he viewed variation. The report helped him to recognize the centrality of variation in statistics by prompting him to consider the different types of variation encountered throughout the problem solving process. As he says, "that focus on variation sort of caught me" (Hudson, Context II, Lines 430-431).

Summary of Personal Factors

Although definitive conclusions cannot be established to determine if the personal factors of interest, motivation, commitment, reflection, confidence, and knowledge base have a causal relationship with the construction of robust understandings, there is definitive evidence that each of the five teachers with robust understandings of variation exhibited similarities with respect to these personal factors. All five teachers expressed an interest in statistics prior to teaching statistics and exhibited motivation, confidence, and commitment for constructing their current statistical understandings. They are reflective in describing their experiences and attribute importance to the role of reflection in their learning. They also provide evidence of critical reflection, which from the perspective of transformation theory is the main mechanism for transforming meaning perspectives (e.g., Cranton, 2006; Mezirow, 1991).

Environmental Influences Related to Learning

In addition to personal factors that may have some effect on a teacher's learning, there are environmental factors that may be more or less conducive to their learning. Each of the teachers cites the importance of a "comfortable" learning environment in which he can feel free to ask questions about content as questions arise. They attribute a sense of community to secondary teachers and statisticians who are active in the AP Statistics program and describe the benefits they attribute to membership in this community, including rational discourse. They also acknowledge the support they receive from their respective districts in terms of the resources, time, and opportunities their districts commit to their professional development.

Comfortable Learning Environment

The teachers in this study identify feelings of safety and comfort as prerequisite for them to feel free in asking questions about content. Blake, for example, describes the electronic discussion group for AP Statistics as a "very comfortable place" where he and others can ask questions that knowledgeable statisticians and secondary teachers will answer (Blake, Context I, Line 773). He notes that at times people disagree about answers, but they always do so "in a very, very nice way" (Blake, Context I, Line 776) so that he and others feel comfortable in posting their questions. He trusts that his questions will be answered in a professionally informative way. The establishment of these types of trusting relationships in which “individuals can have questioning discussions wherein information can be shared openly” (Taylor, 2000, p. 307) is essential for rational discourse.

Rational Discourse

Teacher’s feelings of community and perception of a safe and supportive environment allow teachers opportunities to engage in rational discourse. All five teachers attribute much of their statistics learning to discourse-related activities that include collaboration and interactions with other teachers, interactions with more knowledgeable others, and considerations of alternative points of view.

Collaboration and Interactions With Teachers

Some of the most beneficial learning experiences described by these teachers involve collaboration with or conversations with other statistics teachers. Everett, for example, associates opportunities to interact in pursuing ideas with other teachers with his most pivotal learning experiences.

> The most obvious connection between all of them is that, it— those opportunities, or those events, allowed the opportunity for teachers to just talk with other teachers casually and develop, um, ideas. Pursue ideas. (Everett, Context I, Lines 18-23)

The interactions to which Everett refers and those identified by the other four teachers include those that occur via the electronic discussion group for AP Statistics, in informal conversations with colleagues, and while collaborating on projects.

Prior to the existence of AP Statistics and during the first years of its existence, teachers' interactions with other teachers were limited. During their early years in teaching statistics, Blake, Dustin, Everett, and Isaac each were the sole statistics teachers in their respective schools. Each expresses his gratitude for finding colleagues from other schools with whom he could discuss statistics. Similarly, Hudson describes the importance of having the input of a colleague from his school as he prepared to teach AP Statistics. Isaac, for example, describes a benefit from attending his first statistics institute as "feeling that if I ran into some kind of a problem, there was someone I could call" (Isaac, Context I, Lines 685-686). He suggests that his learning of exploratory data analysis stemmed from "a great deal of conversation, question asking, uh, rather than what we would think of as a class" (Isaac, Context I, Lines 575-577) at this institute. Isaac appears to have taken some consolation from learning alongside others with similar interests who were simultaneously questioning the relative merits of exploring data and how to stimulate statistical learning from such activity—an important aspect of their rational discourse. Teachers' conversations focused on content often stimulated incremental dilemmas, as did their conversations focused on designing activities. Dustin describes his interactions with a colleague from a nearby district and how their conversations transitioned from classroom activities to the content intended to be learned from engagement with the activities.

> Along the way it's like so what's your understanding of this topic? Um, because I keep thinking it's this. And, you know, it's like, and she would say oh, yeah, yeah, yeah. You can't do it that way. Or, no I think it's really this, and then we would go find a textbook and try to figure out what's going on. (Dustin, Context I, Lines 742-747)

Through their willingness to openly confront their potential misconceptions, Dustin and this teacher established a safe and trusting relationship in their developing community during what Dustin describes as their "baptism by fire" (Dustin, Context I, Line 795) in preparing to teach AP Statistics. Mezirow (2000) suggests that this type of trusting relationship allows open and frank discussions—rational discourse—to occur, thus allowing the opportunity for this fundamental component of transformative learning (Cuddapah, 2005; Merriam & Caffarella, 1999). Dustin's description of these early experiences as "baptism by fire" suggests the emotionally-charged atmosphere he and others experienced—atmospheres in which trusting relationships are essential for learning (Taylor, 2000).

Interactions With Statisticians and More Knowledgeable Others

Like teachers in other studies (e.g., Park Rogers et al., 2007; Putnam & Borko, 2000), all five teachers describe educative benefits from interactions within their network of teachers. They seem to place even more value in their interactions with more knowledgeable others, particularly practicing statisticians. Through their involvement in the AP Statistics program, all five teachers have had opportunities to interact one-to-one with statisticians and to engage in rational discourse. Each teacher has either worked with a statistician on designing or implementing professional development sessions, authoring statistics materials, designing activities, or writing AP assessment items, and each describes those experiences as wonderful learning opportunities. Everett indicates that the statisticians are "not shy about saying well that's not really how it is, but they'll also be happy to explain it" (Everett, Context I, Lines 636-639). He describes it as "more of a teammate kind of feeling" (Everett, Context I, Lines 105-106) than one would have in typical student-teacher classroom interactions. Isaac concurs in noting, "she [the statistician] would say I don't think this, this isn't really quite right, what you, what you're saying here. And she'd have a little explanation" (Isaac, Context I, Lines 1501-1504). Statisticians' and knowledgeable others' explanations are important for the valuable information they share, and importance attributed to the type of safe and supportive environment established by these individuals is underscored in much of the research literature framed by transformation theory (e.g., Caswell, 2007; Kitchenham, 2006).

All five teachers engaged in rational discourse with statistics teachers and statisticians to learn statistics and to gain insights into their own assumptions and beliefs related to variation and statistics through the experiences and insights of others. Dustin, for example, describes a time when a statistician told him that one should never block on gender in medical experiments—doctors need to know if men and women react differently to medication. Although blocking by gender may reduce variability and enable isolation of the effects of the medication, in practice, knowledge of a potential interaction effect is of greater importance. It is precisely this type of practical wisdom that these teachers seek to enhance their meaning schemes of statistical concepts. Dustin notes how in designing activities or assessments, statisticians notice details that "you would say oh, I hadn't even thought of that. You know, you think you have the right idea. You would have like a kernel would be, would be the right idea" (Dustin, Context I, Lines 965-967). Isaac describes his interactions with statisticians in a particular capacity as one of his most valuable learning experiences. He indicates that "listening to the statisticians discuss the, you know, various issues, um, uh, what certain, you know, various questions were about and why they were the way they were" (Isaac, Context II, Lines 8-20) provided insights into the premises behind statistical concepts.

Consideration of Alternative Views

Listening to and considering alternative perspectives, particularly through discourse with others, may motivate a teacher to critically question previously unexamined assumptions and beliefs about teaching and learning. Blake, for example, describes how he does not blindly accept statements from knowledgeable statisticians but how he reflects on what is said and tries to make sense from the other person's view by "obsessively think[ing] about that until I have my own resolution" (Blake, Context II, Lines 700-701).

> Some of my incomplete thoughts about design, uh, were challenged by people saying things along the way. And then I had to adjust it, and these people were credible enough that I didn't dismiss what they said. And so now I had to justify what they were saying, and as I justified what they were saying, it took me to a higher level of understanding. And so there was this constant process of, of I guess refining my thoughts, and sometimes finding out, uh, uh, you know, I had changed my thoughts. (Blake, Context II, Lines702-710)

He describes how he asks himself a series of premise-based why questions, such as why does blocking reduce variation? When he reaches a point past which he cannot proceed, he consults a statistician with whom he has conducted numerous workshops or posts his question to the electronic discussion group. The combination of his critical reflection on premises and rational discourse with more knowledgeable others allows Blake to resolve his incremental dilemmas, although resolution may take months or years to achieve.

Blake and the other teachers describe the AP Reading and the AP electronic discussion group as two venues in which they are given the opportunity to consider a variety of opinions on statistical issues. Blake, for example, describes the rubric briefing at the AP Reading as beneficial in terms of "listening to that presentation and that training" (Blake, Context I, Lines 1511-1581). As part of this process, he considers "how they [rubric developers] saw things" and "hearing what people say about that" (Blake, Context II, Lines 838-866). He also describe encounters with different views from reading "arguments" posted to the electronic discussion group, with the arguments sometimes centering on statistical issues that may never have occurred to him. Blake's comments suggest that he engages in premise reflection to transform his meaning schemes related to different concepts.

In addition to considering alternative perspectives through rational discourse, teachers become aware of multiple perspectives by reading textbooks and supplemental materials authored by different people. Everett considers different perspectives presented in written form to be analogous to having multiple instructors teach concepts, each in slightly different form.

> One of the things that's been most interesting for my teaching is teaching out of different books…getting different perspectives, and so you kind of get a sense that, you know, if they only have room for one example in their textbook on a particular topic, you get the sense that that's the way it is. And that's the way that all examples will be, and you don't know that this thing can be applied in so many different ways or taught differently or, and so you have a real limited perspective. (Everett, Context I, Lines 1094-1103)

When textbook authors present concepts in slightly different ways, the reader may experience an incremental dilemma, and eventual resolution of that dilemma results in a richer concept image that reconciles different views.

Reading different textbook presentations of concepts may trigger critical assessment of assumptions related to the concepts and eventually result in transformed meaning schemes. As an example of critically assessing assumptions, consider Dustin's steps to resolve a dilemma triggered from textbooks using seemingly different terms for the same or related concepts. By consulting different textbooks, Dustin encountered the notions of confounding, lurking, and extraneous variables. To reconcile the apparent differences and resolve his incremental dilemma, Dustin compared similarities and differences among authors' descriptions.

> I'd sit there with three or four books open, and I'd read, like say we're doing random variables or design. I'd read all the chapters on design in the books, and then try to figure out, okay, where the points of commonality were, where they were different, and does that different word really matter all that much?…it was just people had different words that they use. I mean do you use confounding, lurking…but the better term is probably extraneous variables…And so extraneous simply says hey, there's all this stuff out here that, you know, you're not dealing with. You just randomize it…And, you know, after a while I realized…they're really talking about the same thing here. (Dustin, Context II, Lines 285-328)

Dustin suggests that by reading and rereading explanations in different textbooks and outlining commonalities in the concepts being described, he eventually transformed his divergent images of lurking, confounding, and extraneous variables into a unified conception of variables different from the independent variable(s) of interest to the study and whose contributions to variability in the dependent variable(s) are controlled through randomization. In resolving his incremental dilemmas, Dustin tried to assume the views of the authors and engaged in rational discourse, albeit with himself, to resolve his dilemma.

Each of the five teachers explicitly sought literature to gain further insights into the area of design. In the process, they claim that they developed insights into not only design but also connections of various concepts to design. For example, Hudson read sections from a design book and deepened his understanding of the relationship between design and variation.

> What I never connected in a way that [the author] did was how the different designs account for, uh, the sources of variation. How completely randomized design deals with things differently than a block design in terms of variation and how you're trying to, um, allocate those sources of variation. So for me, that was—he talked about it in a really simple experimental setting with…hamsters. There were only about 8 of them… as I read it and then reread it, I thought hmm. I'm not sure I really knew that. (Hudson, Context II, Lines 379-466)

Hudson suggests that the author's presentation of different designs and use of simple examples to highlight connections between design and variation coupled with his reading, consideration of examples, and reflection clarified the connections. Although Hudson attributes deepened understandings to this reading, he claims further enlightenment upon reading the GAISE report (Franklin et al., 2007).

> It sort of struck me as I was reading through that the authors really thought that different types of variation were really what was important…I had never, in any of my other experiences before these readings, felt that that was the central theme of the course. I've heard people say statistics is the study of variation, but that hadn't been my experience. (Hudson, Context II, Lines 471-491)

For Hudson, the statistics framework and views of variation expressed by authors of the GAISE report was different from what he encountered in his previous experiences yet contained enough similarities that he could make connections to his understandings. The simplicity of the framework, supporting examples for the framework, multiple readings of the report, and reflections on all of the preceding enabled Hudson to construct a more unified view of statistics. He notes, "I thought how simple…we've always talked in AP about there being four components to the course, but they were never particularly articulated together in this way. And now it's been put together" (Hudson, Context I, Lines 1133-1177).

Community

Related to the development of a safe and supportive environment is the notion of community that may develop from or foster such an environment. The five teachers identify membership in communities as important for their learning. In their roles as learners and teachers of statistics, they are members of a minimum of two communities: the AP Statistics community and the community of practice that includes the community they build with their students. Taking consolation from recognizing similar learning experiences in others is an element of transformative learning identified by Mezirow (1991) and a consolation expressed by other teachers. Blake, for example, expresses comfort in seeing questions asked by others on the electronic discussion group. He indicates that "it's nice to know that other people have questions" (Blake, Context I, Line 767).

Each of the teachers describes the secondary teachers and statisticians with whom they interact as part of a community. Blake describes the environment created by educators in AP Statistics as an "open community" (Blake, Context II, Line 857), and he contrasts it with the "math community" (Blake Context II, Line 860). In the AP Statistics community, he suggests that "nobody is concerned about not knowing something"—secondary teachers and statisticians ask questions freely (Blake, Context II, Lines 858-859) and support each other by professionally sharing answers. In contrast, he suggests that in what he refers to as the math community, which presumably consists of the mathematics teachers with whom he interacts, "you almost have to hide your sins, and hide your ignorance" (Blake, Context II, Lines 861-862). Part of the perceived difference might be based in course requirements for secondary certification. There is an assumption that mathematics teachers know well the mathematics that they teach, as Isaac suggests, but the same is not true for statistics. Isaac posits that [secondary] mathematics teachers' mathematical conversations center on pedagogy more so than content, whereas the reverse is true for statistical conversations.

> When you go and start teaching, well everyone around you pretty much, um, knows the math you do...And so there's not as much discussion of mathematics and perhaps more discussion of well how do you teach this concept or whatever. But in the, in the statistical meetings…at least some part of that is...I don't know how to do this. Does anybody know how to do this? And so, uh, being a member of that community means actually helping each other, uh, learn the mathematics as opposed to, um, um, something where, well we all know for sure algebra…and we don't talk about it very much. (Isaac, Context II, Lines 1253-1271)

Perhaps due in part to what these five teachers conclude from their experiences and from the professional development they conduct with secondary teachers are their beliefs that most secondary teachers in the AP Statistics community view their background in statistics as insufficient in contrast with their perceptions of their background in mathematics. They feel a strong bond with others who struggle to learn the content along with them. Each of the five teachers sees himself as a member of this AP Statistics community.

As part of their community of practice (Imants & van Veen, 2008), teachers build (or not) a community with their students and colleagues. In that community, they sometimes serve the role of the content expert, whereas at other times they serve the role of a learner. Isaac describes how his learning becomes enhanced through his students' questions. "I'd make connections as I learned more, and then as I, as I taught it more and the kids questions would lead me to, scratching my head, then I'd put more and more things together" (Isaac, Context I, Lines 1240-1246). Like Isaac's perception of his students' roles in triggering dilemmas, Leikin and Zazkis (2007) indicate that unexpected and unforeseen events that occur during the course of interacting with students are sources for learning. They note, however, that "it is the teacher's curiosity and deep mathematical knowledge that [lead] to develop[ing] new connections" (p. 124).

Resources

The five teachers allude to the support they receive from their school systems in encouraging them to attend and to conduct professional development. Everett, for example, indicates that his school provides him with the release time he needs to take advantage of the learning opportunities afforded to him through his involvement in the AP Statistics program. In speaking about the regional professional development sessions he leads, he indicates, "it's been, you know, valuable for me, and so—and I've been able to convince my administrators…it's all right for them to pay for a sub for me to go do a workshop because I gain from it also" (Everett, Context I, Lines 1526-1530). Clement and Vandenberghe (2000) note that the "presence of learning opportunities is a necessary, yet not sufficient condition for professional development" (p. 87). Everett and the other four teachers implicitly and explicitly acknowledge the support they receive from their school systems; they also detail their efforts in capitalizing on professional development opportunities for their statistical learning.

Characteristics of Learning Experiences

The five teachers who exhibited robust understanding of variation attribute their learning of variation to a variety of factors. They describe their perceptions of the role of statistical theory in their learning and contrast their theoretical experiences with the understandings they believe were constructed from engagement in active and data-based experiences with statistical applications. From their varied experiences, they identify key problems from the solutions or discussion of which they attribute insights into statistical ideas. They also describe learning benefits related to their advanced leadership roles.

Role of Theory

Prior to teaching statistics, each teacher completed at least one undergraduate or graduate course in statistics or probability and statistics. As the teachers describe their experiences, the focus of their early college courses was mostly theory and procedures. Blake, for example, indicates that in his mathematical statistics course, “I was exposed to the math of probability functions, distributions, etc., but I had no idea where it was going or why” (Blake, Negative CI). By math, he means “theoretically-driven,” with “everything… proven” (Blake, Context I, Lines 105-106). Isaac describes a focus of his introductory course as “how to survive with doing things by hand” (Isaac, Context I, Lines 90-100). His course centered on theory and calculations. These teachers report experiences around *how* to do the procedures or proofs, *what* they were doing, and *what* their results were supposed to tell them without knowing *why* the processes worked or *why* someone was motivated to pursue theoretical results, particularly in connection with application. In Mezirow’s (1991, 2000) terminology, their learning focused on content and process, which mostly limited them to creating or enhancing meaning schemes, not transforming their meaning schemes or perspectives. The courses as teachers recalled them did not support transformative learning, which occurs through a process that begins with critical reflection on *premises*—asking and reflecting on answers to questions of *why* (Mezirow, 1991).

In looking back on their early experiences, the teachers suggest that their knowledge base contained little conceptual understanding. One explanation for their perceived lack of understanding from their college courses lies with the teachers. Dustin, for example, “claim[s] some of the responsibility” (Dustin, Negative CI) for his negative impressions of his learning. He indicates that time constraints related to pre-student teaching field experiences meant that he “had less time that [sic] I needed and probably less patience that I could muster to really devote the time and energy necessary to do this class justice” (Dustin, Negative CI). Isaac also seems to accept responsibility for his learning, indicating, “without suggesting any fault, I really didn’t learn any statistics there” (Isaac, Context I, Lines 59-62). Isaac suggests that his use of “calculation formulas just to do basic stuff” (Isaac, Context I, Lines 73) and his proof writing placed his focus on logic and manipulation but not application. He indicates that at the end of the course, “I didn’t really have an idea what such a thing [inference and the formula for *t*] might be for” (Isaac, Context I, Lines 80-81).

Even though these teachers suggest that their courses did not adequately prepare them to teach AP Statistics, they may not have recognized the limitations they now bemoan while they were completing their courses. The coursework did not trigger dilemmas they needed to resolve. As Everett indicates, “looking back at my college classes, I don’t think that they were very useful, but I didn’t really know that at the time. Because I didn’t know what else was there that I was missing” (Everett, Context I, Lines 826-829). Everett’s recollections of his yearlong probability and statistics sequence center on theory and procedure. He describes how he remembers standard deviation from one of these two courses.

> When we calculated standard deviation, we also used…the sum of the x’s squared, or, you know, that business, minus something or plus something. Um, which to me, gives no understanding of what is being measured. Um, and so I would probably still say that, for instance, my understanding of standard deviation was slim. (Everett, Context I, Lines 384-396)

Everett seems to be describing a variation of the standard deviation formula derived from the “difference of the second moment about the origin and the square of the mean” (Hogg & Tanis, 2001, p. 12) formula for calculating variance, $\sigma = \frac{\sqrt{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}}{n}$. He suggests that this formula yielded little insight towards constructing an understanding of standard deviation. The formula he finds to be more valuable, $\sigma = \sqrt{\dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}}$, is consonant with his current view of standard deviation "as kind of the average distance from the average" (Everett, Context II, Lines 40-41) by making the focus on deviations, $X_i - \bar{X}$, and average, $\dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$, transparent. Everett indicates that it was only after a few years of teaching statistics that he made sense of this interpretation of standard deviation.

Even though none of the teachers identified early courses as their most valuable learning experiences, they do not entirely discount their experiences. Isaac, for example, indicates that he may not "look back on [it] and say oh, this was helpful in my later [statistical experiences]" (Isaac, Context I, Line 98-99), yet he describes his courses as providing him with necessary foundational "tools." He was able to "get an idea of…descriptive statistics. I did get an idea of what a hypothesis test was, and a kind of an idea, although I look back and think not really well developed, of what a confidence interval was" (Isaac, Context I, Lines 103-106). His course experiences provided him with the opportunity to construct a foundational base from which he could build more conceptual and robust understandings, with one exception. The one area in which all of the teachers feel that their backgrounds were insufficient in providing even foundational "tools" is design.

Role of Data and Activity

In their positive critical incident descriptions and during their conversations about their learning, each of the five teachers identifies activities that stimulated what they perceive to be significant learning experiences. A common feature among their learning experiences is their engagement with activities. Dustin, for example, describes collaboratively designing and conducting an experiment.

> What you got to consider are some of the well what if this, what if that, you know, what kinds of things affect the variability? How might this be a better experiment that would give you more, um, more precise results…How can we, um, are there things we can control?...and eliminate some of the variation that exists here. You know, and again, if you're just reading a book, you don't necessarily get that kind of interaction…when you sit down with a group of three or four or even two to say okay, let's talk about, and you actually play with it a while, and then you start to say oh, these kinds of questions need to be answered. (Dustin, Context II, lines 193-211)

Dustin suggests that by actually "playing" with the experimental treatments and groups and working with other teachers to design an experiment, more questions and ideas about variables that contribute variability to results and ways to control variability can be generated than by reading about a similar experiment in a book. Similarly, Dustin suggests that listening to different groups describe their designs provided further stimulation from considering the sources of variation they identified and the ways in which they controlled for variation from those sources.

Teachers also found that conducting sampling or experimental simulations was extremely beneficial for them to develop insights into statistical concepts for learning through their meaning schemes of the concepts. Blake, for instance, describes a simulation that expanded his views of variation and the meaning of *p*-value. He describes how prior to engaging in the simulation, he treated inference procedurally.

> Sort of a stimulus-response mechanism. I know the *p*-value was too low. I know what to do… I know how to calculate conditional probability. I knew what the standard deviation and variance was. I could recognize the formula, and so forth. And…all the statistics that I had learned…dealt with a problem where here's an event that we observed. Run it into this probability model…generate a *p*-value to get me a decision. (Blake, Context II, Lines 113-123)

Blake describes the *p*-value as a calculated result from a formula consisting of known values, with a decision following from the magnitude of the value.

In contrast to his procedural description of how to find a *p*-value, Blake describes how the simulation helped him to "internalize" the meaning of the *p*-value.

> I guess it really hadn't hit home that that probability model was based on an assumption…and the *p*-value was based on the measured variability that's, you know, it's expected… the concept of sampling distribution, if you were to repeat this process over and over and over again under the assumed conditions, um, you're going to see varying results…And so does our result fit into this, uh, sampling distribution or not? Let's make a decision...the idea of sampling distribution became the, uh, became the thing that came to life for me. Uh, because of these simulations…before it was just knowing it, and I knew what to do with it, but I, I didn't see how it fit into our universe, and, and so forth. So, uh, so more internalized. (Blake, Context II, Lines 125-159)

By using a simulation to test a proportion, Blake indicates that the assumption behind the null hypothesis was made explicit—in the case of his example, the population was constructed to have a ratio of success in alignment with "assum[ing] the null hypothesis" (Blake, Context II, Line 72). With multiple workshop participants using graphing calculators to simulate the collection of samples and plotting results on the "dotplot [that] was being created on the board" (Blake, Context II, Lines 90-91), Blake was able to see the development of an empirical sampling distribution and develop a sense for what type of deviation from expectation represented a probable or improbable outcome for a sample. For Blake, he believes that participating in this simulation brought a deeper understanding of sampling distribution and the effects of variation on drawing conclusions from a single sample result in comparison with a theoretical sampling distribution. The visualizations stimulated premise-based reflections behind *why* the inferential process worked.

As Blake's example suggests, there are times when teachers learn by exploring statistical ideas with data and technology. A conceptual focus of the simulation described by Blake was sampling distribution—a concept for which developing an understanding seems elusive for individuals in other studies (e.g., Heid, Perkinson, Peters, & Fratto, 2005). One of the major barriers to understanding seems to be the complexity of the concept. In the case of sampling distributions in particular, Everett stresses the importance of physically acting out a simulation before turning to technology, believing that going straight to the technology equates with learning from a formula if one has not "really figured it out or experienced it" (Everett, Context I, Lines 988-1009) at least vicariously. Hudson also states a belief in "the importance of using physical simulation first before moving to any sort of computer" to better enable making "the connection between the tools and what they represented" (Hudson, Context I, Lines 912-928). He indicates that for him, he needed both to develop deeper understanding of "just by chance," noting that "I think it's harder—it was harder for me to link to the computer simulation if I hadn't had the physical distribution first" (Hudson, Context I, Lines 925-927). Both teachers seem to suggest that physically enacting the collection of one or more samples allows connections between the physical action and the computer's mimicking of the action to become more transparent and believable, thus helping to concretize the abstract notions under exploration.

After describing their experiences in learning statistics, the teachers speculated how different circumstances may have better facilitated their learning of statistics. In general, they suggest that a first experience with an introductory course such as AP Statistics would have provided the familiarity with concepts and applications they needed to make sense of the theory. Everett indicates that his mathematical statistics courses were not "very valuable to me because I didn't have an experience like AP Statistics that dealt with data and gave me the big picture [of statistics] first that I could fit all those other things into" (Everett, Context I, Lines 400-411). Everett does not suggest that he would eliminate theory but instead suggests that a course such as AP Statistics may have given him a framework for statistical thinking to which he could then connect theoretical ideas. Hudson echoes Everett in describing what he would change about his experiences.

> I would have taken a, um, a, a non-calculus based introductory stats course first that painted a better big picture view of the, uh, reasons we're doing all of this. I would have still followed up with the calculus-based one, but I really would have

> benefitted, I think, from having a Stat 101 early, um, myself. (Hudson, Context II, Lines 1242-1247)

Hudson attributes benefit to a two-course sequence similar to the two-course sequence of a data production and analysis course with a survey course of probability and statistics recommended in the *Mathematical Education of Teachers* (CBMS, 2001). Hudson and Everett seem to suggest that following a data analysis course with a theoretical treatment of content might be more effective in enabling preservice teachers to identify connections between the practical and theoretical aspects of statistics and provide a basis for premise-focused reflection.

In general, these five teachers recommend designing instruction and activities that offer students opportunities to make explicit connections between theory and practice. Everett, for example, describes a lecture he attended that helped him to understand how blocking reduces variation. He describes how the presenter started with something with which he was "comfortable," a "two sample situation first, really easy numbers to work with" (Everett, Context I, Lines 198-199). The presenter then moved from the *t*-test to creating a vector of the data represented as a matrix and partitions of the observed data into mean, treatment, and error vectors. The presenter operated on the vectors that represented "sources of variation…different components" (Everett, Context II, Lines 21-22) and compared the approach to the more familiar *t*-test before continuing with more complex examples, including data from a randomized block design. Everett notes that this approach allowed him to put "all those sources of variability together and pulled together a lot of different pieces of understanding that I had had up to that point, but hadn't really had a comprehensive understanding" (Everett, Context II, 24-28). The instructor described the mathematical theory that supported using vectors and partitioning vectors to examine the effects of different sources of variation and connected the theory to the application.

As suggested by the representative examples of Dustin, Blake, and Everett, each of the five teachers describes benefit from concept-focused activity and from experiencing as active learners the statistical content they teach—characteristics of "high quality" professional development (Cohen & Hill, 1998, 2000, 2001; Darling-Hammond & Ball, 1998; Smith, Desimone, & Ueno, 2005) recommended by statistics educators (Heaton & Mickelson, 2002) and generally valued by teachers (Park-Rogers et al., 2007). Their suggestions for preservice teachers' courses in statistics center on data-based, exploratory activities. Their suggestions for preservice teachers support researchers' suggestions that teachers should have opportunities to learn statistics in the same manner in which they are expected to teach statistics (Heaton & Mickelson, 2002; Peck, Kader, & Franklin, 2008) and opportunities to connect their applications to the theory that underlies the content.

Key Examples and Problems Targeting Key Ideas

In addition to participating in statistical activities, teachers described learning benefits from working on problems that draw attention to fundamental statistical concepts and principles and experiencing presentations of key examples. Hudson describes conversation surrounding a particular AP question that asked students to describe an advantage of using a single type of shrimp to determine the effects of different nutrients and salinity on the growth of shrimp. Hudson recorded his response to the problem before attending the briefing on the question's rubric at the AP Reading.

> And as I was listening to her [the presenter] explain more about the motivation for the question, I realized that the answer I would have given, about the, uh, benefit, uh, of using just one type of shrimp…wasn't what she was really focused on, which was that using one type of shrimp would remove one potential source of variation, so it would be easier to isolate the effects of the, uh, the treatments from differences, uh, that you might see as a result of different shrimp species or types of shrimp. And so it was about variation, but in looking at my attempted

> response, my practice response before I went to the Reading, I didn't focus on variation…In this case my failure to notice the primary issue being variation was, uh, was concerning to me. (Hudson, Context II, Lines 59-86)

Hudson identifies "the fact that I misidentified the key issue" (Hudson, Context II, Lines 86-87) as triggering an incremental dilemma that he needed to resolve. The statistician who designed the problem offered an evening session to discuss the issues presented in the briefing for this particular problem. Hudson describes how her use of a key example in that session—one focused on testing deodorant brands for effectiveness that "grabbed me better than the tiger shrimp scenario" (Hudson, Context II, Lines 118-119)—enabled him to see "how the different varia—the different variations, chance variation, variation due to, um, gender differences, variation due to individual differences potentially were interfering with your understanding" (Hudson, Context II, Lines 108-111) and how the total variation could be reduced by focusing on just men and removing variation from gender differences. He suggests that the deodorant example provided a context that "I know a lot about. Deodorant I'm much more familiar with on a personal level" whereas "tiger shrimp is not something I know about" (Hudson, Context II, Lines 123-123). The familiar context made it possible for him to imagine how focusing on men could produce less variable results than an experiment examining the effects of the deodorant on both men and women.

Exploring New Roles

Additional transformative elements observed in the actions of these teachers are experimenting with new roles and building a sense of competence and self-confidence in those roles. When faced with experiences that could lead to transformation, each of the five teachers embraced the opportunities presented to them and took the necessary risks that might seem daunting and inhibit others from achieving transformation (King, 2004). They viewed the triggers they encountered in their new roles as learning opportunities and reacted to the triggers by following a plan of action that incorporated characteristics of prior learning experiences such as those already described. Each of these teachers has become increasingly more involved in the AP Statistics program by conducting professional development, assuming positions as AP readers and positions in increasingly higher-levels of leadership, and authoring statistics publications.

Conducting Professional Development

Since they began teaching statistics to secondary students, Blake, Dustin, Everett, Hudson, and Isaac have expanded their teaching repertoires to include teaching secondary teachers. Each leads numerous workshop sessions and summer institutes focused on both content and pedagogy in statistics education. Although they may be the more knowledgeable others to the teachers who attend their sessions, they recount their own statistical learning from planning for and conducting professional development for inservice teachers. Hudson, for example, indicates that planning is an educative activity for him.

> Because I have to think through the why. Um, why am I picking this activity? When we implement it, what are the points I'm going to emphasize? Why am I picking this activity rather than another one that's similar? Uh, what do I want the teachers to get out of it? So I think, I'm probably learning as much, anyway, in the development as I am in the delivery. (Hudson, Context II, Lines 982-988)

Hudson suggests that he learns as much statistics in thinking through premise-based pedagogical questions as he learns from teaching the sessions. In conducting the sessions, he indicates that teachers' questions about content are most educative for him, particularly because of the alternative views the teachers present. He notes that "the why and the how questions from teachers…start making you think about alternative viewpoints" (Hudson, Context II, Lines 1010-1011). The learning benefits that teachers describe from their experiences in teacher education align with the benefits they describe from planning and enacting lessons for their secondary students.

Assuming Leadership Positions and Authoring Statistics Publications

Beyond conducting professional development and as part of their ongoing involvement in AP Statistics, each of these teachers has pursued or been asked to assume leadership roles in AP Statistics. Their explorations of new roles as part of the process of transformational learning are similar to explorations seen from teachers in other studies (e.g., Caswell, 2007; Whitney, 2008), including roles in leadership (Caswell, 2007). Similar types of involvement have been identified in conjunction with substantial contributions towards "effective" teaching (Poulson & Avramidis, 2003). Each of the teachers expressed trepidation before agreeing to assume new leadership roles in AP Statistics, including those that they themselves pursued. As already discussed, however, they ultimately found their interactions with more knowledgeable others in these capacities to be extremely beneficial for their learning. They emerged from their experiences with greater confidence, which may have contributed to their desire to author statistical publications.

Four of the five teachers have authored publications, including commercial curriculum materials, AP Statistics-specific materials, journal articles, and teacher preparation materials. In the process of writing, these teachers continued to experience triggers. Isaac, for example indicates that he "was immediately struck by how completely unprepared I was to do anything like that" (Isaac, Context I, Line 1428). Even though each of the four authors had been teaching statistics and conducting professional development in statistics for years, they found that the process of writing required even greater understanding of content. As Everett describes the phenomenon, he compares authoring with teaching.

> It forced me to go out and figure all those little details out and make sure I understood it, and was able to – and of course, you know, I've said earlier that having to teach something really forces you to try to understand it well. When you have to write something, you only get—it's even more so because you only get one chance at it. (Everett, Context I, Lines 1456-1462)

Everett suggests that unlike teaching, there are no (immediate) opportunities to respond to questions when explanations are shared with unknown numbers of consumers.

Additionally, unlike teaching for the most part, writing for publication typically is read and critiqued by potentially more knowledgeable others before publication. Each of the four teachers describes the learning benefit of feedback that typically comes from statisticians as well as secondary teachers. As noted earlier, the primary benefit of their feedback did not come from the identification of a mistake but rather from the explanations given for the reasons behind why there was an issue in what had been written. In several cases, the more knowledgeable others were coauthors. Collaborations with others was found to be a factor that influences teacher transformation in a study conducted with teachers in attendance at a summer writing institute (Caswell, 2007), as was the type of active and ongoing involvement displayed by these teachers. The dedicated interest they show to their own development is similar to the type of interest identified as supporting and accelerating teachers' process of change (Polettini, 2000). In the case of these teachers, their interests in their own intellectual development may have contributed to their changed understandings, namely their construction of robust understandings of variation.

Summary of Teachers' Transformations

In describing their learning experiences, the five teachers with robust understandings of variation articulated a variety of characteristics that are indicative of transformative learning (Mezirow, 1991, 2000). They provided evidence that their meaning perspectives for statistics transformed during the time they taught statistics, which also had implications regarding changes to their meaning schemes for variation. Prior to teaching statistics, each teacher had developed a view of statistics based on their experiences as statistics learners. Everett indicates that before he started teaching statistics, if someone had asked him to describe what statistics was, he would have said, “statistics are like things like batting average and…my experiences from sports and things like that…I probably would have drifted into probability a little bit, not really knowing much of a distinction” (Everett, Context II, Lines 1015-1020). He explains that he once thought of statistics as “sort of a thing that you do” (Everett, Context II, Line 1044), which is consistent with a view of statistics as a collection of procedures used to calculate statistical values and probabilities. In contrast, he currently describes statistics as “a decision-making tool,” “a process,…a way to answer questions by designing studies and collecting data and then analyzing the results” (Everett, Context II, Lines 1036-1043), a description which seems to indicate a process different from procedural calculations and one that is consistent with the statistical problem-solving process outlined by the authors of GAISE (Franklin et al., 2007). Everett’s description of statistics transforms from statistics as a procedurally-focused science of calculation to a problem-solving process that he describes as “the art and science of, um, making decisions with data” (Everett, Context II, Lines 1041-1042). Like Everett, the other four teachers describe probability-laden initial views of statistics that over time and through reflection on experiences such as those described in this chapter transformed into what they would term a “big picture” view of statistics as a problem-solving process.

All five teachers experienced a number of incremental dilemmas that prompted reflection, including critical reflection, on content, process, or premises. During the course of resolving their dilemmas, they often sought input from colleagues and more knowledgeable others through engaging in rational discourse. This consultation typically occurred as part of a plan of action enacted to construct the knowledge and skills needed to resolve their dilemmas. As they began to feel greater confidence in their abilities, they experimented with new roles that often corresponded with leadership positions. These positions often triggered further dilemmas that needed resolution. All five teachers currently profess satisfaction with their knowledge of AP Statistics content and are now pursuing knowledge of advanced content that seemingly aligns with their transformed meaning perspectives of statistics.

Along with their transformed views of statistics, in general, the five teachers gained additional awareness of the importance of variation to statistics and throughout the statistical problem-solving process. As part of their meaning perspective transformations for statistics, some of the quintet transformed their meaning schemes for variation, whereas others broadened their meaning schemes. For Dustin, variation changed from something with which he needed to deal to a concept central to every aspect of statistics from design to analysis. Hudson describes his initial surprise at reading the work of statisticians who continually placed variation as the primary focus of design and statistics. He was particularly surprised to see the connection between variation and randomization—that taking a random sample produces a representative sample but also allows for the introduction of "a probability model that you can use to quantify things" (Hudson, Context II, Line 419). Changes in Dustin's and Hudson's views of variation were particularly profound. Perhaps more typical of teachers' change in views, Isaac suggests that he recognized the centrality of variation to statistics early in his learning and thus exhibited learning through his meaning scheme for variation rather than transformative learning. Nonetheless, each teacher experienced changes in their meaning schemes for variation in conjunction with their processes of transforming their meaning perspectives for statistics.

Nature of Teachers' Transformations

Perhaps the most striking feature of the five teachers' transformative learning experiences is that a majority of the experiences that facilitated their transformations are ones that are accessible to most teachers. Why, then, did these experiences stimulate transformative learning for these five teachers? Stated differently, how and why did the five teachers recognize and embrace opportunities to resolve the triggers they encountered during their learning experiences?

All five teachers exhibiting robust understandings of variation attribute their development of foundational views of variation to their early statistics experiences. More importantly for them, they suggest that their interests in statistics as a field were stimulated from their early experiences. It is this interest that may have prompted these teachers to teach statistics. A brief perusal of the reasons the other eleven teachers began teaching statistics reveals that they were prompted to teach statistics from largely nonstatistical or external stimuli, such as suggestions from department chairs or administrators (e.g., Faith, Ivy), their desires to teach upper-level students (e.g., Eden, Jenna), their impressions that statistics content is valuable content that students should know (e.g., Carl, Cheyenne, Dana), or other logistical reasons (e.g., Frank, Gavin). Although other teachers (e.g., Georgia, Haley) also expressed an interest in statistics as a motivating factor for teaching statistics, their interests seem to differ from those of the five teachers. Their interests focus on the structure of the subject as they experienced it rather than on the more general field of statistics and its applications. Even though there may have been other factors that influenced Blake’s, Dustin’s, Everett’s, Hudson’s, and Isaac’s decisions to teach statistics, the overriding rationale they articulate is an interest in the subject. Their early experiences may have triggered overarching touchstone dilemmas for them—ones that remained latent at times but that provoked them to teach statistics and provided internal stimuli to motivate them to learn statistics. Their innate interest in the subject may have provided them with extra motivation to not only face but embrace learning opportunities inherent to the triggers they encounter.

Blake, Dustin, and Isaac each taught statistics before the advent of AP Statistics. Each expresses motivation to learn and commitment to students and teaching, yet in hindsight, it is clear that none of them constructed robust understandings of variation solely from their preparations for teaching and teaching these introductory courses. Part of an explanation for why their understandings did not develop to the point of robustness may lie in the fact that design—one of the most important aspects of statistical work and an area that can be used to frame all statistical applications—was not included in the statistical content they learned or taught previously. Each of the five teachers describes struggling to discover how the area of design connected to other areas that also were part of the course. The AP Statistics course description provided a sort of framework in that it separated content into four relatively distinct areas of exploratory data analysis, design, probability, and inferential statistics, from which they attempted to piece together the areas to form a "big picture" of the course content and of the larger field of statistics. Their desires to have overarching views of the subject matter they taught along with their confidence in their own abilities to learn and their commitments to their students and teaching may explain the stimuli that supported their motivations for learning. It seems they want their students not only to succeed on the AP Statistics examination but are looking for ways to help their students succeed in understanding statistical concepts in ways in which students can see connections among the concepts and procedures they learn in the course. The absence of design, however, may not fully explain deficiencies in teachers' views of variation in other areas such as inferential statistics, for example. It may be that design provided a missing piece of the statistics puzzle for them—a piece that when added to the puzzle revealed smaller holes in knowledge that could not be seen before the missing piece was put in place. Their desire to see the "big picture" may have provided the motivation they needed to react to new content as a learning opportunity more so than an obligation.

Unlike the other four teachers, long before the advent of AP Statistics, Isaac faced a dilemma that caused him to rethink his views of statistics. Isaac's experiences at a four-week leadership institute in statistics parallel his and the other four teachers' experiences with AP

Statistics in several important ways. In particular, the main focus of the institute, exploratory data analysis, differed considerably from Isaac's previous experiences with data. As part of attending the institute, Isaac made a commitment to conduct future professional development sessions with secondary teachers and to implement with his students some of the activities that were designed over the course of the institute. He was faced with the dilemma of teaching content for which he had little to no foundation, much as he and the other teachers were faced with teaching content in experimental design when they committed to teaching AP Statistics. The absence of *any* perceived foundational knowledge for new content may have triggered epochal dilemmas that caused the teachers to question their assumptions and beliefs behind statistics in general rather than incremental dilemmas that may have provoked development of new meaning schemes or transformed meaning schemes related to design or exploratory data analysis in particular without provoking a need to question assumptions and beliefs in statistics more generally. Alternatively, although each of the five teachers suggest they did not have sufficient background knowledge to make sense of the new content, they may have had enough prior experiences that they still were able to recognize some of the subtleties inherent to the content. Their background knowledge and experiences may have focused them on the enormity of learning and truly understanding the content rather than on learning superficial aspects of the content. They may have been able to discern true understanding of content from superficial acquaintances with content.

Other possible explanations behind why the five teachers may not have constructed robust understandings prior to teaching AP Statistics include that they did not encounter dilemmas that required robust understandings for resolution or that potentially triggering events failed to trigger dilemmas in them. In considering experiences that triggered learning in their later experiences, there are noticeable types of experiences absent from their early experiences. In particular, their ruminations do not suggest that they had opportunities to experience multiple activity-based workshop or conference activities, simulations, or rational discourse from one-to-one or small-group interactions with colleagues and statisticians. Student questions also do not appear to have triggered any memorable statistical dilemmas. Their early coursework and teaching experiences apparently allowed them to construct meaning schemes for new statistical procedures and concepts and to enhance their existing meaning schemes in statistics, but they were not prompted to reflect critically on the premises for their meaning schemes and perspectives—the experiences did not challenge their existing views of statistics. Without concrete experiences or virtual contextual experiences with the content, they may not have formed sufficient uncritically assimilated beliefs, assumptions, or perspectives in statistics for triggers to produce epochal dilemmas. The question remains that if, indeed, teachers' early experiences failed to trigger dilemmas that would lead to transformation, what ultimately provoked their transformative experiences? Part of the answer may be evident in the circumstances behind Isaac's pre-AP transformative experience.

Although Isaac could have responded to the leadership institute by discounting the information shared by the institutes' leaders, the leaders were renowned statisticians who described analyses of importance to their work. The leaders may have presented the content in ways that enabled Isaac to see the importance of the content, and the leaders had sufficient stature that Isaac could not easily discount the ideas they shared. Similarly, the designers of AP Statistics are statisticians who focused the course content on involving students in doing statistics. In general, statisticians can provide examples and applications from their own work of doing statistics, which may stimulate incremental dilemmas that focus teachers on some of the subtler ideas of statistics that are important for drawing valid conclusions from data—ideas such as confounding or extraneous variables in design. Statisticians also can present alternative perspectives for underlying rationale behind statistical processes and concepts and thus trigger learning in the communicative domain as teachers attempt to understand the meaning behind the perspectives, or premises, presented by statisticians.

Isaac's leadership institute and the other four teachers' attendance at professional development sessions provided them with opportunities engage in rational discourse with leaders and other teachers to make sense of statistical content. All five teachers were proactive in their education and attended professional development to learn about the content and pedagogy recommended for introductory statistics courses. Opportunities for rational discourse were important because they could recognize some of their own content struggles in others, and rational discourse allowed them to explore new roles as statisticians, relationships with statisticians, and actions in doing statistics by conducting experiments and analyzing context-based data, for example—each characteristic of transformative learning. Their engagement in rational discourse, particularly with statisticians and more knowledgeable others, may provoke questioning of their beliefs and assumptions related to statistics in ways that may not be triggered from interactions with students and peers. In their early professional development experiences with statistics, each of the five teachers met at least one secondary teacher or statistician with whom they developed lasting relationships—relationships that through rational discourse constantly presented triggers of incremental dilemmas as they questioned their underlying beliefs, assumptions, and perspectives to make sense of the meaning communicated by others.

In addition to attending and continuing to attend professional development that focuses on content and not just activities, the teachers interact with others in a number of ways, each of which they find to be valuable for their learning. They each regularly read teachers' and statisticians' postings to the AP electronic discussion group and pay particular attention to responses posted by known and knowledgeable statisticians who not only offer "correct" responses but who provide explanations and reasons, or premises, that underlie the content and processes. They became readers and tables leaders, where they were able to personally interact with other secondary statistics teachers and statisticians. They gained what they describe as valuable insights into important aspects of statistics based on the rationale behind scoring decisions and insights into subtleties in content and language by examining students' closely worded responses that were awarded different scores. Through these activities, they developed a network of teachers with whom they felt part of a comfortable community in which there is no hierarchy among members. They describe how in their interactions with other teachers and statisticians, they feel like teammates who share common goals of understanding. Similar to benefit gained by reading multiple textbook explanations, teachers' interactions with colleagues and more knowledgeable others present them with alternative perspectives and offer additional considerations for their premise-focused reflections on the content. They found similar benefits in their interactions with others in their leadership roles and in their authoring of statistics publications. These five teachers embrace each of these opportunities as learning opportunities; they embrace their experiences with triggers and look forward to resolving the dilemmas they encounter.

At the beginning of this section, I asked two questions that I attempted to answer: Why, then, did these experiences stimulate transformative learning for these five teachers? Stated differently, how and why did the five teachers recognize and embrace opportunities to resolve the triggers they encountered during their learning experiences? I provided some possibilities for why triggers and experiences may have differently affected the teachers in this study, including their interests in the field of statistics, their desires to have an overarching framework for content for themselves and their students, their sufficient foundational knowledge upon which to build deeper understandings, and their embracing of opportunities to engage in rational discourse and potential learning experiences. In most cases, the extent to which they embrace learning opportunities may distinguish them from other teachers; each of the five teachers who exhibited evidence of robust understandings invested considerable time, energy, and effort in their journeys to robustness. Typical of the way they embrace learning opportunities, they see value in learning statistical content beyond the introductory level and are interested in resolving premise-based dilemmas they have—dilemmas that now extend beyond the content in AP Statistics.

Chapter 8

Implications, Limitations, and Future Directions

This study set out to investigate secondary teacher–leaders' conceptions of statistical variation and for those teachers who exhibit robust understandings of variation, the activities and actions that contributed to their current understandings of variation as reflected in their perceptions and recollections of experiences. Data analysis revealed three distinct types of conceptions of statistical variation: Expected but Explainable and Controllable (EEC), Noise in Signal and Noise (NSN), and Expectation and Deviation from Expectation (EDE). Characteristics of these conception types form an image of how advanced knowers view the important concept of statistical variation—an image that complements previous research work detailing young learners' conceptions of variation. Analysis of teachers' reasoning about variation also produced an empirically derived framework for robust understandings of variation. The framework consists of two cycles of levels of reasoning in the formal mode associated with SOLO (Biggs & Collis, 1982, 1991). Robust understandings of variation are indicated from integrated reasoning about variation across three perspectives—design, data-centric, and modeling—in the second cycle of levels. This framework encompasses both the outcomes of this study and previous research results detailing students' sophisticated reasoning about different facets of variation while remaining consistent with statisticians' and statistics educators' expositions about what it means to understand statistical variation. Additionally, this study adds to research literature that investigates students' progressions in learning to reason informally about variation by detailing personal and environmental factors that reportedly influence advanced knowers' journeys towards robust understandings of variation, including elements of reasoning with formal measures of variation and reasoning about variation in relation to formal statistical procedures and methods.

This study represents a first step in a larger research program focused on teacher education in statistics. More generally, the program centers on contributing to the research base in statistics education by providing answers to a coherent set of related questions, including the following four questions. How do interactions with more knowledgeable others, particularly in terms of how they support critical reflection and rational discourse, affect how teachers (and students) construct robust understandings of statistical concepts? What professional development factors or experiences trigger dilemmas and facilitate resolution of dilemmas in ways that contribute to the development of teachers' robust conceptions and understandings of statistical concepts? What strategies and tools are useful in investigating how teachers develop statistical and pedagogical knowledge needed for teaching statistical concepts? What characterizes prospective and inservice teacher education courses and programs that facilitate teachers' development of robust statistical conceptions and pedagogical strategies for successfully teaching statistics? In succeeding sections, each of these questions is examined individually, with discussion focusing on questions for future research and implications from this study for investigating answers to those questions. Limitations of this study appear as part of the discussion.

How do interactions with more knowledgeable others, particularly in terms of how they support critical reflection and rational discourse, affect how teachers (and students) construct robust understandings of statistical concepts?

Viewing teachers' perceptions of beneficial learning experiences through the lens of transformation theory offers insights into mechanisms behind teachers' learning about variation, which arguably parallels teachers' learning in statistics more generally. Central to teachers' transformative experiences was engagement in critical reflection (e.g., Caswell, 2007; Mezirow, 1985, 1990; Saavedra, 1996). Characteristics of critical reflection seem to differ from characteristics of reflection described by some researchers. Many mathematics education researchers extol the importance of reflection for learning (e.g., Goodell, 2000; Roddick, Becker, & Pence, 2000), but some descriptions of reflective practice focus on reflections related to content (e.g., "what I learned this week" [Goodell, 2000, p. 50]) or process (e.g., "how I learned it" [Goodell, 2000, p. 50]) without considering reflections on premises underlying content or processes (e.g., "why did I learn from this process?"). Because reflection on premises may result in transformative learning and critical reflection was a factor in teachers' constructions of robust understandings of variation, results from this study suggest that reflection on premises, indicative of critical reflection, is an important consideration for investigations of adult learning.

The five teachers' individual reflections on content, processes, and premises were enhanced through engagement in rational discourse with others, particularly with individuals who have more advanced knowledge of statistics content. For example, interactions with statisticians provide opportunities to assume perspectives for how statistics is used in practice, which may cause teachers to question their assumptions and beliefs about statistics content. Many preservice and inservice mathematics teachers have a long history of deterministic mathematical experiences (e.g., Cobb & Moore, 1997; Meletiou-Movrotheris & Stylianou, 2003); attempting to understand the perspectives offered by more knowledgeable others such as statisticians might possibly lead to probabilistic considerations like the occurrence of "chance" events. Involvement of statisticians in teacher preparation is relatively new (Franklin & Mewborn, 2006; Peck, Kader, & Franklin, 2008) but seemingly offers great promise for teacher education in statistics. The quintet of teachers who exhibited robust understandings of variation described the importance of considering alternative perspectives gleaned from reading multiple textbooks and rational discourse with statisticians and more knowledgeable others. The learning benefits attributed to these types of interactions with practitioners and more knowledgeable others merit further consideration from statistics education research.

Although it may not be practical for statisticians to co-teach statistics content courses with mathematics educators, it seems that other methods could provide teachers with opportunities to consider alternative perspectives and engage in rational discourse, such as the use of cases. Cases have been used in teacher education to present real problems, issues, and challenges in teaching (e.g., Lin, 2002; McGraw, Lynch, Koc, Budak, & Brown, 2007) and seemingly could be used to engage teachers in rational discourse that focuses on the content, processes, and premises of the statistics content at the heart of the case.

What professional development factors or experiences trigger dilemmas and facilitate resolution of dilemmas in ways that contribute to the development of teachers' robust conceptions and understandings of statistical concepts?

Important factors of learning reported by the five teachers who exhibited robust understandings of variation relate to important components of professional development identified by researchers (e.g., Weissglass, 1994) and valued by teachers (e.g., Rogers et al., 2007). Learning dilemmas for the teachers in this study stemmed from activities characteristic of "high quality" professional development, including activities focused on content (e.g., Cohen & Hill, 1998, 2000, 2001; Darling-Hammond & Ball, 1998; Goos, Dole, & Makar, 2007; Smith, Desimone, & Ueno, 2005), instructional tasks (e.g., Ball & Cohen, 1999), and examinations of student work (e.g., Darling-Hammond & Ball, 1998). An important question that arises from teachers' descriptions of professional development activities is: What factors trigger the dilemmas for which these activities facilitate resolution? For the teachers in this study, incremental dilemmas were triggered when teachers engaged in learning activities such as using physical manipulatives in combination with technology to concretize complicated statistical concepts through simulation. Other incremental dilemmas were triggered from considerations of key examples or problems that seemed to be particularly effective for stimulating thought.

This study begins the process of identifying characteristics of experiences that successfully triggered dilemmas resolved through content learning. It seems important for statistics education researchers and curriculum developers to identify additional triggers for learning in order to design activities that facilitate learning in resolution to the dilemmas triggered by the activities. Additional questions that are important for teacher learning in statistics include questions regarding characteristics of triggers that may not have been recalled by the teachers in this study, characteristics of triggers for different populations of teachers, and characteristics of triggers for learning different statistics content.

Questions also remain about whether factors or experiences that contribute to the development of robust understandings of variation differ according to individuals' conceptions of variation. Results described in detail in Chapter 5 point to three distinct types of conceptions of variation, but because teachers' conceptions were not uncovered until after data collection was completed, questions asked during teachers' context interviews did not target development of these conceptions. If research finds that important factors for learning differ by conception types, then it would make sense to investigate whether conception types align with correspondingly different views for other statistical concepts and how those different views might affect teachers' learning.

Although this study examined learning experiences for the five teachers who exhibited robust understandings of variation, additional data for the variation-related learning experiences of the remaining eleven teachers was intentionally collected even though it was not needed to address the questions that guided this study. Analysis of the larger body of data offers potential insights into differences in perceptions of learning between those who develop robust understandings of variation and those who do not. Subsequent analysis can offer a possible starting point for identification of characteristics that facilitate or inhibit teachers' resolution of dilemmas for developing robust understandings of variation.

What strategies and tools are useful in investigating how teachers develop statistical and pedagogical knowledge needed for teaching statistical concepts?

This study describes a framework for what it means to have robust understandings of variation and investigates teachers' perceptions of experiences that contributed to their development of robust understandings of variation. In so doing, this study informs research focused on developing teachers' statistical content knowledge. Further research is needed to uncover what knowledge is sufficient for teaching statistics to secondary students, what pedagogical strategies and instruction enable teachers to construct that knowledge, and how teachers can draw on that knowledge to help their students develop similar knowledge to eventually achieve educators' visions of a statistically literate society. These large goals require a collection of powerful strategies and tools to investigate the nature and development of teacher knowledge.

This study offers insights into the benefits of retrospective study for investigating the answers to the question of what pedagogical strategies and instruction enable teachers to construct that knowledge. Although this study relied on teachers' memories to gain insights into factors influential in their learning of statistical variation, particular instruments such as event history calendars (e.g., Martyn & Belli, 2002) and critical incidents descriptions (e.g., Butterfield, Borgen, Amundson, & Maglio, 2005) were used to stimulate teachers' recollection. Several teachers suggested that completing their calendars and checking their undergraduate and graduate transcripts and professional development materials reminded them about some of their early learning experiences that they had forgotten. Although there were some experiences that teachers were not able to describe in great detail, giving an incomplete picture of their learning experiences, the details that teachers were able to provide about their positive and negative learning experiences and the consistency of their stories over the duration of the study suggests that they have detailed recollections of events and experiences that made a difference—positive or negative—in their learning. Because the instruments used to stimulate recall seem to have had positive benefit for eliciting details of teachers' experiences, event history calendars and critical incidents descriptions offer promise for studies to obtain detailed background information about teachers' prior experiences with learning content and pedagogy.

What characterizes prospective and inservice teacher education courses and programs that facilitate teachers' development of robust statistical conceptions and pedagogical strategies for successfully teaching statistics?

A typical consideration in the content preparation of teachers is the nature of the course work required. Although more than 5 of the 16 teachers in this study may have constructed robust understandings of variation that were not fully evoked from the content interview tasks and questions used in this study, there were teachers who exhibited faulty reasoning with respect to variation—teachers who had completed one or more statistics courses at the time of their interviews. Considering the backgrounds of all 16 teachers, this study raises questions about the merits of encountering theoretical probability and statistics courses in one's first introduction to statistics. Each of the five teachers who exhibited robust understandings attributes little credit to his theoretical coursework for his current understandings. By the same token, these five teachers now seek understanding of the theory that underlies the concepts and procedures they teach in their statistics classes. They suggest that a data-based exploratory course that includes design and consideration of the artistic aspects of statistics may have prepared them better for understanding the scientific theory that underlies the content. They still attribute importance to theoretical courses but only if the theory builds from application. Their suggestions seem to align with the two-course sequence recommended for preservice teachers by the Conference Board of the Mathematical Sciences (2001). While it seems unrealistic to think that years of positive experiences can be captured in one or two courses, this study does provide insights into what kind of course experiences may be useful for prospective teachers.

The five teachers with robust understandings of variation stress the importance of focusing statistics instruction on the “big picture” of statistics, which seems to equate with focusing on the statistical problem-solving process articulated by the authors of the GAISE report (Franklin, et al., 2007) in its entirety and considering connections among the statistical concepts and procedures inherent in the process. These teachers suggest that an introductory course should begin with the area of design and use design to motivate the need for exploratory data analysis and inferential techniques. They also suggest infusing design issues throughout the course.

This study found that teachers with all three types of conceptions view the main purpose of design to be controlling variation. The rationale behind their attributions differs among conception types, but what is clear is their views suggest that the main instructional focus in design should be on controlling variability. Teachers exhibiting different types of conceptions articulate different purposes for exploring data, but all describe exploring data to consider relationships among data and variables and all use summary values to help in describing the relationships. For those teachers with EEC, NSN, and EDE conceptions, exploratory data analysis is a means to an end of making inferences from data. Teachers’ movement from exploratory towards confirmatory analysis follows the flow of the statistical problem-solving process—a process that places variation at the center of attention. These descriptions for how teachers with different conception types approach data suggests that instruction can capitalize on teachers’ views of variation by providing data exploration opportunities that are just that—explorations.

Suggested explorations include creating displays of data that highlight relationships among data and variables and use summary values, including measures of variation, that best describe relationships seen in data. To account for different purposes for exploration, data from both familiar and unfamiliar contexts can be explored using data for multiple variables to uncover the relationships among variables.

This study provides support and means for connecting exploratory data analysis and inference to design and introducing students to statistics via issues of design. It adds to the statistics education research literature by suggesting that a combined focus on design and variation may benefit students' learning more than a focus on either design or variation in the absence of the other.

Concluding Thoughts

Investigating teachers' conceptions of statistical variation and for those secondary AP Statistics leaders who exhibit robust understandings of variation, investigating the activities and actions that contributed to their current understandings of variation as reflected in their perceptions and recollections of experiences is just one part of a larger research program. This dissertation study provides an image of how advanced knowers view the multifaceted concept of variation and offers insights into how a subset of teachers believe they constructed their conceptions and robust understandings of variation. Through the theory and methods used to answer its two research questions, this study provides a firm theoretical and methodological base for future studies. Results from this study provide the basis for formulating hypotheses about teacher learning for prospective secondary mathematics teachers and formulating hypotheses of secondary teachers' learning for statistics in general. In short, this study provides background information needed in research to address important questions that remain for teacher education in the content area of statistics. By motivating and investigating the four questions highlighted in the preceding discussion to build on the results of this dissertation study and the larger body of existing work in which this research resides, this study and the larger research program can contribute to the scholarship and practice of teaching within statistics education and teacher education.

References

Arnold, P. (2008). Developing new statistical content knowledge with secondary school mathematics teachers. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey, Mexico: International Commission on Mathematical Instruction and the International Association for Statistical Education. Retrieved October 12, 2008, from http://www.ugr.es/~icmi/iase_study/

Artzt, A. F., Curcio, R. F., & Sultan, A. (2004). The design and implementation of an innovative four-year secondary mathematics teacher preparation program. In V. V. Orlov (Ed.), *Problems of theory and practice in teaching mathematics* (pp. 61-71). St. Petersburg, Russia: Hertzen State Pedagogical University.

Asiala, M., Brown, A., Devries, D. J., Dubinsky, E., Mathews, D., & Thomas, K. (1996). A framework for research in curriculum development in undergraduate mathematics education. In J. Kaput, A. Schoenfeld, & E. Dubinsky (Eds.), *CBMS issues in mathematics education: Vol. 6. Research in collegiate mathematics education II* (Vol. 6, pp. 1-32). Providence, RI: American Mathematical Society.

Bailar, B. A. (1988). Statistical practice and research: The essential interactions. *Journal of the American Statistical Association, 83*(401), 1-8.

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147-168). Dordrecht, The Netherlands: Kluwer.

Baldwin, W. (2000). Information no one else knows: The value of self-report. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. A. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 3-8). Mahwah, NJ: Erlbaum.

Ball, D. L. (1990). Reflections and deflections of policy: The case of Carol Turner. *Educational Evaluation and Policy Analysis, 12*, 247-259.

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes & L. Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3-32). San Francisco, CA: Jossey-Bass.

Batanero, C., Burrill, G., Reading, C., & Rossman, A. (Eds.). (2008). *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey, Mexico: International Commission on Mathematical Instruction and the

International Association for Statistical Education. Retrieved October 12, 2008, from http://www.ugr.es/~icmi/iase_study/

Batanero, C., Garfield, J., Ottaviani, M. G., & Truran, J. (2000). Research in statistical education: Some priority questions. *Statistical Education Research Newsletter, 1*(2), 2-6.

Baumgartner, L. M. (2001). An update on transformational learning. *New Directions for Adult and Continuing Education, 89*, 15-24.

Behr, M., Harel, G., Post, T., & Lesh, R. A. (1992). Rational number, ratio, and proportion. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 296-333). Reston, VA: National Council of Teachers of Mathematics.

Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal, 3*(2), 42-63.

Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics, 45,* 35-65.

Ben-Zvi, D., & Friedlander, A. (1997). Statistical thinking in a technological environment. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 International Association for Statistics Education (IASA) Round Table Conference* (pp. 45-55). The Netherlands: International Statistical Institute.

Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 3-16). Dordrecht, The Netherlands: Kluwer.

Ben-Zvi, D., Gil, E., & Apel, N. (2007, August). What is hidden beyond the data? Young students' reason and argue about some wider universe. In D. Pratt & J. Ainley (Eds.), *Reasoning about informal inferential statistical reasoning: A collection of current research studies. Proceedings of the fifth international research forum on Statistical Reasoning, Thinking, and Literacy*. Warwick, UK: University of Warwick.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. New York: Academic Press.

Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and the quality of intelligent behavior. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 57-76). Hillsdale, NJ: Erlbaum.

Bock, D. E., Velleman, P. F., & DeVeaux, R. D. (2004). *Stats: Modeling the world.* Boston: Pearson Education.

Bock, D. E., Velleman, P. F., & DeVeaux, R. D. (2006). *Stats: Modeling the world* (2nd ed.). Boston: Pearson Education.

Bohan, J. F. (2000). *AP Statistics: Preparing for the Advanced Placement examination*. New York: Amsco.

Borman, K., & Kersaint, G. (2002). *Assessing the impact of the National Science Foundation's Urban Systemic Initiative*. Tampa, FL: University of South Florida.

Borovenik, M. (2005, February). *Probabilistic and statistical thinking*. Paper presented at the fourth conference of the European Society for Research in Mathematics Education. Sant Feliu de Guixols, Spain.

Boyle, B., White, D., & Boyle, T. (2004). A longitudinal study of teacher change: What makes professional development effective? *The Curriculum Journal, 15*, 45-68.

Brewer, W. F. (1994). Autobiographical memory and survey research. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 11-20). New York: Springer-Verlag.

Broekmann, I. (1998). *The Ratlor effect.* Unpublished short paper presented at the 22nd conference of the International Group for the Psychology of Mathematics Education, Stellenbosch, South Africa.

Brookfield, S. (1990). Using critical incidents to explore learners' assumptions. In J. Mezirow and Associates (Eds.), *Fostering critical reflection in adulthood: A guide to transformative and emancipatory learning* (pp. 177-193). San Francisco, CA: Jossey-Bass.

Burrill, G. (2008). Fundamental ideas in teaching statistics and how they affect the training of teachers. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey, Mexico: International Commission on Mathematical Instruction and the International Association for Statistical Education. Retrieved October 12, 2008, from http://www.ugr.es/~icmi/iase_study/

Burrill, G., Franklin, C., Godbold, L., & Young, L. J. (2003). *Navigating through data analysis in grades 9-12*. Reston, VA: National Council of Teachers of Mathematics.

Butterfield, L. D., Borgen, W. A., Amundson, N. E., & Maglio, A.S. T. (2005). Fifty years of the critical incident technique: 1954-2004 and beyond. *Qualitative Research, 5*, 475-497.

Callingham, R. A. (1997). Teachers' multimodal functioning in relation to the concept of average. *Mathematics Education Research Journal, 9*, 205-224.

Callingham, R. A., Watson, J. M., Collis, K. F., & Moritz, J. B. (1995). Teacher attitudes towards chance and data. In B. Atweh & S. Flavel (Eds.), *Proceedings of the eighteenth annual conference of the Mathematics Education Research Group of Australasia* (pp. 143-150). Darwin, NT: Mathematics Education Research Group of Australasia.

Canada, D. (2004). *Elementary preservice teachers' conceptions of variation.* Unpublished doctoral dissertation, Portland State University, Portland.

Cantwell, R. H., & Scevak, J. (2004). *Discrepancies between the "ideal" and "passable" doctorate: Supervisor thinking on doctoral standards.* Paper presented at the annual conference of the Australian Association for Research in Education, Melbourne.

Chance, B., & Rossman, A. (2006). *Investigating statistical concepts, applications, and methods*. Belmont, CA: Duxbury.

Chatfield, C. (1988). *Problem solving: A statistician's guide*. London: Chapman and Hall.

Clark, J. M., Kraut, G., Mathews, D., & Wimbish, J. (2007). *The "fundamental theorem" of statistics: Classifying student understanding of basic statistical concepts*. Unpublished manuscript.

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*, 947-967.

Cobb, G. W. (2005). *The introductory statistics course: A saber tooth curriculum*. Plenary talk presented at the United States Conference on Teaching Statistics, Columbus, OH.

Cobb, G., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 204*, 801-823.

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning, 1*, 5-43.

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis, 12*, 327-345.

Cohen, D. K., & Ball, D. L. (1990). Policy and practice: An overview. *Educational Evaluation and Policy Analysis, 12*, 233-239.

Cohen, D. K., & Hill, H. C. (1998). *State policy and classroom performance: Mathematics reform in California*. Philadelphia, PA: Consortium for Policy Research in Education.

Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record, 102*, 294-343.

Cohen, D. K., & Hill, H. C. (2001). *Learning policy: When state education reform works*. New Haven, CT: Yale University Press.

Collegeboard.com Inc. (2007a). The Advanced Placement program. Retrieved March 12, 2007, from http://apcentral.collegeboard.com/apc/public/program/index.html

Collegeboard.com Inc. (2007b). AP Statistics course home page. Retrieved March 12, 2007, from http://apcentral.collegeboard.com/apc/public/courses/teachers_corner/2151.html

Collegeboard.com Inc. (2007c). Become an AP exam reader. Retrieved February 12, 2007, from http://apcentral.collegeboard.com/apc/public/teachers/opportunities/4137.html

Collegeboard.com Inc. (2007d). FAQs about the AP Reading. Retrieved March 12, 2007, from http://apcentral.collegeboard.com/apc/public/teachers/opportunities/4139.html

Collis, K. F., Romberg, T. A., & Jurdak, M. E. (1986). A technique for assessing mathematical problem-solving ability. *Journal for Research in Mathematics Education, 17*, 206-221.

Conference Board of the Mathematical Sciences. (2001). *CBMS issues in mathematics education: Vol. 11. The mathematical education of teachers*. Providence, RI: American Mathematical Society.

Coutinho, C. Q. S. (2008). Teaching statistics in elementary and high school and teacher training. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey, Mexico: International Commission on Mathematical Instruction and the International Association for Statistical Education. Retrieved October 12, 2008, from http://www.ugr.es/~icmi/iase_study/

Cranton, P. (1994). Self-directed and transformative instructional development. *The Journal of Higher Education, 65*, 726-744.

Cranton, P. (1996). *Professional development as transformative learning: New perspectives for teachers of adults*. San Francisco, CA: Jossey-Bass.

Cranton, P. (2006). *Understanding and promoting transformative learning: A guide for educators of adults*. San Francisco, CA: Jossey-Bass.

Crawford, F. M. (2005). *Characteristics of Advanced Placement Environmental Science Reading participants and their perceptions of the Reading as a professional development experience.* Unpublished doctoral dissertation, University of Florida, Gainesville.

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

Cuddapah, J. L. (2005). *Exploring first-year teacher learning through the lens of Mezirow's transformative learning theory (Jack Mezirow).* Unpublished doctoral dissertation, Teachers College, Columbia University, New York.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*, 382-393.

D'Ambrosio, B. S., Kastberg, S., McDermott, G., & Saada, N. (2004). Beyond reading graphs: Student reasoning with data. In P. Kloosterman & F. K. Lester, Jr. (Eds.), *Results and interpretations of the 1990 through 2000 mathematics assessments of the National Assessment of Educational Progress* (pp. 363-382). Reston, VA: National Council of Teachers of Mathematics.

Darling-Hammond, L., & Ball, D. (1998). *Teaching for high standards: What policymakers need to know and be able to do*. New York: Consortium for Policy Research in Education.

delMas, R. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 79-96). Dordrecht, The Netherlands: Kluwer.

delMas, R., Garfield, J., & Ooms, A. (2006). Using assessment items to study students' difficulty reading and interpreting graphical representations of distributions In K. Makar (Ed.), *Reasoning about distribution: A collection of current research studies. Proceedings of the fourth international research forum on Statistical Reasoning, Thinking, and Literacy*. Auckland, New Zealand: Statistical Reasoning, Thinking, and Literacy.

delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal, 4*(1), 55-82.

Department of Education Training and Youth Affairs. (2000). The impact of educational research. Retrieved April 25, 2007, from http://www.dest.gov.au/archive/highered/respubs/impact/pdf/impact.pdf

Derry, S. J., Levin, J. R., Osana, H. P., Jones, M. S., & Peterson, M. (2000). Fostering students' statistical and scientific thinking: Lessons learned from an innovative college course. *American Educational Research Journal, 37,* 747-773.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis, 24*, 81-112.

Eisenhower, D., Mathiowetz, N. A., & Morganstein, D. (1991). Recall error: Sources and bias reduction techniques. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 127-144). New York, NY: Wiley.

Erickson, D. M. (2007). A developmental re-forming of the phases of meaning in transformational learning. *Adult Education Quarterly, 58*, 61-80.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*, 403-434.

Finzer, W. (1999). Fathom Dynamic Data Software (Version 1.0) [Computer software]. Emeryville, CA: Key Curriculum Press

Finzer, W. (2006). Fathom Dynamic Data Software (Version 2.1) [Computer software]. Emeryville, CA: Key Curriculum Press.

Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*, 95-113.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education, 28*, 96-105.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327-358.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., et al. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework*. Alexandria, VA: American Statistical Association.

Franklin, C., & Mewborn, D. (2006). The statistical education of grades pre-K–12 teachers: A shared responsibility. In G. Burrill (Ed.), *Thinking and reasoning with data and chance: 68th NCTM yearbook* (pp. 335-344). Reston, VA: National Council of Teachers of Mathematics.

Freedman, D., Thornton, A., Camburn, D., Alwin, D., & Young-DeMarco, L. (1988). The life history calendar: A technique for collecting retrospective data. *Sociological Methodology, 18*, 37-68.

Friel, S. N. (2008). The research frontier: Where technology interacts with the teaching and learning of data analysis and statistics. In G. W. Blume & M. K. Heid (Eds.), *Research on technology in the learning and teaching of mathematics: Cases and perspectives* (pp. 279-332). Greenwich, CT: Information Age.

Friel, S. N., Curcio, R. F., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*, 124-158.

Gal, I., & Garfield, J. B. (1997). Curricular goals and assessment challenges in statistics education. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 1-13). Amsterdam: IOS Press.

Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal, 4*(1), 92-99.

Garfield, J., & Ben-Zvi, D. (2008). Preparing school teachers to develop students' statistical reasoning. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey, Mexico: International Commission on Mathematical Instruction and the International Association for Statistical Education. Retrieved October 12, 2008, from http://www.ugr.es/~icmi/iase_study/

Garfield, J. B., & Ben-Zvi, D. (with Chance, B., Medina, E., Roseth, C., & Zieffler, A.). (2008). *Developing students' statistical reasoning: Connecting research and teaching practice,* New York: Springer.

Garfield, J., delMas, R., & Chance, B. (2007). Using students' informal notions of variability to develop an understanding of formal measures of variability. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 117-148). New York, NY: Erlbaum.

Gfeller, M. K., Niess, M. L., & Lederman, N. G. (1999). Preservice teachers' use of multiple representations in solving arithmetic mean problems. *School Science and Mathematics, 99*, 250-257.

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

Goodell, J. (2000). Learning to teach mathematics for understanding: The role of reflection. *Mathematics Teacher Education and Development*, *2,* 48-60.

Goos, M., Dole, S., & Makar, K. (2007). Designing professional development to support teachers' learning in complex environments. *Mathematics Teacher Education and Development* [Special Issue], *8,* 23-47.

Graven, M. (2004). Investigating mathematics teacher learning within an in-service community of practice: The centrality of confidence. *Educational Studies in Mathematics, 57*, 177-211.

Gould, R. (2004). Variability: One statistician's view. *Statistics Education Research Journal, 3*(2), 7-16.

Groth, R. (2003). High school students' levels of thinking in regard to statistical study design. *Mathematics Education Research Journal, 15*, 252-269.

Groth, R. E. (2005). An investigation of statistical thinking in two different contexts: Detecting a signal in a noisy process and determining a typical value. *Journal of Mathematical Behavior, 24,*109-124.

Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education, 38*, 427-437.

Groth, R. E., & Bergner, J. A. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median, and mode. *Mathematical Thinking and Learning, 8*, 37-63.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice, 8*, 381-391.

Habermas, J. (1971). *Knowledge and human interests*. Boston: Beacon Press.

Habermas, J. (1984). *The theory of communicative action. Vol. 1: Reason and the rationality of society*. Boston: Beacon Press.

Hammerman, J. K., & Rubin, A. (2004). Strategies for managing statistical complexity with new software tools. *Statistics Education Research Journal, 3*(2), 17-41.

Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic enquiry with data: Critical barriers to classroom implementation. *Educational Psychologist, 27*, 337-364.

Hattie, J., & Purdie, N. (2003). The power of the SOLO Model to address fundamental measurement issues. Retrieved April 14, 2007, from http://coe.winthrop.edu/ncate/links_from_the_menu/stan2.linkedpages/stan2.B.3.html

Heaton, R. M., & Mickelson, W. T. (2002). The learning and teaching of statistical investigation in teaching and teacher education. *Journal of Mathematics Teacher Education, 5*, 35-59.

Heid, M. K., Perkinson, D., Peters, S. A., & Fratto, C. L. (2005). Making and managing distinctions: The case of sampling distributions. In G. M. Lloyd, M. Wilson, J. L. M. Wilkins, & S. Behm (Eds.), *Proceedings of the 27th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Roanoke, VA: Virginia Polytechnic Institute and State University.

Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's Mathematics Professional Development Institutes. *Journal for Research in Mathematics Education, 35*, 330-351.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*, 371-406.

Hoekstra, A., Beijaard, D., Brekelmans, M., & Korthagen, F. (2007). Experienced teachers' informal learning from classroom teaching. *Teachers and Teaching: Theory and Practice, 13*, 189-206.

Hogg, R.V., & Tanis, E.A. (2001), *Probability and Statistical Inference* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Horvath, J. K., & Lehrer, R. (1998). A model-based perspective on the development of children's understanding of chance and uncertainty. In S. P. LaJoie (Ed.), *Reflections on statistics: Agendas for learning, teaching and assessment in K-12* (pp. 121-148). Mahwah, NJ: Elrbaum.

Imants, J., & van Veen, K. (in press). Teacher learning as workplace learning. In N. Verloop (Ed.), *International encyclopedia on education* (3rd ed).

Jack Mezirow and Associates (Eds.). (2000). *Learning as transformation: Critical perspectives on a theory in progress*. San Francisco, CA: Jossey-Bass.

Jacobs, V. R. (1999). How do students think about statistical sampling before instruction? *Mathematics Teaching in the Middle School, 5*, 240-263.

Karpadia, R. (1980). Developments in statistical education. *Educational Studies in Mathematics, 11*, 443-461.

Kegan, R. (2000). What "form" transforms? A constructive-developmental approach to transformative learning. In J. Mezirow and Associates (Eds.), *Learning as transformation: Critical perspectives on a theory in progress* (pp. 35-70). San Francisco, CA: Jossey-Bass.

Kelchtermans, G., & Vandenberghe, R. (1994). Teachers' professional development: A biographical perspective. *Journal of Curriculum Studies, 26*, 45-62.

Kelly, B. A., & Watson, J. M. (2002). Variation in a chance sampling setting: The lollies task. In B. Barton, K. C. Irwin, M. Pfannkuch, & M. O. J. Thomas (Eds.), *Mathematics education in the South Pacific. Proceedings of the 25th annual conference of the Mathematics Education Research Group of Australasia, Auckland* (pp. 366-373). Sydney: Mathematics Education Research Group of Australasia.

Kennedy, R. L., & Wyrick, A. M. (1995). *Teaching as reflective practice.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Biloxi, MS.

Kilpatrick, J. (1985). Reflection and recursion. *Educational Studies in Mathematics, 16*, 1-26.

King, K. P. (2000). *Educational technology that transforms: Educators' transformational learning experiences in professional development*. Paper presented at the Annual Adult Education and Research Conference, University of British Columbia. Retrieved March 15, 2009, from http://www.edst.educ.ubc.ca/aerc/2000/kingk-final.PDF

King, K. P. (2002). *A journey of transformation: A model of educators' learning experiences in educational technology*. Raleigh, NC: Adult and Community College Education.

King, K. P. (2004). Both sides now: Examining transformative learning and professional development of educators. *Innovative Higher Education, 29*, 155-174.

Kitchener, K. S., & King, P. M. (1990). The reflective judgment model: Transforming assumptions about knowing. In J. Mezirow and Associates (Eds.), *Fostering critical reflection in adulthood* (pp. 159-176). San Francisco, CA: Jossey-Bass.

Kitchenham, A. (2006). Teachers and technology: A transformative journey. *Journal of Transformative Education, 4*, 202-225.

Knowles, M. S. (1984). *Andragogy in action: Applying modern principles of adult learning*. San Francisco: Jossey-Bass.

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers and Mathematics Learning, 12*, 217-230.

Konold, C., & Higgins, T. L. (2002). Highlights of related research. In S. J. Russell, D. Schifter, & V. Bastable (Eds.), *Working with data* (pp. 165-201). Parsippany, NJ: Dale Seymour.

Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 193-215). Reston, VA: National Council of Teachers of Mathematics.

Konold, C., & Miller, C. (2004). TinkerPlots Dynamic Data Exploration [Computer software]. Emeryville, CA: Key Curriculum Press.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*, 259-289.

Konold, C., & Pollatsek, A. (2004). Conceptualizing an average as a stable feature of a noisy process. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 169-200). Dordrecht, The Netherlands: Kluwer.

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM, 40*, 873-892.

Lann, A., & Falk, R. (2003). What are the clues for intuitive assessment of variability? In C. Lee (Ed.), *Reasoning about variability: A collection of research studies. Proceedings of the third international research forum on Statistical Reasoning, Thinking, and Literacy*. Mount Pleasant: Central Michigan University.

Lee, H. S., & Hollebrands, K. F. (2008). Preparing to teach data analysis and probability with technology. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey, Mexico: International Commission on Mathematical Instruction and the International Association for Statistical Education. Retrieved October 12, 2008, from http://www.ugr.es/~icmi/iase_study/

Leikin, R., & Zazkis, R. (2007). A view on the teachers' opportunities to learn mathematics through teaching. In J.-H. Woo, H.-C. Lew, K.-S. Park, & D.-Y. Seo (Eds.), *Proceedings of the 31st conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 122-127). Seoul, Korea: The Korea Society of Educational Studies in Mathematics.

Lin, P. (2002). On enhancing teachers' knowledge by constructing cases in classrooms. *Journal of Mathematics Teacher Education, 5,* 317-349.

Lipson, K. (2002). The role of computer based technology in developing understanding of the concept of sampling distribution. In B. Philips (Ed.), *Developing a statistically literate society. Proceedings of the sixth International Conference on Teaching Statistics*. Cape Town, South Africa.

Liu, Y., & Thompson, P. W. (2005). Teachers' understanding of hypothesis testing. In G. M. Lloyd, M. Wilson, J. L. M. Wilkins, & S. L. Behm (Eds.), *Proceedings of the 27th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Roanoke, VA: Virginia Polytechnic Institute and State University.

Liu, Y., & Thompson, P. W. (2009). Mathematics teachers' understandings of proto-hypothesis testing. *Pedagogies, 4*, 126-138.

Loftus, E. F. (2000). Suggestion, imagination, and the transformation of reality. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. A. Cain (Eds.), *The science* *of self-report: Implications for research and practice* (pp. 201-210). Mahwah, NJ: Erlbaum.

Lohman, M. C., & Woolf, N. H. (2001). Self-initiated learning activities of experienced public school teachers: Methods, sources, and relevant organizational influences. *Teachers and Teaching: Theory and Practice, 7*, 59-74.

Loosen, F., Lioen, M., & Lacante, M. (1985). The standard deviation: Some drawbacks of an intuitive approach. *Teaching Statistics, 7*, 2-5.

Lortie, D. (1975). *School teacher: A sociological study*. Chicago: University of Chicago Press.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Erlbaum.

MacCullough, D. L. (2007). *A study of experts' understanding of arithmetic mean.* Unpublished doctoral dissertation, The Pennsylvania State University, State College.

Makar, K. (2004). *Developing statistical inquiry: Prospective secondary mathematics and science teachers' investigations of equity and fairness through analysis of accountability data.* Unpublished doctoral dissertation, University of Texas at Austin, Austin.

Makar, K., & Confrey, J. (2002). *Comparing two distributions: Investigating secondary teachers' statistical thinking.* In B. Philips (Ed.), *Developing a statistically literate society. Proceedings of the sixth International Conference on Teaching Statistics*. Cape Town, South Africa.

Makar, K., & Confrey, J. (2003). Chunks, clumps, and spread out: Secondary preservice teachers' informal notions of variation and distribution. In C. Lee (Ed.), *Reasoning about variability: A collection of research studies. Proceedings of the third international research forum on Statistical Reasoning, Thinking, and Literacy*. Mount Pleasant: Central Michigan University.

Makar, K., & Confrey, J. (2004). Secondary teachers' statistical reasoning in comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 353-374). Dordrecht, The Netherlands: Kluwer.

Makar, K., & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27-54.

Marsick, V. J. (1990). Action learning and reflection in the workplace. In J. Mezirow and Associates (Eds.), *Fostering critical reflection in adulthood* (pp. 23-46). San Francisco, CA: Jossey-Bass.

Marsick, V. J., & Watkins, K. E. (2001). Informal and incidental learning. *New Directions for Adult and Continuing Education, 89,* 25-34.

Martyn, K. K., & Belli, R. F. (2002). Retrospective data collection using event history calendars. *Nursing Research, 51*, 270-274.

Masnick, A. M., & Klahr, D. (2001). *Elementary school children's understanding of experimental error*. Paper presented at the 23rd annual conference of the Cognitive Science Society, Edinburgh, Scotland.

Mathews, D., & Clark, J. M. (2007). Successful students' conceptions of mean, standard deviation, and the central limit theorem. Unpublished manuscript.

McClain, K. (2005). The evolution of teachers' understandings of statistical data analysis: A focus on distribution. In K. Makar (Ed.), *Reasoning about distribution: A collection of current research studies. Proceedings of the fourth international research forum on Statistical Reasoning, Thinking, and Literacy*. Auckland, New Zealand: Statistical Reasoning, Thinking, and Literacy.

McGraw, R., Lynch, K., Koc, Y., Buday, A., & Brown, C. (2007). The multimedia case as a tool for professional development: An analysis of online and face-to-face interaction among mathematics pre-service teachers, in-service teachers, mathematicians, and mathematics teacher educators. *Journal for Mathematics Teacher Education, 10,* 95-121.

Means, B., Swan, G. E., Jobe, J. B., & Esposito, J. L. (1991). An alternative approach to obtaining personal history data. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 167-184). New York: Wiley.

Meijer, P. C., Verloop, N., & Beijaard, D. (1999). Exploring language teachers' practical knowledge about teaching reading comprehension. *Teaching and Teacher Education, 15*, 59-84.

Meletiou-Mavrotheris, M., & Lee, C. (2003). Studying the evolution of students' conceptions of variation using the transformative and conjecture-driven research design. In C. Lee (Ed.), *Reasoning about variability: A collection of research studies. Proceedings of the third international research forum on Statistical Reasoning, Thinking, and Literacy*. Mount Pleasant: Central Michigan University.

Meletiou-Mavrotheris, M., & Stylianou, D. A. (2003). *On the formalist view of mathematics: Impact on statistics instruction and learning*. Paper presented at the third conference of the European Society for Research in Mathematics Education, Bellaria, Italia.

Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco, CA: Jossey-Bass.

Merriam, S. B. (2001). Andragogy and self-directed learning: Pillars of adult learning theory. In S. B. Merriam (Ed.), *The new update on adult learning theory* (pp. 3-14). San Francisco, CA: Jossey-Bass.

Merriam, S. B. (2004a). The changing landscape of adult learning theory. *Review of Adult Learning and Literacy: Connecting Research, Policy, and Practice, 4*, 199-220.

Merriam, S. B. (2004b). The role of cognitive development in Mezirow's transformational learning theory. *Adult Education Quarterly, 55*, 60-68.

Merriam, S. B., & Caffarella, R. S. (1999). *Learning in adulthood: A comprehensive guide*. San Francisco, CA: Jossey-Bass.

Mezirow, J. (1985). A critical theory of self-directed learning. In S. Brookfield (Ed.), *Self-directed learning: From theory to practice* (Vol. 25, pp. 17-30). San Francisco, CA: Jossey-Bass.

Mezirow, J. (1990). How critical reflection triggers transformative learning. In J. Mezirow and Associates (Eds.), *Fostering critical reflection in adulthood* (pp. 1-20). San Francisco, CA: Jossey-Bass.

Mezirow, J. (1991). *Transformative dimensions of adult learning*. San Francisco, CA: Jossey-Bass.

Mezirow, J. (1994). Understanding transformation theory. *Adult Education Quarterly, 44*, 222-232.

Mezirow, J. (2000). Learning to think like an adult: Core concepts of transformation theory. In J. Mezirow and Associates (Eds.), *Learning as transformation: Critical perspectives on a theory in progress* (pp. 3-34). San Francisco, CA: Jossey-Bass.

Mezirow, J. (2003). Transformative learning as discourse. *Journal of Transformative Education, 1*, 58-63.

Moerer-Urdahl, T., & Creswell, J. (2004). Using transcendental phenomenology to explore the "ripple effect" in a leadership mentoring program [Electronic Version]. *International Journal of Qualitative Methods*, 3, Article 2. Retrieved February 12, 2007, from http://www.ualberta.ca/!iiqm/backissues/3_2/pdf/moerercreswell.pdf

Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education, 26*, 20-39.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*, 125-145.

Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95-137). Washington, DC: National Academy Press.

Moore, D. S. (1997). Probability and statistics in the core curriculum. In J. Dossey (Ed.), *Confronting the core curriculum* (pp. 93-98). Washington, DC: Mathematical Association of America.

Moore, D. S. (1998). Statistics among the liberal arts. *Journal of the American Statistical Association, 93*, 1253-1259.

Moore, D. S. (1999). Discussion: What shall we teach beginners? *International Statistical Review, 67*, 250-252.

Moore, D. S., & Cobb, G. (2000). Statistics and mathematics: Tension and cooperation. *The American Mathematical Monthly, 107*, 615-630.

Moustakas, C. (1994). *Phenomenological research methods*. Thousand Oaks, CA: Sage.

Nathan, M. J., & Knuth, E. J. (2003). A Study of whole classroom mathematical discourse and teacher change. *Cognition and Instruction, 21*, 175-207.

National Assessment Governing Board [NAGB]. (2004). Mathematics framework for the 2005 National Assessment of Educational Progress. Retrieved April 9, 2007, from http://www.nagb.org/pubs/m_framework_05/chap2.html

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Peck, R., Olsen, C., & Devore, J. (2001). *Introduction to statistics and data analysis.* Pacific Grove, CA: Duxbury.

Peck, R., Olsen, C., & Devore, J. (2005). *Introduction to statistics and data analysis* (2nd ed.). Belmont, CA: Duxbury.

Pegg, J. (2003). Assessment in mathematics: A developmental approach. In J. M. Royer (Ed.), *Mathematical cognition: A volume in current perspectives on cognition, learning, and instruction* (pp. 227-259). Greenwich, CT: Information Age.

Pegg, J., & Tall, D. (2001). Fundamental cycles in learning algebra: An analysis. In H. Chick, K. Stacey, J. Vincent, & J. Vincent (Eds.), *The future of the teaching and learning of algebra: Proceedings of the 12th ICMI study conference*. Melbourne: The University of Melbourne.

Pegg, J., & Tall, D. (2004). Fundamental cycles of cognitive growth. In A. D. Cockburn & E. Nardi (Eds.), *Proceedings of the 26th conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 369-376). Norwich, UK: University of East Anglia.

Pegg, J., & Tall, D. (2005). The fundamental cycle of concept construction underlying various theoretical frameworks. *International Reviews on Mathematical Education (Zentralblatt fur Didaktik der Mathematik), 37*, 468-475.

Peters, J. (1991). The Action-Reason-Thematic technique: Spying on the self. In J. Mezirow and Associates (Eds.), *Fostering critical reflection in adulthood: A guide to transformative and emancipatory learning* (pp. 314-335). San Francisco, CA: Jossey-Bass.

Peters, S. (in press). Engaging with the art and science of statistics. *Mathematics Teacher.* Reston, VA: National Council of Teachers of Mathematics.

Peterson, P. L. (1990). Doing more in the same amount of time: Cathy Swift. *Educational Evaluation and Policy Analysis, 12*, 261-280.

Petrosino, A. J., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning, 5*, 131-156.

Pfannkuch, M. (1997). Statistical thinking: One statistician's perspective. In F. Biddulph & K. Carr (Eds.), *People in mathematics education. Proceedings of the twentieth annual conference of the Mathematics Research Group of Australasia* (Vol. 2, pp. 406-413). Rotorua, New Zealand: Mathematics Research Group of Australasia.

Pfannkuch, M. (2005). Probability and statistical inference: How can teachers enable learners to make the connection? In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 267-294). New York: Springer.

Pfannkuch, M., & Wild, C. J. (2000). Statistical thinking and statistical practice: Themes gleaned from professional statisticians. *Statistical Science*, *15*, 132-152.

Polettini, A. F. F. (2000). Mathematics teaching life histories in the study of teachers' perceptions of change. *Teaching and Teacher Education, 16,* 765-783.

Pollatsek, A., Lima, S., & Well, A. D. (1981). Concept or computation: Students' understanding of the mean. *Educational Studies in Mathematics, 12*, 191-204.

Poulson, L., & Avramidis, E. (2003). Pathways and possibilities in professional development: Case studies of effective teachers of literacy. *British Educational Research Journal, 29*, 543-560.

Pressler, M. W. (2006, October 11). The handwriting is on the wall: Researchers see a downside as keyboards replace pens in schools. *The Washington Post,* p. A01.

Prodromou, T., & Pratt, D. (2006). The role of causality in the co-ordination of two perspectives on distribution within a virtual simulation. *Statistics Education Research Journal*, *5*(2), 69-88.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*(1), 4-15.

Research and Education Association. (1997). *AP Statistics*. Piscataway, NJ: Author.

Reading, C. (2004). Student description of variation while working with weather data. *Statistics Education Research Journal, 3*(2), 84-105.

Reading, C., & Reid, J. (2004). *Consideration of variation: A model for curriculum development*. Paper presented at the International Association for Statistical Education roundtable on curricular development in statistics education, Lund, Sweden.

Reading, C., & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: From a variation perspective. *Statistics Education Research Journal, 5*(2), 46-68.

Reading, C., & Shaughnessy, J. M. (2000). *Student perceptions of variation in a sampling situation.* In T. Hakahar & M. Koyama (Eds.), *Proceedings of the 24th conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 89-96). Hiroshima, Japan.

Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 201-226). Dordrecht, The Netherlands: Kluwer.

Rees, K. (2007). Life as an "Acorn": My experience as a first-time AP Statistics reader. Retrieved March 12, 2007, from http://apcentral.collegeboard.com/apc/members/features/9927.html

Reid, J., & Reading, C. (2005). Developing consideration of variation: Case studies from a tertiary introductory service statistics course. In *Proceedings of the 55th session of the International Statistical Institute, Sydney.* Voorburg, The Netherlands: International Statistical Institute.

Reid, J., & Reading, C. (2006). A hierarchy of tertiary students' consideration of variation. In A. Rossman & B. Chance (Eds.), *Proceedings of the seventh International Conference on Teaching Statistics, Salvador.* Voorburg, The Netherlands: International Statistical Institute.

Reid, J., & Reading, C. (2008). Measuring the development of students' consideration of variation. *Statistics Education Research Journal, 7*(1), 40-59.

Richardson, V. (1998, September). How teachers change: What will lead to change that most benefits students learning? *Focus on Basics: Connecting Research & Practice, 2*(C) Retrieved May 6, 2009, from http://www.ncsall.net/?id=395

Richardson, V., & Placier, P. (2001). Teacher change. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 905-947). Washington, DC: American Educational Research Association.

Rider, R. L. (2004). *The effect of multi-representational methods on students' knowledge of function concepts in developmental college mathematics.* Unpublished doctoral dissertation, North Carolina State University, Raleigh.

Roddick, C., Becker, J. R., & Pence, B. J. (2000). Capstone courses in problem solving for prospective secondary teachers: Effects on beliefs and teaching practices. In T. Nakahara & M. Koyama (Eds.), *Proceedings of the 24th conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 97-104). Hiroshima: Hiroshima University.

Rogers, M. P., Abell, S., Lannin, J., Wang, C-Y., Musikul, K., Barker, D., et al. (2007). Effective professional development in science and mathematics education: Teachers' and facilitators' views. *International Journal of Science and Mathematics Education, 5*, 507-532.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review, 96*, 341-357.

Rossman, A. J. (1997). *Workshop statistics: Discovery with data*. New York: Springer-Verlag.

Rossman, A. J., Chance, B. L., & Lock, R. H. (2001). *Workshop statistics: Discovery with data and Fathom.* Emeryville, CA: Key College.

Rubin, A., Bruce, B., & Tenney, Y. (1990). *Learning about sampling: Trouble at the core of statistics.* In D. Vere-Jones (Ed.), *Proceedings of the third International Conference on Teaching Statistics* (Vol. 1, pp. 314-319). Voorburg, The Netherlands: International Statistical Institute.

Rubin, A., Hammerman, J., Campbell, C., & Puttick, G. (2005). The effect of distributional shape on group comparison strategies. In K. Makar (Ed.), *Reasoning about distribution: A collection of current research studies. Proceedings of the fourth international research forum on Statistical Reasoning, Thinking, and Literacy*. Auckland, New Zealand: Statistical Reasoning, Thinking, and Literacy.

Saavedra, E. (1996). Teachers study groups: Contexts for transformative learning and action. *Theory into Practice, 35*, 271-277.

Saldanha, L. A., & Thompson, P. W. (2001). Students' reasoning about sampling distribution and statistical inference. In R. Speiser & C. Maher (Eds.), *Proceedings of the 23rd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. Vol. 1, pp. 449-454). Snowbird, Utah: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Saldanha, L. A., & Thompson, P. W. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*, 257-270.

Scheaffer, R. L. (2006). Statistics and mathematics: On making a happy marriage. In G. Burrill (Ed.), *Thinking and reasoning with data and chance: 68th NCTM yearbook* (pp. 309-321). Reston, VA: National Council of Teachers of Mathematics.

Schorr, R. Y., Firestone, W. A., & Monfils, L. (2003). State testing and mathematics teaching in New Jersey: The effects of a test without other supports. *Journal for Research in Mathematics Education, 34,* 373-405.

Schwarz, N., Hippler, H. J., & Noelle-Neumann, E. (1994). Retrospective reports: The impact of response formats. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 187-202). New York, NY: Springer-Verlag.

Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. New York, NY: Teachers College Press.

Serow, P. (2006). Triangle-property relationships: Making the connections. In J. Novotna, H. Moravova, M. Kratka, & N. Stehlikova (Eds.), *Proceedings of the 30th conference of the International Group for the Psychology of Mathematics Education* (Vol. 5, pp. 89-969). Prague: Psychology of Mathematics Education.

Shaughnessy, J. M. (1992). Research on probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddulph & K. Carr (Eds.), *People in mathematics education. Proceedings of the twentieth annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 6-22). Waikato, NZ: Mathematics Education Group of Australasia.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester, Jr. (Ed.), *Handbook of research on mathematics teaching and learning* (2nd ed., pp. 957-1009). Greenwich, CT: Information Age.

Shaughnessy, J. M., Canada, D., & Ciancetta, M. (2003). Middle school students' thinking about variability in repeated trials: A cross-task comparison. In N. A. Pateman, B. J. Dougherty, & U. Zilliox (Eds.), *Proceedings of the joint meeting of PME 27 and PME-NA XXV* (Vol. 4, pp. 159-165). Honolulu, HI: University of Hawaii.

Shaughnessy, J. M., Ciancetta, M., Best, K., & Canada, D. (2004). *Students' attention to variability when comparing distributions.* Paper presented at the Research Presession of the 82nd annual meeting of the National Council of Teachers of Mathematics, Philadelphia, PA.

Shaughnessy, J. M., Ciancetta, M., & Canada, D. (2004). Types of student reasoning on sampling tasks. In M. J. Hoines & A. B. Fuglestad (Eds.), *Proceedings of the 28th conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 177-184). Bergen, Norway: Psychology of Mathematics Education.

Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 205-238). Dordrecht, The Netherlands: Kluwer.

Shields, P. M., Marsh, J. A., & Adelman, N. E. (1998). *The SSI's impacts on classroom practice*. Menlo Park, CA: SRI International.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Silva, C. B., & Coutinho, C. Q. S. (2006). The variation concept: A study with secondary school mathematics teachers. In A. Rossman & B. Chance (Eds.), *Proceedings of the seventh International Conference on Teaching Statistics, Salvador*. Voorburg, The Netherlands: International Statistical Institute.

Silva, C. B., & Coutinho, C. Q. S. (2008). Reasoning about variation of a univariate distribution: A study with secondary math teachers. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey, Mexico: International Commission on Mathematical Instruction and the International Association for Statistical Education. Retrieved October 12, 2008, from http://www.ugr.es/~icmi/iase_study/

Senger, E. S. (1998-1999). Reflective reform in mathematics: The recursive nature of teacher change. *Educational Studies in Mathematics, 37*, 199-221.

Smith, T. M., Desimone, L. M., & Ueno, K. (2005). "Highly qualified" to do what? The relationship between teacher quality mandates and the use of reform-oriented instruction in middle school mathematics. *Educational Evaluation and Policy Analysis, 27*, 75-109.

Snee, R. D. (1999). Discussion. Development and use of statistical thinking: A new era. *International Statistical Review, 67*, 255-258.

Sorto, M. A. (2004). *Prospective middle school teachers' knowledge about data analysis and its applications to teaching.* Unpublished doctoral dissertation, Michigan State University, Lansing.

Spielman, L. J., & Lloyd, G. M. (2004). The impact of enacted mathematics curriculum models on prospective elementary teachers' course perceptions and beliefs. *School Science and Mathematics, 104*, 32-44.

Spillane, J. P. (2000a). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction, 18*, 141-179.

Spillane, J. P. (2000b). A fifth-grade teacher's reconstruction of mathematics and literacy teaching: Exploring interactions among identity, learning, and subject matter. *Elementary School Journal, 100*, 307-330.

Spillane, J. P. (2002). Local theories of teacher change: The pedagogy of district policies and programs. *Teachers College Record, 104*, 377-420.

Stanage, S. M. (1987). *Adult education and phenomenological research: New directions for theory, practice, and research*. Malabar, FL: Robert E. Krieger.

Sternstein, M. (1998). *Barron's AP Statistics*. Hauppauge, NH: Barron's.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: The Free Press.

Tabak, J. (2005). *Probability and statistics: The science of uncertainty*. New York: Checkmark Books.

Tarr, J. E., & Shaughnessy, J. M. (2007). Student performance in data analysis, statistics, and probability. In P. Kloosterman & F. K. Lester, Jr. (Eds.), *Results and interpretations of* *the 2003 mathematics assessment of the National Assessment of Educational Progress* (pp. 139-168). Reston, VA: National Council of Teachers of Mathematics.

Taylor, E. W. (1997). Building upon the theoretical debate: A critical review of the empirical studies of Mezirow's transformative learning theory. *Adult Education Quarterly, 48*, 34-60.

Taylor, E. W. (1998). *The theory and practice of transformative learning: A critical review. Information series no. 374*. Columbus, OH: ERIC Clearinghouse on Adult, Career, and Vocational Education.

Taylor, E. W. (2000). Analyzing research on transformative learning theory. In J. Mezirow and Associates (Eds.), *Learning as transformation: Critical perspectives on a theory in progress* (pp. 285-328). San Francisco, CA: Jossey-Bass.

The College Board. (2004). *The College Board AP Advanced Placement program: Statistics course description*. New York: College Entrance Examination Board.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). Reston, VA: National Council of Teachers of Mathematics.

Thompson, C. L., & Zeuli, J. S. (1999). The frame and the tapestry: Standards-based reform and professional development. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 341-375). San Francisco, CA: Jossey-Bass.

Tourangeau, R. (2000). Remembering what happened: Memory errors and survey reports. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. A. Cain (Eds.), *The Science of self-report: Implications for research and practice* (pp. 29-48). Mahwah, NJ: Erlbaum.

Van der Vaart, W., Van der Zouwen, J., & Dijkstra, W. (1995). Retrospective questions: Data quality, task difficulty, and the use of a checklist. *Quality & Quantity, 29*, 299-315.

van Driel, J. H., Verloop, N., & de Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching, 35*, 673-695.

Von Glasersfeld, E. (1990). An exposition of constructivism: Why some like it radical. *Constructivist views on the teaching and learning of mathematics. Journal for Research in Mathematics Education: Monograph No. 4* (pp. 19-30). Reston, VA: National Council of Teachers of Mathematics.

Watson, J. M. (1998). Professional development for teachers of probability and statistics: Into an era of technology. *International Statistical Review, 66*, 271-289.

Watson, J. M. (2000). Preservice mathematics teachers' understanding of sampling: Intuition or mathematics. *Mathematics Teacher Education and Development, 2*, 121-135.

Watson, J. M. (2001). Profiling teachers' competence and confidence to teach particular mathematics topics: The case of chance and data. *Journal of Mathematics Teacher Education, 4*, 305-337.

Watson, J. M. (2002). Lessons from variation research I: Student understanding. In M. Goos & T. Spencer (Eds.), *Mathematics—Making waves. Proceedings of the nineteenth biennial conference of the Australian Association of Mathematics Teachers, Geelong* (pp. 261-268). Sydney: Mathematics Education Group of Australasia.

Watson, J. M., Callingham, R. A., & Kelly, B. A. (2007). Students' appreciation of expectation and variation as a foundation for statistical understanding. *Mathematical Thinking and Learning, 9*, 83-130.

Watson, J. M., Collis, K. F., Callingham, R. A., & Moritz, J. B. (1995). A model for assessing higher order thinking in statistics. *Educational Research and Evaluation: An International Journal on Theory and Practice, 1*, 247-275.

Watson, J. M., & Kelly, B. A. (2002a). *Can grade 3 students learn about variation?* In B. Philips (Ed.), *Developing a statistically literate society. Proceedings of the sixth International Conference on Teaching Statistics*. Cape Town, South Africa.

Watson, J. M., & Kelly, B. A. (2002b). Grade 5 students' appreciation of variation. In A. D. Cockburn & E. Nardi (Eds.), *Proceedings of the 26th conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 385-392). Norwich, UK: University of East Anglia.

Watson, J. M., & Kelly, B. A. (2003a). Predicting dice outcomes: The dilemma of expectation versus variation. In L. Bragg, C. Campbell, G. Herbert, & J. Mousley (Eds.), *Mathematics education research: Innovation, networking, opportunity. Proceedings of the 26th annual conference of the Mathematics Education Research Group of Australasia* (pp. 728-735). Geelong: Mathematics Education Research Group of Australasia.

Watson, J. M., & Kelly, B. A. (2003b). Statistical variation in a chance setting. In N. A. Pateman, B. J. Dougherty, & U. Zilliox (Eds.), *Proceedings of the joint meeting of PME 27 and PME-NA XXV* (Vol. 4, pp. 387-394). Honolulu, HI: University of Hawaii.

Watson, J. M., & Kelly, B. A. (2004a). Expectation versus variation: Students' decision making in a chance environment. *Canadian Journal of Science, Mathematics, and Technology Education, 4*, 371-396.

Watson, J. M., & Kelly, B. A. (2004b). Statistical variation in a chance setting: A two-year study. *Educational Studies in Mathematics, 57*, 121-144.

Watson, J. M., & Kelly, B. A. (2005). The winds are variable: Student intuitions about variation. *School Science and Mathematics, 105*, 252-269.

Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology, 34*, 1-29.

Watson, J. M., & Moritz, J. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education, 31*, 44-70.

Weissglass, J. (1992). Changing the culture of mathematics instruction. *Journal of Mathematical Behavior, 11*, 195-203.

Weissglass, J. (1994). Changing mathematics teaching means changing ourselves: Implications for professional development. In D. B. Aichele (Ed.), *Professional development for teachers of mathematics* (pp. 67-78). Reston, VA: National Council of Teachers of Mathematics.

Wheatley, G. H. (1992). The role of reflection in mathematics learning. *Educational Studies in Mathematics, 23,* 529-541.

Whitelaw, C., Sears, M., & Campbell, K. (2004). Transformative learning in a faculty professional development context. *Journal of Transformative Education, 2*(1), 9-27.

Whitney, A. (2008). Teacher transformation in the National Writing Project. *Research in the Teaching of English, 43*(2), 144-187.

Wiemers, N. J. (1990). Transformation and accommodation: A case study of Joe Scott. *Educational Evaluation and Policy Analysis, 12*, 281-292.

Wild, C. J. (2005). A statistician's view on the concept of distribution. In K. Makar (Ed.), *Reasoning about distribution: A collection of current research studies. Proceedings of the fourth international research forum on Statistical Reasoning, Thinking, and Literacy*. Auckland, New Zealand: Statistical Reasoning, Thinking, and Literacy.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-265.

Wilensky, U. (1995). Learning probability through building computational models. In L. Meira & D. Carraher (Eds.), *Proceedings of the 19th conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 152-159). Recife, Brazil: Program Committee of the 19th PME Conference.

Wilensky, U. (1997). What is normal anyway? Therapy for epistemological anxiety. *Educational Studies in Mathematics, 33,* 171-202.

Wilson, S. M. (1990). A conflict of interests: The case of Mark Black. *Educational Evaluation and Policy Analysis, 12*, 293-310.

Yates, D. S., Moore, D. S., & McCabe, G. P. (1999). *The practice of statistics: TI-83 graphing calculator enhanced*. New York: W. H. Freeman.

Yates, D. S., Moore, D. S., & Starnes, D. S. (2003). *The practice of statistics: TI-83/89 graphing calculator enhanced* (2nd ed.). New York: W. H. Freeman.

Zawojewski, J. S., & Shaughnessy, J. M. (2000). Data and chance. In E. A. Silver & P. A. Kenney (Eds.), *Results from the seventh mathematics assessment of the National Assessment of Educational Progress* (pp. 235-268). Reston, VA: National Council of Teachers of Mathematics.

Appendix A

Initial Questionnaire

Please complete the following questionnaire as soon as possible, save the file, and return the file as an attachment to sap233@psu.edu.

Name: ____

Gender: _________ Male _________ Female

Age: _________

Current City and State of Residence: ____

Years Teaching: _________

Years Teaching Statistics: _________

Please check the category that applies to you.

Years as AP Statistics reader: _______ 1 _______ 2 _______ 3 _______ 4 _______ 5 _______ 6

Years as AP Statistics table leader: _______ 0 _______ 1 _______ 2 _______ 3 _______ 4 _______ 5 _______ 6

Please list the year of completion for all degrees that you have.

Undergraduate	Mathematics Education	Mathematics	Statistics	Other (please specify)	
Major					
Minor					

Graduate	Mathematics Education	Mathematics	Statistics	Other (please specify)	Degree Type (e.g., MA, MS, Ph.D.)
Major					
Minor					

Please check those items that apply to you.

Teaching Certification: _______ Subject Area(s) - Specify ______________________________
Grade Level(s) - Specify ______________________________
State(s) - Specify ______________________________

Other Certification: _______ Subject Area(s) - Specify ______________________________
Grade Level(s) - Specify ______________________________
State(s) - Specify ______________________________

Professional Development: _______ Attended non-AP professional development in statistics
Approximate number of experiences: _______

_______ Attended AP Statistics professional development
Approximate number of experiences: _______

_______ Conducted non-AP professional development in statistics
Approximate number of experiences: _______

_______ Conducted AP Statistics professional development
Approximate number of experiences: _______

Other Statistical Experiences: _______ Served as statistical consultant, non-AP - Specify ______________________

_______ Publication(s) related to statistics (authored) - Specify ______________________

_______ Provide a brief description of other statistical experiences, including publications read ______________________

Please specify the type of graphing calculator or statistical software with which you are most comfortable ______________________________

Appendix B

Sample Event History Calendar

	Year	1981				1982				1983				1984				1985				1986				1987				1988				1989			
Formal and Informal Learning Experiences	1 Undergrad education related to statistics																					X	X														
																						Intro prob/stat course															
	2 Graduate education related to statistics																																			X	X
																																		Math-stat course			
	3 Non-AP PD related to statistics																																				
	4 PD attendance related to AP Statistics																																				
Formal and Informal Teaching Experiences	5 Teaching mathematics																											X	X	X	X	X	X	X	X	X	X
																										Twin Valley HS				Twin Valley HS				Twin Valley HS			
	6 Teaching non-AP Statistics																																				
	7 Teaching AP Statistics																																				
	8 Conducting non-AP PD related to statistics																																				
	9 Conducting AP PD related to statistics																																				

		1981	1982	1983	1984	1985	1986	1987	1988	1989
AP Roles	10 AP Reader									
	11 AP Table leader									
Miscellaneous	12 Other (Specify)									
	13 Other (Specify)									
	14 Other (Specify)									
	Year	1981	1982	1983	1984	1985	1986	1987	1988	1989

	Year	1990	1991	1992	1993	1994	1995	1996	1997	1998
Landmark AP Statistics/Statistics									First AP Statistics Exam	Free Response: Butterfly Problem
									AP Reading at The College of New Jersey	AP Reading at University of Nebraska at Lincoln
									Workshop Statistics (Rossman, 1997) published	Yates, Moore,& McCabe (1999) book published
									REA (1997) Test-prep book published	Barron's Test-prep book (Sternstein, 1998) published
								TI-83 Graphing calculator released		
Formal and Informal Learning Experiences	1 Undergrad education related to statistics									
	2 Graduate education related to statistics									
	3 Non-AP PD related to statistics									
	4 PD attendance						X	X		

	related to AP Statistics																					AP Statistics Planning Conf				AP institute week-long institute											
Formal and Informal Teaching	5 Teaching mathematics	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
		Twin Valley HS				Twin Valley HS				Twin Valley HS				Twin Valley HS				Twin Valley HS				Twin Valley HS				Twin Valley HS				Twin Valley HS				Twin Valley HS			
	6 Teaching non-AP Statistics																							X	X	X	X	X	X	X	X	X	X	X	X	X	X
	7 Teaching AP Statistics																											X	X	X	X	X	X	X	X	X	X
	8 Conducting non-AP PD related to statistics																																				
	9 Conducting AP PD related to statistics																																				
AP Roles	10 AP Reader																														X				X		
	11 AP Table leader																																				
Miscellaneous	12 Other (Specify)																																				
	13 Other (Specify)																																				
	14 Other (Specify)																																				
	Year	1990				1991				1992				1993				1994				1995				1996				1997				1998			

	Year	1999	2000	2001	2002	2003	2004	2005	2006	2007
Landmark AP		Free Response: Dental/Apple Problem	Free Response: Drug A versus Drug B	Free Response: Blocking by Trees	Free Response: Einstein versus Newton	Free Response: Type I, Type II Errors: Law Firm Class Action Suit	Free Response: Pharmaceutical company, one-sided confidence interval	Free Response: Children in Daycare Centers, Comparison of Means	Free Response: Thermostats, Chi-squared distribution, sampling dist	
		AP Reading at University of Nebraska at Lincoln	AP Reading at University of Nebraska at Lincoln	AP Reading at University of Nebraska at Lincoln	AP Reading at University of Nebraska at Lincoln	AP Reading at University of Nebraska at Lincoln	AP Reading at University of Nebraska at Lincoln	AP Reading at University of Nebraska at Lincoln	AP Reading at University of Nebraska at Lincoln	AP Reading in Louisville, Kentucky

						Peck, Olsen, & Devore (2001) book published				Workshop Statistics with Fathom (Ross, Chance, & Lock, 2001) published				Yates, Moore, & Starnes (2003) book published								Bock, Velleman, & DeVeaux (2004) book published				Peck, Olsen, Devore (2005) Edition 2 released				Bock, Velleman, & DeVeaux (2006) Edition 2 released							
						Bohan (2000) Test-prep book published																															
		TI-83 Plus graphing calculator released								TI-83 Plus Silver Edition graphing calculator released												TI-84 Graphing calculator released															
		Fathom software released (Finzer, 1999)																								Fathom Version 2 released (Finzer, 2006)											
Formal and Informal Learning Experiences	1 Undergrad education related to statistics																																				
	2 Graduate education related to statistics																																				
	3 Non-AP PD related to statistics			X								X																									
		Attend NCSSM Statistics Leadership Institute								Attend NCSSM Statistics Leadership Institute																											
	4 PD attendance related to AP Statistics																																				
Formal and Informal Teaching Experiences	5 Teaching mathematics	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X																		
		Twin Valley HS				Twin Valley HS				Twin Valley HS				Twin Valley HS				Twin Valley HS																			
	6 Teaching non-AP Statistics	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X																		
	7 Teaching AP Statistics	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X																				
	8 Conducting non-AP PD related to statistics											X				X																					
										Governor's School – Data Driven Decision Making				Governor's School – Data Driven Decision Making																							
	9 Conducting AP PD related					X			X	X							X	X																			

	to statistics					AP Beginner Workshop (1) Conduct 1-day AP stats workshop at BCIU (1)				AP Experienced Workshop				AP Math Specialty Conference				AP Math Specialty Conference																			
AP Roles	10 AP Reader		X				X				X				X																						
	11 AP Table leader																		X				X					X									
Miscellaneous	12 Other (Specify)					X										X				X																	
						PCTM: Conduct Session – AP Statistics (1)								Co-author of AP Statistics Web Guide				Co-author of AP Test-prep book as supplement to Bock, Velleman, DeVeaux (2004) text								Co-author of 2nd Edition AP Test-prep book as supplement to Bock, Velleman, DeVeaux (2006) text											
	13 Other (Specify)															X																					
														JSM: Panel Member to discuss AP Statistics																							
	14 Other (Specify)																																				
	Year	1999				2000				2001				2002				2003				2004				2005				2006				2007			

Sample Experience Description Table

Year	Item	Where possible, provide a brief description of the experience and the people, places, and feelings associated with the experience. Also briefly describe the salient characteristics of the experience
1986	1	The course was an undergraduate introductory course taken in my junior year. The other students in the class were mostly undergraduate secondary education majors with whom I took a majority of my classes. The instructor taught the course as skill-driven, with calculations to be completed by hand. There was little focus on the meaning of the calculated values or on applications. The course did not make me want to take any more courses in statistics.
	2	
	3	
	4	
	5	
	6	
	7	
	8	
	9	
	10	
	11	
	12	
	13	
1986	14	

Appendix C

Event History Calendar

The Event History Calendar (EHC) provides a chronological table to record and view your experiences with learning and teaching statistics. You will be completing an EHC to detail your experiences in learning and teaching statistics. While the table looks rather daunting at first, my friends assure me that after reading the directions and filling out information for one or two years, completing the calendar can be accomplished rather quickly.

In the table, *columns* represent time periods, with each year from 1972 to 2007 divided into four seasons (Winter, Spring, Summer, and Fall). *Rows* represent particular experiences with statistics and are grouped into three categories of learning statistics, teaching statistics, and grading AP Statistics exams. Additional space is provided to record experiences with statistics that are not listed in the table.

Landmark experiences related to statistics are included in the table and shown in lavender font to help you in your recall of experiences. As you complete the EHC, please feel free to contact me via email ([e-mail address]) or phone ([phone number]) if you have any questions.

An example calendar is appended to this document. The easiest way to see the sample calendar is to use the link that appears at the end of the directions.

How to enter an event

- Begin with your first experience with statistics, which might be your first undergraduate course or an experience prior to when you started your undergraduate program.
- If your first experience is an undergraduate course, scroll down in the document until you see the year in which you were enrolled in the course. For example, if you were enrolled in an undergraduate statistics course for the fall semester of 1980, you would scroll down to 1980.
  - Type an "X" in the boxes that correspond with the seasons during which you were enrolled in the course. For example, if you were enrolled in an undergraduate statistics course for the fall semester of 1980, you would type an "X" in the third and fourth boxes for item 1 in the column for 1980.

		X	X
Intro Stats Course			

  - You should also provide a brief description of the experience, such as "Intro Stats Course," by typing the description in the second row of the block, as shown to the right.

		X	X
Intro Stats Course*			

- If you feel the experience was pivotal to your development in learning statistics or as a statistician or as a statistics educator, mark the experience with a “*” in the description, as shown to the right.

- If you had experiences with statistics before you enrolled in your undergraduate program, then you should begin the calendar with the year of your first experience. For example, if you were in enrolled in a high school statistics course during the 1989-1990 school year, you would scroll down to 1989.
  - Type an “X” in the boxes that correspond with the seasons during which you had the experience. For example, if you were enrolled in a high school statistics course during the 1989-1990 school year, you would type an “X” in the third and fourth boxes for the first “Other” category, item 12, in the column for 1989. You would also enter an “X” in the first and second boxes for item 12 in the column for 1990.
  - You should also provide a brief description of the experience, such as “Intro Stats Course,” by typing the description in the second row of the blocks, as shown to the right.
  - If you feel the experience was pivotal to your development in learning statistics or as a statistician or as a statistics educator, mark the experience with a “*” in the description, as shown to the right.

		X	X	X	X		
Intro Stats Course				Intro Stats Course			

		X	X	X	X		
Intro Stats Course*				Intro Stats Course*			

The descriptions

- After you record the timing and nature of your experience, you will need to record a short description of the experience. This description will be entered in a table different from the EHC but linked to the EHC. Directions are given in the next session.
- If possible, you should include information about the people, places, and feelings associated with the experience.
- Note those events that were particularly positive, such as events when you learned something new or realized something about your understanding of statistical concepts.
- Also note those events that were particularly negative, such as events where you realized you did not have the knowledge of statistics to engage in productive dialogue about statistics.

How to enter a description

- To move to the table for recording your description, you will control-click on a link within the EHC if you have a PC or click on the link if you have a Macintosh. If you control-click (or click) on the year of the experience in the EHC, for example control-click (click) on “1980,” a table for entering descriptions for the items of experiences from 1980 will show on your screen, as shown below.
- You should enter a description of your experience in the third column of the row that corresponds with the item number for the experience. Continuing with the previous example, a description for an undergraduate introductory course taken in 1980 would be entered in the third column of the first item for 1980, as shown below.
- Similarly, if your experience was entered in the “Other” category for item 12, you should control-click on the year of the experience to be linked to the table if you have a PC or click on the year of the experienced to be linked to the table if you have a Macintosh.

- When you are finished entering a description of the experience, you can control-click (click) on the year of the experience in the description table to return to the EHC.

Year	Item	Where possible, provide a brief description of the experience and the people, places, and feelings associated with the experience. Also briefly describe the salient characteristics of the experience
1980	1	The course was an undergraduate introductory course taken in my junior year. The other students in the class were mostly undergraduate secondary education majors with whom I took a majority of my classes. The instructor taught the course as skill-driven, with calculations to be completed by hand. There was little focus on the meaning of the calculated values or on applications. The course did not make me want to take any more courses in statistics.

Completing the process

- Please continue by entering other experiences, represented by different item numbers, which may have occurred during the year of your first experience in statistics.
- For each subsequent year, please continue entering events until the present time.

What to do about multiple experiences

- If you have more than one experience that applies to a category in any given year, then indicate each time period that applies to the collection of experiences.
- List each experience, and the number of experiences if a particular experience was repeated in the same year, in the second block.
- For example, if you conducted three AP workshops between summer and fall in 1980, you would type an "X" in the third and fourth boxes for item 9 in the column for 1980.
- You would also provide a brief description of the workshops, such as "AP Beginner Workshop" and "AP Experienced Workshop," by typing the description in the second row of the block.
- Of the three workshops, if two were for experienced AP teachers and one was for beginning AP teachers, the number of each type of experience can be recorded in parentheses next to the name of the experience, as shown to the right.

		X	X
AP Beginner Workshop (1) AP Experienced Workshop (2)			

Experiences to record in "other"

- Statistics-related professional presentations, like NCTM sessions
- Publications (statistics or statistics-related)
- Statistics-related informal experiences
  - Conversations with colleagues
  - Reading statistics books
  - Reading professional journals

- Pedagogical professional development that had an impact on statistics teaching
- Any other experience that contributed to growth or learning in statistics or statistics teaching

If you would like to see an example of an EHC, a sample calendar is also appended to this calendar. If you have a PC, control-click on the word “sample” to see the sample. If you have a Macintosh, click on the word “sample” to see the sample. When you are finished looking at the sample, you can scroll to the end of the document and control-click on the word “here” to return to the directions.

	Year	1972	1973	1974	1975	1976	1977	1978	1979	1980
Formal and Informal Learning Experiences	1 Undergrad education related to statistics									
	2 Graduate education related to statistics									
	3 Non-AP professional development related to statistics									
	4 Professional development attendance related to AP Statistics									
Formal and Informal Teaching Experiences	5 Teaching mathematics									
	6 Teaching non-AP Statistics									
	7 Teaching AP Statistics									
	8 Conducting non-AP professional development related to statistics									
	9 Conducting AP professional development related to statistics									
AP Roles	10 AP Reader									
	11 AP Table leader									
Miscellaneous	12 Other (Specify)									
	13 Other (Specify)									
	14 Other (Specify)									
	Year	1972	1973	1974	1975	1976	1977	1978	1979	1980

	Year	1981	1982	1983	1984	1985	1986	1987	1988	1989
Formal and Informal Learning Experiences	1 Undergrad education related to statistics									
	2 Graduate education related to statistics									
	3 Non-AP professional development related to statistics									
	4 Professional development attendance related to AP Statistics									
Formal and Informal Teaching Experiences	5 Teaching mathematics									
	6 Teaching non-AP Statistics									
	7 Teaching AP Statistics									
	8 Conducting non-AP professional development related to statistics									
	9 Conducting AP professional development related to statistics									
AP Roles	10 AP Reader									
	11 AP Table leader									
Mi	12 Other (Specify)									
	13 Other (Specify)									
	14 Other (Specify)									
	Year	1981	1982	1983	1984	1985	1986	1987	1988	1989

	Year	1990	1991	1992	1993	1994	1995	1996	1997	1998
andmark AP Statistics/Statistics									First AP Statistics Exam	Free Response: Butterfly Problem
									AP Reading at The College of new Jersey	AP Reading at University of Nebraska at Lincoln
									Workshop Statistics (Rossman, 1997) published	Yates, Moore, & McCabe (1999) book published
									REA (1997) Test-prep book published	Barron's Test-prep book (Sternstein, 1998) published
								TI-83 Graphing calculator released		
ormal and Informal Learning Experiences	1 Undergrad education related to statistics									
	2 Graduate education related to statistics									
	3 Non-AP professional development related to statistics									
	4 Professional development attendance related to AP Statistics									
ormal and Informal Teaching Experiences	5 Teaching mathematics									
	6 Teaching non-AP Statistics									
	7 Teaching AP Statistics									
	8 Conducting non-AP professional development related to statistics									
	9 Conducting AP professional development related to statistics									
P	10 AP Reader									

	11 AP Table leader									
iscellaneous	12 Other (Specify)									
	13 Other (Specify)									
	14 Other (Specify)									
	Year	1990	1991	1992	1993	1994	1995	1996	1997	1998

	Year	1999	2000	2001	2002	2003	2004	2005	2006	2007
Landmark AP Statistics/Statistics Experiences		Free Response: Dental/Apple Problem	Free Response: Drug A versus Drug B	Free Response: Blocking by Trees	Free Response: Einstein versus Newton	Free Response: Type I, Type II Errors: Law Firm Class Action Suit	Free Response: Pharmaceutical company, one-sided confidence interval	Free Response: Children in Daycare Centers, Comparison of Means	Free Response: Thermostats, Chi-squared distribution, sampling dist	
		AP Reading at University of Nebraska at Lincoln	AP Reading at University of Nebraska at Lincoln	AP Reading at University of Nebraska at Lincoln	AP Reading at University of Nebraska at Lincoln	AP Reading at University of Nebraska at Lincoln	AP Reading at University of Nebraska at Lincoln	AP Reading at University of Nebraska at Lincoln	AP Reading at University of Nebraska at Lincoln	AP Reading in Louisville, Kentucky
			Peck, Olsen, & Devore (2001) book published	Workshop Statistics with Fathom (Ross, Chance, & Lock, 2001) published	Yates, Moore, & Starnes (2003) book published		Bock, Velleman, & DeVeaux (2004) published	Peck, Olsen, & Devore (2005) Edition 2 released	Bock, Velleman, & DeVeaux (2006) Edition 2 released	
			Bohan (2000) Test-prep book published							
		TI-83 Plus graphing calculator released		TI-83 Plus Silver Edition graphing calculator released			TI-84 Graphing calculator released			
		Fathom software released (Finzer, 1999)						Fathom Version 2 released (Finzer, 2006)		
Formal and Informal Learning	1 Undergrad education related to statistics									
	2 Graduate education related to statistics									
	3 Non-AP professional development related to statistics									

	4 Professional development attendance related to AP Statistics									
Formal and Informal Teaching Experiences	5 Teaching mathematics									
	6 Teaching non-AP Statistics									
	7 Teaching AP Statistics									
	8 Conducting non-AP professional development related to statistics									
	9 Conducting AP professional development related to statistics									
AP Roles	10 AP Reader									
	11 AP Table leader									
Miscellaneous	12 Other (Specify)									
	13 Other (Specify)									
	14 Other (Specify)									
	Year	1999	2000	2001	2002	2003	2004	2005	2006	2007

EHC Description of Experiences

Year	Item	Where possible, provide a brief description of the experience and the people, places, and feelings associated with the experience. Also briefly describe the salient characteristics of the experience
1972	1	
	2	
	3	
	4	
	5	
	6	
	7	
	8	
	9	
	10	
	11	
	12	
	13	
1972	14	

1 Undergrad education related to statistics	2 Graduate education related to statistics	3 Non-AP professional development related to	4 Professional development	5 Teaching mathematics	6 Teaching non-AP Statistics	7 Teaching AP Statistics	8 Conducting non-AP professional development related to statistics	9 Conducting AP professional development related to statistics	10 AP Reader	11 AP Table leader	12-14 Other

Appendix D

Critical Incident Description

Think about your experiences learning statistics, and in particular learning about variation. From your experiences, identify one particularly positive experience and one particularly negative experience related to your informal or formal study of variation. Please provide written responses to the information requested below. In general, your response to each experience should be approximately one single-spaced page long. You may be asked to expand upon your responses when we meet to discuss your experiences.

Describe **one positive experience** related to your informal or formal study of variation – an experience that you recall as being particularly good or that you feel resulted in significant learning on your part. Elaborate on this experience and the timing of the experience. To the extent possible, please address all of the questions listed below in your written response.

Describe **one negative experience** related to your informal or formal study of variation – an experience that you recall as being particularly bad or that you feel affected your perception of your understanding or knowledge of variation in a negative way. Elaborate on this experience and the timing of the experience. To the extent possible, please address all of the questions listed below in your written response.

List of Questions for each Incident

Details of the experience:
- When, where, and for how long did the experience occur?*
- What events or circumstances precipitated the experience or caused the experience to occur in the way in which it did?
- What other people or circumstances played an influential role in the experience?
- How did the experience end?

Reflections on the experience:
- As you reflect on the experience, why do you believe you viewed the events surrounding this experience positively or negatively?
- What emotions did you recall feeling during the experience?*
- In response to the experience, what actions did you take?
- What do you believe you learned from the experience?

Effects of the experience:
- How has the experience affected your understanding of variation?
- How has the experience affected your statistics teaching?

Beyond the experience:
- If you could change past events surrounding the experience, what would you change and why?
- If you were to encounter the experience under identical circumstances to those surrounding the original experience, what effect do you believe the experience would have on you today?

* These questions overlap with some of the information requested in the Event History Calendar. If you already provided the information in your Event History Calendar, you don't need to duplicate your responses to these questions.

Appendix E

Abbreviated Context I Interview Schedule

*Note: In general, whatever the participant responds, **probe for details about the experience, particularly with respect to variation or experiences indicative of transformational learning related to statistics.***

Before we talk about what you included in your calendar, I would like to know if you remembered any additional experiences you had with statistics that should be recorded on the calendar?

- **Could you tell me when the event occurred?**
- **Over what time period did the event take place?**
- **Please provide a brief description of the event, and the people, places, and feelings associated with the event.**
- **Please describe any particularly salient characteristics of the experience.**

General Questions for Each Experience Type Listed in EHC

Think about the statistics courses you took (or other type of experience). Were there any in which you feel you learned a great deal about statistics or variation or in which you grew a great deal as a statistician or a statistics educator?

- **Which of the statistics courses you took was the course where you learned the most about statistics or variation or where you grew the most as a statistician or a statistics educator?**
- **What did you learn in the course?**

  In general, if the participant mentions variation, probe for details about what they learned about variation. If the person describes an event that may have triggered a disorienting dilemma, probe for details of the experience in light of potential indicators for transformation, including

  - *a description of the disorienting dilemma,*
  - *self-examination,*
  - *critical assessment of assumptions,*
  - *recognition that others have had similar experiences,*
  - *exploring new roles through rational discourse with others,*
  - *planning a course of action,*
  - *constructing the knowledge and skills needed to enact the plan,*
  - *experimenting with new roles,*
  - *building a sense of competence, and*
  - *reintegrating into life based on new roles.*
- **Three major areas of statistics are exploratory data analysis, study design, and inferential statistics. Which, if any, of these were part of this course?**
  - **What do you remember learning about the role of variation in [*exploratory data analysis, planning a study, inferential statistics*]?**

  - **How is what you learned in that course different from or similar to how you now see variation in [exploratory data analysis, planning a study, inferential statistics]?**
- **As you reflect on your experiences in the course, what, if any, experiences seem particularly important to your development of knowledge about statistics and variation or important to your development as a statistician or as a statistics teacher?**
  - **Why do you believe [paraphrase experiences] influenced your development of knowledge about statistics and variation or as a statistician or as a statistics teacher?**
  - **What development resulted from [paraphrase experiences]?**
- **What do you remember feeling about [paraphrase experience or characteristic]?**
- **What conversations did you have about [paraphrase experience or characteristic]?**
  - **In relation to the experience, what was the role of the person you had this conversation with?**
- **What did you do in response to [paraphrase experience or characteristic]?**
- **As a result of [paraphrase experience or characteristic], how, if at all, did you see a change in how you thought about variation?**
- **In how you thought about statistics in general?**
- **In how you view yourself as a statistician or statistics teacher?**

How, if at all, does your knowledge of statistics now differ from your knowledge of statistics when you [had this experience]?

- **How does the way you now see variation differ from or agree with the way you saw variation when you [had this experience]?**

Teaching question: **How does the way you now teach variation differ from the way you taught variation when you first started teaching statistics?**

- **How is your knowledge of pedagogy and use of pedagogical strategies affected by your knowledge of statistics?**

Professional development question: **How did you learn about this program, and what factors influenced your decision to attend this session [*program*]?**

Questions Related to Critical Incidents

If any of the requested questions were not addressed in the critical incident description, ask the participant to answer the unanswered questions from this list:

Details of the experience:

- **When, where, and for how long did the experience occur?***
- **What events or circumstances precipitated the experience or caused the experience to occur in the way in which it did?**
- **What other people or circumstances played an influential role in the experience?**
- **How did the experience end?**

Reflections on the experience:

- **As you reflect on the experience, why do you believe you viewed the events surrounding this experience positively or negatively?**
- **What emotions did you recall feeling during the experience?***
- **In response to the experience, what actions did you take?**

- **What do you believe you learned from the experience?**

Effects of the experience:
- **How has the experience affected your understanding of variation?**
- **How has the experience affected your statistics teaching?**

Beyond the experience:
- **If you could change past events surrounding the experience, what would you change and why?**
- **If you were to encounter identical circumstances to those surrounding the experience, what affect do you believe the experience would have on you today?**

Additional Questions:

After the incident occurred, what conversations did you have about the incident?
  - **In relation to the experience, what was the role of the person you had this conversation with?**
  - **How, if at all, did those conversations impact your understanding of the incident?**
  - **Using the title of the person's position or the person's relationship to you, from whom, if anyone, did you seek input that you believed might differ from your perspective of the incident?**
    - **What perspective did they offer for the incident?**

Describe how often you think about or thought about this incident.
  - **About how often, and when did you reflect on the incident?**
  - **Describe the form of your reflection, e.g., thinking, writing, or talking.**
    - **How, if at all, did your reflection on the incident change your interpretation or understanding of the incident or strengthen your initial interpretation or understanding of the incident?**
      - **Why do you believe your interpretation or understanding of the incident changed?**

What, if any, relationship do you see between the critical incident you documented in your critical incident document and the other experiences you had surrounding this incident? [*Point to events on the EHC that surround the incident*.]

Appendix F

Abbreviated Context II Interview Schedule

Note: In general, whatever the participant responds, probe for details about the experience, particularly with respect to variation or experiences indicative of transformational learning.

For today's session, I would like to explore some of the experiences we discussed last time. Since our last conversation and subsequent to your reflections on that conversation, what, if any, other thoughts, events, or experiences related to your learning of statistics do you feel should be added to your Event History Calendar?

- **Could you tell me when the [event or experience] occurred?**
- **Over what time period did the [event or experience] take place?**
- **Could you provide a brief description of the [event or experience] and the people, places, and feelings associated with the [event or experience]?**
- **Please describe any particularly salient characteristics of the experience.**

We are now going to talk in depth about some of the events you identified as pivotal on your calendar. Some of the questions that I ask may seem to be repetitive, but I want to be sure that we have touched upon important characteristics of some of your experiences.

Of all of your experiences with learning and teaching statistics, which two or three experience do you believe precipitated the greatest change in your understanding of variation or resulted in the greatest change in your understanding of variation?

- **Did you view the [name] experience as pivotal at the time of the experience, and if so, how?**
- **How do you currently view the experience?**
- **What features of this experience do you believe were essential for your learning?**
  - **Why do you believe [name feature] was particularly effective for you?**
  - **Do you believe [name features] might have a common effect among statistics teachers, and if so, why?**

What features of this experience do you believe were largely ineffective for your learning?

- **Why do you believe [name feature] was particularly ineffective for you?**
- **Do you believe [name features] might have a common effect among statistics teachers, and if so, why?**

In response to the experience, what actions did you take?

- **Why do you believe you [describe the action]?**
- **How did [describe the action] affect your learning?**
- **How did [describe the action] affect your teaching?**

In our last session, you indicated that [*insert name of professional development/class/ teaching experience*] was an experience that was particularly educative for you. In particular, you mentioned that you believed [*name experience or characteristic*] was instrumental in your learning about variation. Did you view the experience as pivotal at the

time of the experience, and if so, how? *Repeat the preceding series of questions for this experience.*

Over the course of the past few months, you've spent a considerable amount of time reflecting on your teaching and learning in statistics. We've talked about each of these experiences individually, but as you've been reflecting on your experiences, did you feel there were any patterns in your experiences? If so, what are they?

Can you describe how you believe this collective group of experiences might have contributed to your development of an understanding of variation?

- **Why do you believe these collective experiences were particularly effective for you?**
- **Which grouping(s) of these experiences do you believe would have a similar learning effect on other statistics teachers, and why?**

In your educational experiences, what hindrances to your development of an understanding of variation do you believe existed?

- **Why do you believe these characteristics were a hindrance for you?**
- **What personal characteristics do you believe a person would need to have in order for these characteristics to not be a hindrance?**

If you could change something in your learning experiences with statistics and with variation in particular, what would you change and why?

How, if at all, did your knowledge of variation change as a result of this collective group of experiences?

How, if at all, did you see a change in how you thought about variation?

- **Why do you believe that to be the case?**

How, if at all, did you see a change in how you thought about statistics in general?

- **Why do you believe that to be the case?**

How, if at all, did you see a change in how you view yourself as a statistician or statistics teacher?

- **Why do you believe that to be the case?**

What else you would like to add about your experiences that would help me to understand your learning experiences related to variation?

VITA

Susan A. Peters

Education

Ph.D. Curriculum and Instruction, The Pennsylvania State University, August, 2009
M.A.S. Master of Applied Statistics, The Pennsylvania State University, August, 2007
M.A. Mathematics, West Chester University of Pennsylvania, August, 1991
B.S. Secondary Education: Mathematics, Kutztown University of Pennsylvania, May, 1987

Academic and Professional Experience

Assistant Editor, *Journal for Research in Mathematics Education*, 2007-2009
Supervisor, CI 495C: Clinical Applications of Instruction–Secondary Education, 2006-2007
Instructor, MTHED 412: Teaching Secondary Mathematics II, 2005-2006
Mathematics Teacher, Twin Valley School District, 1987-2003

Selected Publications

Peters, S. A. (in press). Engaging with the art and science of statistics. *Mathematics Teacher*. Reston, VA: National Council of Teachers of Mathematics.

Zbiek, R. M., Peters, S. A., Boone, T., Johnson, K., & Foletta, G. (2009, April). *Locally logical mathematics: An emerging teacher honoring both students and mathematics*. Paper presented at the 2009 American Educational Research Association annual meeting, San Diego, CA

Carroll, A., Carver, R., Peters, S., & Ricks, J. (2003). *Preparing for the statistics AP® exam with Stats: Modeling the World*. Boston, MA: Pearson.

.

Selected Presentations

Zbiek, R. M., Heid, M. K., Blume, G., & Peters, S. A. (2009, April). Mathematical processes lens for prospective secondary teachers' mathematics, Research Presession of the National Council of Teachers of Mathematics 2009 Annual Meeting and Exposition, Washington, DC.

Peters, S. A. (2008, July). Robust understanding of variation: An interaction of three perspectives, Research presentation at the 32nd conference of the International Group for the Psychology of Mathematics Education and the 30th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Morelia, Mexico.

Peters, S. A., Fratto, C. L., Heid, M. K. (2007, October). Investigating statistical concepts from a distributional perspective, National Council of Teachers of Mathematics (NCTM) Regional Conference, Richmond, VA.

Fellowships and Awards

Alumni Society Research Initiation Grant
Mid-Atlantic Center for Mathematics Teaching and Learning Fellowship
Robert E. and Virginia L. Mountz Scholarship in Mathematics Education