ROBUST UNDERSTANDING OF STATISTICAL VARIATION

SUSAN A. PETERS
University of Louisville
s.peters@louisville.edu

ABSTRACT

This paper presents a framework that captures the complexity of reasoning about variation in ways that are indicative of robust understanding and describes reasoning as a blend of design, data-centric, and modeling perspectives. Robust understanding is indicated by integrated reasoning about variation within each perspective and across perspectives for four elements: variational disposition, variability in data for contextual variables, variability in relationships among data and variables, and effects of sample size on variability. This holistic image of robust understanding of variation arises from existing expository and empirical literature, and additional empirical study.

Keywords: *Statistics education research; Understanding variation; Framework for robust understanding of variation; SOLO Taxonomy*

1. INTRODUCTION

"*If there was no variation, there would be no need for statistics and statisticians*" (Snee, 1999, p. 257).

As the opening quote suggests, variation is central to statistics and plays a crucial role throughout statistical problem solving. The ability to think statistically—viewed by many as an instructional goal for introductory courses (e.g., Ben-Zvi & Garfield, 2004; Chance, 2002)—requires consideration of variation (Franklin et al., 2007; Shaughnessy, 1997; Wild & Pfannkuch, 1999). As the study of statistics gains prominence in content recommendations (e.g., National Council of Teachers of Mathematics, 1989, 2000), so, too, should the study of variation. Statisticians recognize the primacy of variation to statistics (e.g., Cobb & Moore, 1997; Moore, 1998; Snee, 1999; Wild & Pfannkuch, 1999), but not all statistics textbooks make the primary role of variation in statistics explicit (e.g., Porter, 2001). When coupled with the perception that many teachers lack experiences with statistics (e.g., Shaughnessy, 2007), a need arises for clear explication of what is encompassed in robust statistical understandings, particularly understanding of statistical variation, for teachers and their students to develop those understandings, and to attain the goal of a statistically literate population.

Research to examine reasoning about and understanding variation offers insights into statistical connections needed to understand variation (e.g., Reading & Shaughnessy, 2004; Watson, Callingham, & Kelly, 2007), and expository literature outlines views deemed necessary for understanding variation (e.g., Garfield & Ben-Zvi, 2005). Although insightful and informative, collectively this work falls short of providing a holistic image of robust understanding of variation. We still need articulation of important aspects of variation and important connections among the ideas—articulation that can be informed

by considering what robust understanding of variation means. Emerging from a larger study to examine secondary teacher-leaders' development of robust understandings of variation (Peters, 2009), this paper articulates an organized and interconnected set of actions and reasoning indicative of robust understandings of variation to provide a holistic image of that understanding.

2. BACKGROUND

Reasoning about variation can be captured from three perspectives: a design perspective that integrates acknowledgement and anticipation of variation in the design of quantitative studies; a data-centric perspective that integrates processes of representing, measuring, and describing variation in exploratory data analysis; and a modeling perspective that integrates reasoning to fit models to patterns of variability in data and statistics, judging goodness of fit, and transforming data to improve fit. These descriptions of data-centric and modeling perspectives for variation are adaptations and extensions of data-centric and modeling perspectives on distribution (Prodromou & Pratt, 2006). Addition of a design perspective provides a means to acknowledge requisite reasoning about variation (Wild & Pfannkuch, 1999) that is not captured by data-centric and modeling perspectives. Empirical and expository literature considered from these three perspectives offers a foundation for describing reasoning and behaviors indicative of understanding variation on which the framework described in this paper is built. Throughout this paper, the terms "variation" and "variability" are used interchangeably. Although some researchers make a distinction between the two terms, there exists no general agreement in statistics education research on the use of these terms (Garfield & Ben-Zvi, 2008).

2.1. DESIGN PERSPECTIVE

Research to investigate reasoning about variation illuminates some aspects of understanding and reasoning from a design perspective. Fundamental to understanding variation is recognizing the omnipresence of variability (Cobb & Moore, 1997), recognizing that variability can be seen in everything and everywhere. This recognition leads to the search for potential sources of variation that can then be controlled through design. Although novices may not necessarily recognize the omnipresence of variability, they exhibit potential precursors such as recognizing and expecting variation in events such as tossing a die (Watson & Kelley, 2002a). The importance accorded to recognizing the omnipresence of variability and the need for design that emerges from this recognition is underscored in expository literature (e.g., Franklin et al., 2007; Garfield & Ben-Zvi, 2005; Moore, 1990).

A main focus of statistics is inferring characteristics of a population through analysis of data collected from a sample of the population. Valid inferences stem from appropriate sampling methods and designs that include randomization (Cobb & Moore, 1997) to control variability and allow determination of whether observed data characteristics are likely due to chance (Franklin et al., 2007). Research on reasoning about samples and variation focuses on sampling methods that produce representative samples and on sample variability instigated from random processes (e.g., Shaughnessy, Ciancetta, & Canada, 2004; Watson, 2002; Watson & Kelly, 2002a, 2002b). Research suggests that sophisticated reasoning about variation and samples requires balancing two notions: sample representativeness and sample variability (Rubin, Bruce, & Tenney, 1990). Representative samples have characteristics similar to those of the population, whereas

sample variability results from samples that are not all identical and do not match the population exactly. Also important are considerations of the effects of sample size on the variability of samples and sample statistics to explain and describe variability in data (Garfield & Ben-Zvi, 2005).

Experimental design introduces randomization to control systematic and random variability (Derry, Levin, Osana, Jones, & Peterson, 2000). Coordinating systematic, between-group variation and random, within-group variation provides a critical link between variation and formal inference (Reid & Reading, 2005), one deemed necessary for sophisticated reasoning about variation (Reid & Reading, 2008). Individuals struggle to control variation in studies, which may preclude them from establishing unbiased and significant results, but they can consider and identify different sources of variation (Petrosino, Lehrer, & Schauble, 2003) and anticipate sources of error in data (Masnick & Klahr, 2003). Sophisticated reasoning is needed to design experiments and observational studies well; this includes consideration of sources of variation and ways to control variation from those sources (Garfield & Ben-Zvi, 2005; Moore, 1990).

Synthesis of the aforementioned expository and empirical literature devoted to observational and experimental design leads to four broad aspects of reasoning about variation within the design perspective that are consistent with sophisticated reasoning about and understanding of variation. The four items shown in Figure 1 are expanded in the Results section.

- Recognition of the omnipresence of variability and anticipation and acknowledgement of natural variability when designing a study, drawing conclusions from a study, or critiquing a study
- Anticipation of possible sources of variability (such as measurement variability) in the context of a study and description of the effects various sources of variability may have on the measured variability of characteristics under study
- Anticipation of the effects of sample size on the variability of the sample and on the variability of statistics that characterize the sample while designing a study or in critiquing a study design
- Acknowledgement of controllable and uncontrollable variability, such as explicating the benefit of using random assignment or random sampling of observational/experimental units in the context of a particular study, and the need for control to isolate systemic variation from random variation

Figure 1. Initial indicators of reasoning about variation within the design perspective

2.2. DATA-CENTRIC PERSPECTIVE

After a study is designed and data are collected, statisticians typically explore data by examining graphical displays and numerical summaries. Sophisticated data exploration includes representing and reading data by changing the form of data to extract descriptive information (Curcio, 1987) and rearranging data to interpret and identify data trends (Friel, Curcio, & Bright, 2001). Statistical thinking embodied in the "dynamic process of changing representations to engender understanding" (Wild & Pfannkuch, 1999, p. 227)—transnumeration—is considered necessary for understanding variation (Garfield & Ben-Zvi, 2005). Also important while examining data is flexible movement between pointwise and aggregate views of data (Bakker & Gravemeijer, 2004). Pointwise views of data allow for calculation of summary values such as average absolute deviation, whereas views of data as aggregate collections of values (e.g., Ben-Zvi & Arcavi, 2001; Hammerman & Rubin, 2004; Hancock, Kaput, & Goldsmith, 1992; Konold, Harradine, &

Kazak, 2007) focus reasoning on global characteristics of data such as patterns of variability in data.

Transnumeration (Wild & Pfannkuch, 1999) includes reasoning with summary statistics such as interquartile range, mean absolute deviation, and standard deviation as numerical representations of data. Research suggests that individuals experience difficulty when moving beyond intuition and calculation to examine the meanings of statistical measures (e.g., Clark, Kraut, Mathews, & Wimbish, 2007; Mokros & Russell, 1995). Individuals need to move beyond intuitive, rule-based, and procedural conceptions (e.g., Shaughnessy, Ciancetta, Best, & Canada, 2004; Silva & Coutinho, 2006, 2008; Sorto, 2004) and deterministic comparisons of measures or graphs (delMas & Liu, 2005; Lann & Falk, 2003). Understanding entails thoughtful consideration of representations and measures, including which representations and measures best represent data (Garfield & Ben-Zvi, 2005) and reasoning about spread relative to center (Garfield, delMas, & Chance, 2007; Reading & Shaughnessy, 2004; Shaughnessy, Canada, & Ciancetta, 2003). For example, understanding standard deviation or mean absolute deviation necessitates 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). This dynamic conception of standard deviation also allows consideration of the effects of outliers on summary measures.

Summary measures describe data and allow for comparisons between data sets or between samples and populations. Much research focuses on group comparisons (e.g., Makar & Confrey, 2004, 2005; Watson et al., 2007) and reveals individuals' nontrivial struggles to describe variation between groups to reason about group differences (Makar & Confrey, 2002, 2004). Sophisticated understandings of variation are partially exhibited when individuals use global summary measures of variation to compare groups and to distinguish between within-group and between-group variation (Garfield & Ben-Zvi, 2005). Linking within-group variation to between-group variation to make inferences indicates sophisticated reasoning about variation (Reid & Reading, 2008).

Summary measures such as correlation and the coefficient of determination allow for determination of the strength of relationships between variables and also provide information about variance in data. Students struggle to reason about these bivariate measures as measures descriptive of association and variance (Truran, 1995). Students also struggle to reason covariationally (Zieffler & Garfield, 2009). Those who exhibit sophisticated covariational reasoning appear to perceive bivariate data in three dimensions, in which the third dimension stems from "the relative density of the data points" (Cobb, McClain, & Gravemeijer, 2003, p. 67). Results of research focused on reasoning about bivariate data along with research related to reasoning about variation from the data-centric perspective more generally lead to the behaviors and reasoning consistent with variation from the data-centric perspective displayed in Figure 2.

2.3. MODELING PERSPECTIVE

Important for considering variation is modeling the variation in data "for the purposes of prediction, explanation, or control" (Wild & Pfannkuch, 1999, p. 226); ideas that entail reasoning about variation from the modeling perspective. The modeling perspective includes reasoning about formal inference through comparisons with theoretical models.

The normal distribution plays a central role in standard parametric methods. Work with normal distributions frequently involves the 68-95-99.7 rule for normal distributions and using information about the mean and standard deviation to reason about specific characteristics of standard and nonstandard normal distributions. Often individuals

- Creation, use, or interpretation of various data representations to highlight patterns of variability in data and to focus on the aggregate features of the data
- Calculation of summary statistics values or acknowledgement of the utility in measuring the variability in the data and use and interpretation of appropriate summary values (including measures of variation such as range, interquartile range, and standard deviation for univariate sets or correlation and coefficient of determination of bivariate sets) to describe holistic features of the distribution
- Estimation of measures of variability for data based upon characteristics of the distribution, including shape, center, and outliers for univariate sets of data, or correlation and coefficient of determination for bivariate sets of data
- Use of summary statistics measures, including measures of variation, to make comparisons and to examine the variability within and among groups

Figure 2. Initial indicators of reasoning about variation within data-centric perspective

overgeneralize conditions for using the empirical rule and assume normality when doing so makes little sense within context (delMas, Garfield, Ooms, & Chance, 2007), a mistake that may originate from describing characteristics of normal distributions without considering why normality is appropriate (Wilensky, 1995, 1997). In general, using appropriate data transformations and models to account for the variability in univariate or multivariate data and assessing the goodness of a model's fit by examining deviations from the model (Garfield & Ben-Zvi, 2005), for example, illustrate reasoning consistent with understanding variation from the modeling perspective.

A second category of theoretical distributions at the center of inferential methods are sampling distributions. Researchers note individuals' tendencies to confuse sample distributions with sampling distributions (e.g., Saldanha & Thompson, 2002) and to confuse the variation of individual samples with the variation of sample means in sampling distributions (e.g., Garfield et al., 2007; Meletiou-Mavrotheris & Lee, 2003). Reasoning about sampling distributions is foundational for reasoning formally about inference and seemingly requires reasoning proportionally about the likelihood of sample results by considering an observed sample statistic in relation to the distribution of statistics for samples of a given size from the population (Saldanha & Thompson, 2002). Difficulties in recognizing the effects of sample size on sampling distributions (e.g., Chance, delMas, & Garfield, 2004; Fischbein & Schnarch, 1997), particularly on the standard deviation of sampling distributions, abound. Also important is the probabilistic, nondeterministic form in which conclusions—predictions or inferences about a population—should be stated (e.g., delMas, 2004; Meletiou-Mavrotheris, 2007). Figure 3 details the aspects of understanding variation within the modeling perspective that arise from empirical and expository literature on inference and on modeling data.

- Use of a normal distribution to model patterns of variation for symmetric, bell-shaped data distributions and use of the characteristics of a normal distribution to examine characteristics of the data, including invocation of the empirical rule: 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
- Use of appropriate models or transformations to account for the variability in data and to explore characteristics and relationships of univariate or bivariate sets of data
- Use of deviations from the model to describe the goodness of fit of the model
- Use of models to make predictions or statistical inferences from the data while allowing for variability in predictions or interpretations

Figure 3. Initial indicators of reasoning about variation within modeling perspective

3. THEORETICAL UNDERPINNINGS

The study from which this paper originates uses the Structure of the Observed Learning Outcome (SOLO) Model (Biggs & Collis, 1982, 1991) as a general framework to investigate and describe characteristics of reasoning indicative of robust understandings of variation. The SOLO Model is an empirically derived, neo-Piagetian model of cognitive development (Pegg & Tall, 2001, 2005) consisting of five modes of functioning that correspond closely to Piaget's developmental stages (Biggs & Collis, 1982, 1991): sensorimotor, ikonic, concrete symbolic, formal, and postformal modes. Each successive mode reveals increased degrees of abstraction in reasoning (Biggs & Collis, 1982). This paper focuses strictly on reasoning in the formal mode; reasoning that generates speculations that both incorporate and transcend particular situations. Although some adults never develop sufficient understanding to reason in the formal mode for particular areas, formal reasoning corresponds with thinking characteristically exhibited by undergraduates and professionals (Biggs & Collis, 1982; Groth & Bergner, 2006) and is appropriate for considering reasoning indicative of robust understandings.

A cycle of three levels of response describes the underlying cognitive structure in each mode and reveals a hierarchy of cognitive complexity (Pegg, 2003). At the unistructural level of the formal mode, an individual focuses on one relevant aspect in the formal mode for the construct under consideration, whereas at the multistructural level the individual focuses on more than one relevant aspect without integrating the aspects (Biggs & Collis, 1991). At the relational level, the individual integrates all relevant aspects to reveal coherent meaning. Researchers identify more than one cycle of levels within a mode (Callingham, 1997; Pegg, 2003; Watson, Collis, Callingham, & Moritz, 1995). A second cycle becomes apparent when relational reasoning in the first cycle consolidates to a single construct more sophisticated than relational reasoning in the first level that then becomes unistructural in the second level of cycles (Pegg & Tall, 2001). The second cycle of levels represents advanced development that maintains a comparable degree of abstraction with the first cycle of levels.

The framework put forth in this paper examines reasoning about variation using the general framework of the SOLO Model as its foundation. Figure 4 represents the first and second cycles of levels of reasoning indicative of understanding variation. The first cycle, depicted by subscripted levels of "1," lies within each perspective of variation. The arrows represent increased sophistication in reasoning from each perspective. For example, an individual who evidences relational reasoning (R_1) within the design perspective displays integrated reasoning about aspects of variability related to design. Reasoning in the second cycle of levels in the formal mode involves reasoning that integrates perspectives. The subscripted levels of "2" depict this second cycle. An individual who reasons relationally from each perspective and integrates reasoning from the three perspectives reasons at a relational level (R_2) in the second cycle of levels in the formal mode. Relational reasoning within the second cycle of levels, that is, relational reasoning within all three perspectives (R_1 within design, data-centric, and modeling perspectives) and relational reasoning across all three perspectives (R_2), indicates *robust understanding of variation*. Although individuals who reason at unistructural and multistructural levels in the second cycle exhibit relational reasoning from one or more perspectives, they do not exhibit relational reasoning across perspectives and hence do not exhibit reasoning consistent with robust understanding of variation.

Understanding Variation in Formal Mode

Relational Reasoning Within the Design Perspective	Relational Reasoning Within the Data-Centric Perspective	Relational Reasoning Within the Modeling Perspective

First Cycle of Levels

U_1 U_1 U_1

M_1 M_1 M_1

R_1 R_1 R_1

U_2

M_2 Relational Reasoning About Variation Across Perspectives

R_2
Reasoning Indicative of Robust Understanding of Variation
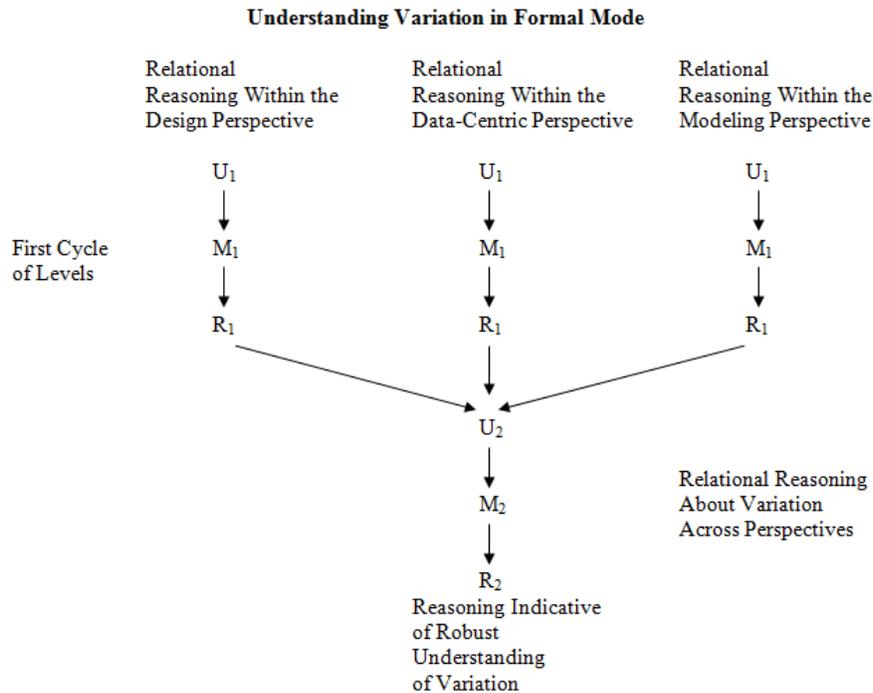
Figure 4. Robust understanding of variation

4. DATA COLLECTION

The main study from which this paper originates focuses on secondary mathematics/statistics teacher-leaders' development of robust understandings of variation (Peters, 2009). With a primary goal of investigating the phenomenon of developing robust understandings of variation, Advanced Placement® (AP) Statistics teacher-leaders were selected for participation in the main study under the assumption that they were more likely to exhibit robust understandings of variation than secondary mathematics teachers in general (R. Peck, Personal Communication, May 21, 2007). Teacher-leaders who differed in the number of years they had taught statistics and who represented a variety of educational and statistical experiences were selected from those who agreed to participate. This purposeful sample consisted of 16 teacher-leaders from across the United States. A 90- to 120-minute semi-structured content interview was conducted with each teacher to investigate teachers' reasoning about variation. Teachers responded to three main tasks developed specifically for this study. Task design was informed and influenced by the synthesis of statistics education literature described in the Background section. Each task statement was purposefully vague to allow multiple approaches and was designed to elicit formal, abstract reasoning from the design, data-centric, and modeling perspectives and integrated reasoning across perspectives. This interview yielded the primary source of evidence to inform the developing framework for robust understanding of variation. Secondary sources included teachers' statistics course syllabi and content-related passages from additional teacher interviews focused on learning experiences.

Throughout the task descriptions presented in the next three sections, discussion focuses on the rationale behind task components, how teachers' responses to the tasks

were probed for depth of understanding, and how teachers were prompted to reason about elements that might not have been addressed without prompting. Included are examples of the types of questions that accompanied task components. The order of questions depended on the direction taken by the teacher in response to task components.

4.1. CONSULTANT TASK

Figure 5 displays the first task statement presented to teachers, the Consultant Task. With no information about how administrators selected exams, teachers could reason from the design perspective to conclude that potential bias issues might prevent them from forming conclusions. Because average scores for each sample are the only measures given, teachers could reason from the data-centric perspective to ask for additional information about the data. Finally, by having values for the means to compare consultant scores, teachers could reason from the modeling perspective to suggest that inferential methods could help them to form conclusions.

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 5. Consultant Task description

Teachers who approached the task with a focus on missing information about design (e.g., sampling technique) prompted interview questions related to why the information was important, how the information helped to answer the administrators' question, and what conclusions could be drawn in the absence of that information. Addressing the importance of additional information provided an opportunity for teachers to express the omnipresence of variability, concerns related to the effects of potential sources of variation, and a need to know how sources of variation were controlled. From the general way in which questions were posed, teachers could reason about design elements in terms that transcended contexts to reason in the formal mode. Attention to what conclusions followed from the information could offer further insight into teachers' thinking about variation. Additional questions prompted teachers to describe and defend the design they would use to answer the administrators' question, providing an opportunity for teachers to reason about design elements, consider controlling strategies, and describe the effects of sample size.

Teachers' requests for information about variation in scores prompted interview questions about why the information was needed. Values for measures of variation and dotplots of the data, supplied to teachers separately upon request, are shown in Figure 6. To inform how teachers described variation and interpreted standard deviation, questions

focused on describing the expected distributions from the summary values. There is a discrepancy between the summary measures and the dotplot for Consultant Two's scores that stems from a score of 150 incorrectly entered instead of 15. The summary measures were calculated with the value of 150, but the dotplot only displayed scores from 0 to 15. This discrepancy provided an opportunity for teachers, upon prompting, to use the dotplot to estimate and reason about the correct values for mean and standard deviation. Responses could reveal how teachers used data to reason about variation. Additional opportunities to reason about variation from the data-centric perspective arose from comparable questions for the corrected summary measures and dotplot for Consultant Two's scores.
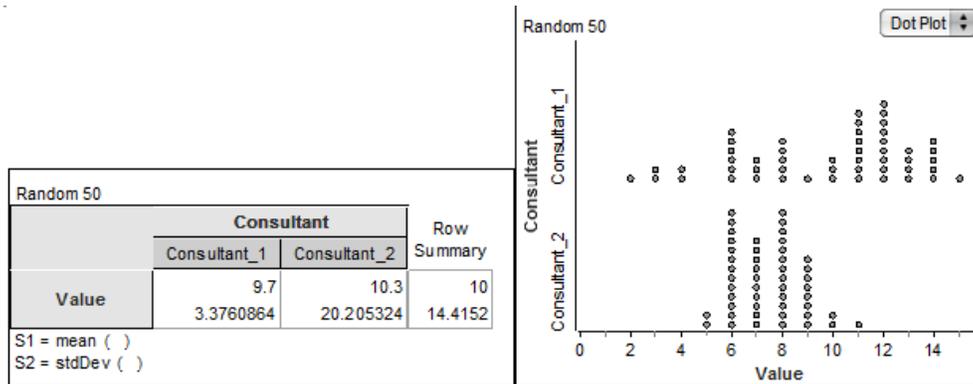

Figure 6. Summary values and dotplots for sample exam scores

A second part of the Consultant Task included questions to describe expected differences between size-15 samples and size-50 samples to ascertain teachers' perceptions of the effects of sample size on variability. Teachers were asked to evaluate the scores of the 15 randomly selected exams shown in Figure 7 from exams randomly assigned to and scored by each consultant to determine whether there was a difference in scoring. Because sample data were in tabular form and available on a graphing calculator, teachers could choose among summary measures, graphs, and other strategies to analyze data and reason about representing, measuring, and describing variation and comparing variation between consultants' scores.

Teachers who initially suggested using inferential methods to form conclusions prompted questions about what information was needed for the methods and conditions under which the methods were appropriate for drawing conclusions. Subsequent questions explored what conclusions could be drawn from using summary measures and graphs to reason about variation within each distribution and between the two distributions. Inferential comparisons provided opportunities for teachers to model the pattern of variability in data and to model sampling distributions for samples of a given size, indicative of reasoning from the modeling perspective. Interview questions related to expectations for additional samples of varying sizes selected from the same population could yield information about how teachers balanced the ideas of sample representativeness and sample variability (Rubin et al., 1990). Through the variety of situations and questions presented in the Consultant Task, teachers could consider models for data and statistics, and represent, measure, describe, and reason about variation in multiple ways.

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

Figure 7. Scores for 15 randomly selected exams

4.2. CALIPER TASK

Figure 8 and Figure 9 show the Caliper Task question and graph, respectively. Like the Consultant Task, the Caliper Task could be approached from any of the three perspectives. In the absence of context, teachers could express a need for context to consider the nature of or expected pattern of variability to reason about variation from the design perspective. Alternatively, teachers could use the scatterplot to describe the pattern of variability in the data or the strength of the relationship between x and y to exhibit reasoning from the data-centric perspective. Because the task statement asks for a response to a question, teachers could initially react to the student's question by attending to considerations for predicting the value of y when x is four to exhibit reasoning from the modeling perspective. Context was revealed to teachers after they reasoned about the data sans context. For the data shown in Figure 9, the explanatory values are measurements of objects manufactured to have a specific length measured in centimeters. Corresponding response values are students' Vernier caliper measurements to the nearest thousandth of an inch for that object.

> 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 8. Initial question for the Caliper Task

Teachers who approached the task by expressing a need for context prompted design questions about reasons for the request and the perceived legitimacy of any articulated concerns. By describing why context was important, teachers could reveal how they used context to consider sources of and types of variability for each variable. The small sample size offered additional opportunities for teachers to reason about the effects of sample size on variability and how the absence of context in combination with the small sample
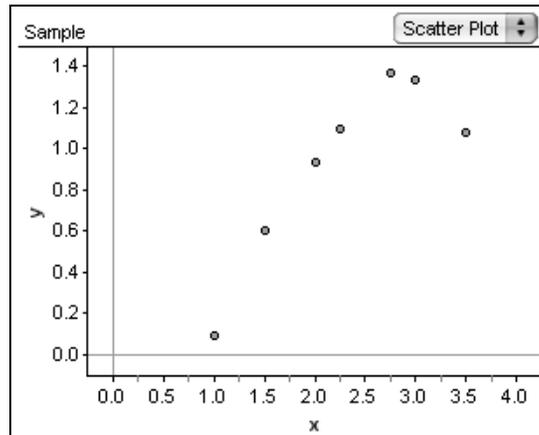
Figure 9. Initial graph for the Caliper Task

size might affect conclusions to offer insights into how teachers reason about controlling variation. After context was introduced, interview questions focused on describing specifically how the context and expectations of variation based on context affected analysis and why the data did not exactly match the known relationship between inches and centimeters. Context thus provided additional opportunities to reason about sources of variability and to offer suggestions for controlling variability from those sources. Additional evidence of how teachers reasoned about the effects of sample size could arise from reasoning about the larger bivariate data set shown in Figure 10 in response to questions similar to those asked for the initial graph.
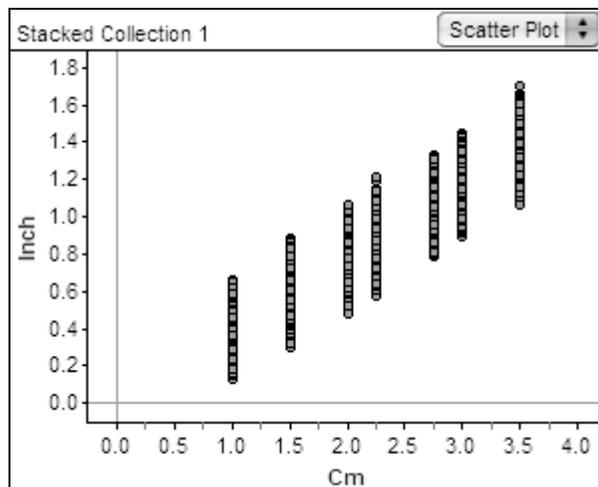

Figure 10. Scatterplot resulting from a larger data set

Interpreting the initial scatterplot to describe the pattern of variability in the data or the strength of the relationship between *x* and *y* could yield data-centric information about how teachers reasoned about the aggregate of data and how they used the measures of correlation and the coefficient of determination to reason about variability in bivariate data. Information about context prompted interview questions for teachers to use the

scatterplot as an aid to reason about reasonable variation for each variable, to describe the pattern of variability between variables, and to interpret the correlation coefficient and coefficient of determination in conjunction with the context. Additional evidence about how teachers reasoned from the data-centric perspective could stem from responses to questions about the larger set of data, the regression output (including the coefficient of determination and the standard deviation of residuals), and how residual plots could be used to reason about random variability for proposed models.

Speculation about different models for the original data prompted interview questions to describe the conditions under which each model would be appropriate and how the student could choose among models. Responses to these questions could inform how teachers modeled the pattern of variability in data and their considerations in finding a "best" fit for the data. With the context known, interview questions focused on describing specifically how context affected model selection. For the larger set of data shown in Figure 10, teachers were asked questions to describe a reasonable model for the univariate distributions formed at the seven discrete explanatory values to gather information about the extent to which context influenced expectation. Teachers' conceptions of "good" fit could be illuminated from responses to interview questions related to what univariate patterns of variability would produce different lines of best fit. Through the questions asked for the Caliper Task, teachers were afforded opportunities to reason about variation from each perspective in univariate and bivariate settings.

4.3. HANDWRITING TASK

The Handwriting Task was designed to elicit reasoning from the design perspective and stemmed from an electronic group discussion posting (Bullard, 2006). In the quote displayed in Figure 11, Pressler (2006) conjectures that a relationship exists between handwriting quality and composition scores assigned by adults. Interview questions centered on designing a study to test the conjecture and providing explanations and rationale for the designs. Questions targeted issues of replication, randomization, and control to elicit reasoning about expected variation, consideration of variation sources, attempts to control variability, and reasoning about sample size. If teachers did not mention sample size or blocking, interview questions prompted teacher to consider them to inform how teachers reasoned about the effects of control and sample size on the variability of samples and sample statistics.

> "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 11. Handwriting Task excerpt for study design

4.4. TASK SUMMARY

As the preceding descriptions of the interview protocol show, aspects of reasoning from the design, data-centric, and modeling perspectives can arise in multiple ways. Teachers had multiple opportunities to integrate reasoning about variation within and across the three perspectives and to reason about variation in different contexts. Teachers' responses to these tasks offered insights into reasoning about variation from each of the perspectives and how reasoning was integrated among perspectives to inform a detailed framework of reasoning indicative of robust understandings of variation.

5. DATA ANALYSIS

Pilot-study work with six AP Statistics teachers not considered for participation in the main study was conducted for the purpose of determining the extent to which tasks and interview questions elicited relational-level reasoning about variation within the design, data-centric, and modeling perspectives and across perspectives in the formal mode. Preliminary analysis led to changes in the tasks and the initial indicators listed in Figures 1, 2, and 3.

5.1. TASK CHANGES

Analysis of the six pilot-study teachers' video recorded interviews and annotated interview transcripts began by first matching passages of reasoning about variation to the indicators in Figures 1, 2, and 3. Indicators not addressed by a majority of teachers led to adaptations and additions to the interview tasks and questions. For example, one change to the Consultant Task occurred for the data-centric indicator that includes "creation, use, or interpretation of various data representations." The original task presented teachers with summary measures and dotplots for size-50 samples and dotplots for size-15 samples. Teachers reasoned from the size-50 dotplots and summary measures but reasoned strictly from the dotplots displayed in Figure 12(a) for the size-15 samples. To achieve a higher probability of teachers creating representations and reasoning about that creation, data were presented in tabular form for the main study [Figure 12(b)].

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
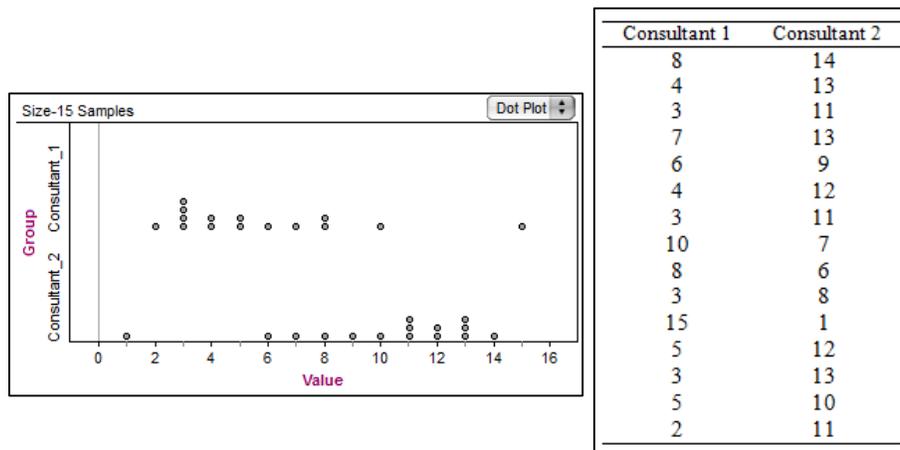
Figure 12. (a) Original presentation of size-15 samples and (b) adapted presentation of size-15 samples

5.2. INDICATOR CHANGES

The focus of analysis for the six pilot-study interviews shifted from considering interview questions and tasks to considering how well the emerging list of indicators captured reasoning about variation. Analysis of teachers' reasoning for the six teachers in the pilot study and the 16 teachers in the main study followed the constant comparative method (Glaser & Strauss, 1967) and consisted of identifying passages aligned with indicators and identifying aspects of reasoning about variation not captured by the set of indicators. If a teacher's reasoning was valid but did not align with established indicators,

the essence of the reasoning was noted, and indicators were added, changed, or refined. For example, aspects of teachers' reasoning during the Consultant Task included anticipating or acknowledging variability in ways unrelated to design. The design indicator of "anticipation and acknowledgement of natural variability" and all other indicators did not fully capture reasoning to compare summary measures and size-50 sample dotplots. In one instance, after looking at the standard deviation for Consultant Two's scores, a teacher stated that he was "concerned about how large the scale is. It only goes to 15.... I don't see how the average distance could be 20 if the scale is up to 15 … the standard deviation gives us information about variability around the mean." The teacher did not seem to be surprised by the existence of variation, as he earlier stated a clear expectation for scores to vary. His identification of an unreasonable standard deviation value suggests the value was not aligned with what he anticipated for this setting. His reasoning contributed to the addition of an indicator for the data-centric perspective: anticipating reasonable variability by recognizing unreasonable variability. More detail about the final list of indicators appears in the Results section.

5.3. ELEMENT EVOLUTION

Initial analyses informed the final framework that includes four considerations or aspects of variability that transcend perspectives—hereafter referred to as *elements*—and observable indicators for elements from each perspective. Elements are indicative of aspects of reasoning that connect across perspectives—integrated reasoning. The element of "effects of sample size on variability" was the first to be identified. The importance that pilot-study teachers attributed to sample size and their accompanying reasoning was not reflected in the initial list of indicators shown in Figures 1, 2, and 3. Only one indicator in the design perspective explicitly referenced sample size. When pilot-study teachers reasoned about how outliers affected the mean and variability of a sample by describing a particular sample and how the effects of an outlier lessen as sample sizes increased, their reasoning extended to the data-centric perspective. Similarly, when teachers described how a collection of statistics was affected by sample size and related that collection to inference, their reasoning was from the modeling perspective. Teachers' reasoning from all three perspectives led to the element of considering the effects of sample size on variability.

Consideration of other aspects of reasoning that cut across perspectives led to identification of three more elements: anticipating and acknowledging variability, reasoning about variability for contextual variables, and reasoning about variability in relationships among data and variables. Reasoning with a *variational disposition* occurs when expectation of variation accompanies a need to consider and implement design strategies for collecting data. Recognizing unreasonable variability in data or allowing for reasonable variability in predictions provide additional indication of a variational disposition. Reasoning about *variability in data for contextual variables* includes anticipating potential sources and types of variability in study design, exploring contextual data by identifying characteristics of data through representations and measurements, and fitting models to data. Reasoning about *variability in examining relationships among data and variables* 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. More detail about the elements appears in the Results section.

5.4. ANALYSIS IN THE MAIN STUDY

Analysis of data from the 16 teachers in the main study included a first pass through annotated interview transcripts to identify passages aligned with indicators and aspects of reasoning about variation that were not captured by the emerging set of indicators. Indicators continued to be added, changed, or refined. A second pass included constructing tables for each teacher, with columns labeled by perspective and rows labeled by element. Interview passages were placed in temporal sequence across perspectives for each element. Tables for each teacher provided means to note relational reasoning within and across perspectives. Throughout analysis, indicator descriptions were revised from response characteristics that were not consistent with emerging descriptions. Continued analysis using the constant comparative method (Glaser & Strauss, 1967) involved multiple readings of teachers' responses, adjustments to tables in response to revised descriptions, and further refinement of indicators until all conflicts were resolved and all aspects of reasoning about variation were captured by the collective set of indicators. Annotated interview transcripts and tables for some teachers were discussed with another mathematics education researcher until agreement was reached on the reasoning and integrated reasoning evidenced in passages.

6. RESULTS

Figure 13 shows the perspectives, elements, and reasoning indicators that emerged from analysis of the pilot and main study interviews and shows how general indicators and elements correspond with the SOLO Model. For the first cycle of levels, attention is on elements from individual perspectives. The relational level (R_1) corresponds with integrated reasoning about variation among elements within a given perspective. Robust understandings of variation are indicated by relational reasoning among elements within each perspective and relational reasoning across and among elements from the design, data-centric, and modeling perspectives in the second cycle of levels (R_2).

Table 1 displays the complete list of indicators for each element. To facilitate discussion and to avoid duplication of long phrases, labels are given for each of the perspectives, elements, and indicators in Table 1. The design, data-centric, and modeling perspectives are notated as DP, DCP, or MP, respectively. The four elements are numbered according to the order in which they appear in the table. For each element from each perspective, cell headings indicate *general indicators* of reasoning about variation for the given element from that perspective. For example, the general indicator of DP1 is "acknowledging the existence of variability and the need for study design," which represents a variational disposition from the design perspective. The cell headings in Table 1 align with the phrases in the cells shown in Figure 13.

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
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
First SOLO Cycle of Levels	$U_1 \rightarrow M_1 \rightarrow R_1$	$U_1 \rightarrow M_1 \rightarrow R_1$	$U_1 \rightarrow M_1 \rightarrow R_1$
Second SOLO Cycle of Levels		U_2 $\downarrow$ M_2 $\downarrow$ R_2	

Figure 13. Elements and reasoning indicative of robust understanding of variation

  Specific indicators are the lettered components in the table. For example, 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 specific way to exhibit a variational disposition from the design perspective. Because reasoning from the design perspective can appear in reasoning about designing studies or in critiquing studies designed by others, a numbering system for designing DP#(1x) and critiquing DP#(2x) studies is combined with lettering, where # represents element 2, 3, or 4 and x represents indicator a or b. Similarly, to discuss the large number of data-centric indicators that involve creation, use, or interpretation of data representations, a numbering system for creation DCP#(1x), use DCP#(2x), and interpretation DCP#(3x) is combined with lettering. For these indicators, # represents element 2 or 3 and x represents indicator a, b, or c.

Table 1. Indicators of robust understandings of variation—italicized indicators stem from previous empirical and expository literature

	Design Perspective	Data-Centric Perspective	Modeling Perspective
Variational disposition	DP1: 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	DCP1: 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)	MP1: Anticipating and allowing for reasonable variability in data when using models for (a) making predictions from data; or (b) making inferences from data
Variability in data for contextual variables	DP2: 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	DCP2: 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)	MP2: Identifying the pattern of variability in data or the expected pattern of variability for contextual variables by (a) modeling data to explain variability in data; (b) considering contextual variables in the formulation of appropriate data models; (c) considering contextual variables in modeling data to describe holistic features of data; or (d) considering or creating distribution-free models or simulations to explore contextual variables

Variability and relationships among data and variables	DP3: Controlling variability when (1) designing studies or (2) critiquing the extent to which variability was controlled in studies by (a) 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 (b) 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 (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	MP3: 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
Effects of sample size on variability	DP4: 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	DCP4: 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 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 (a) model the sampling distribution; or (b) consider significance, practical or statistical significance, of inferences

Problem contexts and questions and interview time limitations are likely to prohibit reasoning about every indicator. As a result, reasoning was considered to be indicative of robust understanding of variation if it included evidence of general and specific indicators appropriate to the tasks under consideration for each element within each perspective, was relational within each of the three perspectives, was relational across the three perspectives, and transcended context. Reasoning was not consistently faulty in that it might include minor misstatements that were corrected or were otherwise consistently stated correctly, and unsubstantiated claims were not made. Discussion of results primarily includes examples from Blake, Dustin, Everett, Hudson, and Isaac (pseudonyms). These five teachers exhibited reasoning consistent with robust understanding of variation, as reported in Peters (2009). Passages and descriptions of teachers' reasoning presented here are intended to illustrate relational reasoning within perspectives or across perspectives but are not intended to be sufficient evidence of robust understanding for any one teacher.

To explicate the framework and to illustrate relational reasoning about variation in the formal mode, the discussion of results includes examples of teachers' responses to the tasks described in Section 4. Discussion is organized around reasoning within perspectives in the first cycle of levels before considering reasoning across perspectives in the second cycle of levels.

6.1. RELATIONAL REASONING WITHIN PERSPECTIVES

Relational reasoning within a perspective manifests itself in reasoning that includes indicators for each element and is integrated among elements. Limitations inherent to contexts constrain reasoning to different combinations of indicator subsets. For example, relational reasoning within the design perspective might be observed throughout the course of designing and enacting a study, yet indicators for critiquing a design may be absent.

Design perspective Relational reasoning about variation within the design perspective evidences indicators in close proximity and in a coordinated and cohesive manner. Although reasoning may occur in a long, temporal sequence, the nature of design is such that considerations typically occur early in a study when designing a study, at the conclusion of a study when contemplating design strengths and weaknesses, or when critiquing a study implemented by others. Because none of the interview tasks require both designing and enacting a study, teachers reasoned about creating a design or critiquing a design but not both for any design. Reasoning in the formal mode extends beyond the context of a particular problem or setting.

The Handwriting Task provides an ideal setting for reasoning about variation when designing a study. In one teacher's design considerations, aspects of the four elements appear in a cohesive sequence that reveals a flowing thought process. The teacher's initial reaction suggests aspects of all four elements: "I guess my first thought is to create a group of homogeneous graders" (Blake). The suggestion for homogeneity reveals implicit expectation for an otherwise varied and heterogeneous group and thus acknowledges the omnipresence of variability to reveal a variational disposition [DP1]. Through homogeneity, the effects of one potential source of variability for this context [DP2]— variability from differences in adults that might affect grading [DP2(b)]—are minimized or controlled to explore the relationship between the contextual variables of adults' scores [DP3(1b)].

While reasoning through the Consultant Task, the teacher also suggests controlling for time and "any other thing that might affect the score given" (Blake) to educe the relationship between consultants' scores [DP3(1b)]. Controlling for the variable of time [DP3(1b)] minimizes the probability of encountering variability from samples selected from one grader's first set of scored exams and the other grader's last set of scored exams, the teacher's example of controlling [DP3(1b)] a potential source of variability [DP2(1b)] to reveal sophisticated reasoning about experimental design (Groth, 2003). Requesting a group of graders alludes to the effects of sample size [DP4]. This allusion becomes explicit when the teacher observes that the "small [sample size] just makes it harder to find a significant difference…because of variability" (Blake), and explicates the effects of sample size [DP4] on probable sample compositions [DP4(1a)] and sample characteristics [DP4(1b)] for samples of the same size. As exemplified by this teacher's reasoning, reasoning within the design perspective that is consistent with robust understanding of variation incorporates both general and specific indicators appropriate to the situation under consideration.

Indicators of reasoning about variation that appear when critiquing studies differ from those when designing studies; focus changes from controlling variation in the latter case to evaluating how well variation is controlled in the former. Aspects of reasoning while critiquing are exemplified in teachers' reasoning about the Consultant Task. Indicators in relational reasoning are not standalone statements but are interwoven throughout reasoning about design, as illustrated by the following passage that includes indicators from at least three of the four elements.

> I'd like to know if…both consultants scored the same exam papers…or, uh, at least that the 50 exam papers that each consultant scored were a random sample from all the exam papers…that they [administrators] had available, [which] would allow me to decide whether I thought the six tenth of a point difference on the scale…[is] due to differences in grading practices, applying the rubric, or…due to the random selection of the 50 papers…a different group of 50 papers scored by the same consultant probably would have yielded a somewhat different score. (Hudson)

Acknowledging that a different group of 50 papers is likely to produce a different mean score and that the design determines whether the difference can be attributed to chance evidences a general variational disposition [DP1] while also alluding to the importance of design for inference [DP1(2c)]. The same reasoning reveals contextual consideration for the nature of variability in data, specifically, sampling variability or variability in grading [DP2(2a)]. A suggestion that random sampling allows one to determine whether the relationship between the consultants scores is that of a difference in grading practices alludes to the controlling effects of randomization [DP3(2a)] without indicating how randomization controls variability. Although the teacher's reasoning fails to address control explicitly, the teacher's question about whether consultants scored the same exams hints at control [DP3]. The teacher later reveals that he prefers a matched-pairs design to reduce the effects of different student responses and reveal the relationship between consultants' scores [DP3(1b)]. Similarly, explicit mention of the effects of sample size does not appear in this passage but appears as the teacher continues to consider grading practices [DP4(2a) and DP4(2b)]. Although some indicators are not immediately evident in the passage, they become clearer in the teacher's extended reasoning about design aspects related to the problem. The collective set of passages suggests that indicators not explicitly addressed are considered along with those clearly present, and indicators for each element from the design perspective appear in relatively short succession.

These illustrations of relational reasoning within the design perspective stand in stark contrast to reasoning at the multistructural level and, consequently, at the unistructural level. Rather than aspects of the four elements appearing in close proximity, multistructural reasoning about variation is segmented in nature and may not address all elements. For example, a teacher, Georgia, identifies potential sources of variability for the variable of consultants' scores [DP2(2b)] but does not seem to consider ways to control variability from the sources to establish the relationship [DP3(1b)] when designing an alternative study. Georgia suggests that consultants' scores might reveal systematic bias because they "could be grouped together such that higher, you know, my honors verses low level … where the honors should have higher scores" (Georgia). Sophisticated reasoning appears in advocating for random and representative samples (Groth, 2003) [DP3(1a)]. The teacher indicates, "if they [scored exams] were not randomly selected, then I would say redo it randomly" (Georgia). Her reasoning focuses on randomly selecting exams from those scored [DP3(1a)], which does not preclude consultants scoring exams from different homogeneous groups of students. Although the teacher's reasoning for both elements is valid, each element seems to be differently instigated, and reasoning about them does not occur in tandem. Implementing the teacher's design as recommended could produce the biased samples the teacher argued against. In general, by disjointedly addressing elements, valid indicators may appear in reasoning for multiple elements and yet produce inconsistencies when examined holistically.

Data-centric perspective Unlike relational reasoning within the design perspective, relational reasoning within the data-centric perspective may not be segmented, yet it may not be clearly integrated either. Reasoning during data analysis may undergo multiple iterations to consider or reveal different relationships in or aspects of data and thus may transpire over an extended period of time. Although individuals may not overtly reason about different elements almost simultaneously to suggest relational reasoning, each thought follows from and considers prior expressed thoughts to reveal a coordinated and cohesive progression of thought indicative of relational reasoning.

Integrated reasoning about variation within the data-centric perspective is illustrated using a teacher's focus on univariate data and characteristics of the data while reasoning about the Consultant Task. A variational disposition [DCP1] is both implicit and explicit in the teacher's request for and subsequent reasoning about standard deviation values for both consultants' scores: "This seems [points to 20.2], if it's out of 15 points that seems unrealistic for a standard deviation" (Everett). Requesting measures of variation alludes to anticipating variability and acknowledging that the mean is insufficient for forming conclusions [DCP1(b)], whereas identifying an unreasonable standard deviation value based on possible scores [DCP1(c)] more explicitly indicates a variational disposition. Reasoning about other elements flows from considering a more reasonable value using the dotplots of consultants' scores. The dotplots allow for reasoning with both numerical and graphical representations—transnumeration (Wild & Pfannkuch, 1999)—to estimate Consultant Two's standard deviation [DCP2(2a)].

From the dotplot, the teacher considers the mean as a balance point—sophisticated reasoning about the mean (Mokros & Russell, 1995)—to surmise that Consultant Two's given mean is incorrect and should be approximately 7.5 before he considers standard deviation. His reasoning about variation follows from and is consistent with earlier observations about unreasonable variation. He notes, "I could go two units either way and capture 45 out of the 50 [see Figure 14] … it's reasonably normal [traces normal path over dotplot for Consultant 2] … a quick estimate would be about 2 units" (Everett). The

estimate results from considering deviation from the mean [DCP2(2a)]—sophisticated reasoning about variation (e.g., Reading & Shaughnessy, 2004)—that leads to the observation that 45 of 50 scores fall within two units of the mean. Observing the relatively normal pattern of variability in the data—indicative of distributional reasoning (e.g., Bakker & Gravemeijer, 2004; Ben-Zvi, Gil, & Apel, 2007) that focuses on aggregate data features [DCP2(b)]—lends further support for the conservative estimate of two.
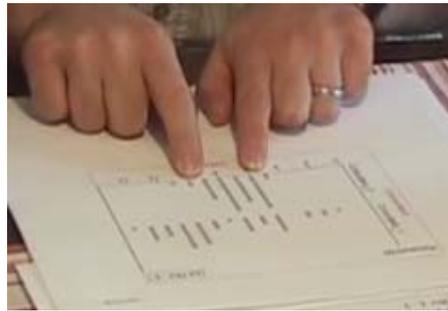
Figure 14. Examining the spread of data relative to center

Similar sophisticated reasoning surfaces when the teacher considers the variability within Consultant One's scores [DCP2(1c)] before considering the variability between distributions (Reid & Reading, 2008) when reasoning about the relationship between consultants' scores [DCP3(d)]. The teacher observes, "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 …. The standard deviation's larger … the range is obviously larger" (Everett). Informal inferential reasoning becomes apparent in comparisons using aggregate measures of average, standard deviation, and range [DCP3(2c) and DCP3(d)] to determine whether consultants' scores are consistent with each other [DCP3]. This reasoning leads to two difference relationships between consultants' scores: variability within each distribution and variability between distributions [DCP3], particularly variability in means. Further inferential reasoning appears from comparing the difference in means relative to standard error to consider the effects of sample size [DCP4] on the variability of the differences in means [DCP4(1b)].

This teacher's reasoning addresses each element within the data-centric perspective. He identifies unreasonable variation [DCP1] and estimates a more reasonable value from the dotplot [DCP2] to reason about the variation within each distribution [DCP2] before comparing the variation within and between distributions [DCP3] while considering the effects of sample size on the characteristics about which he reasons [DCP4]. Although his reasoning spans over 15 to 20 minutes and includes specific indicators as described in preceding paragraphs, it is connected. Throughout, general and specific indicators of the four elements appear in ways that continually connect to and build from prior reasoning.

Data-centric reasoning extends beyond explorations with univariate distributions to investigations of relationships between variables in bivariate and multivariate data. The Caliper Task provides a bivariate setting for reasoning about variation from the data-centric perspective. Aspects of relational reasoning can be seen in a teacher's descriptions of patterns in the data. Upon reading the task statement and after making assumptions to compensate for the lack of context, the teacher reasons about the relationship between explanatory and response variables [DCP3(2a)]. He examines patterns in the variability of the data, patterns such as "the path of a ball … the general form of a quadratic … we

could actually, uh, do a, uh piecewise graph [see Figure 15]" (Dustin). A variational disposition emerges from subsequent reasoning that a prediction [DCP1(a)] would follow only from patterns and subsequent models of the patterns that "reasonably approximate" the true relationship between variables.
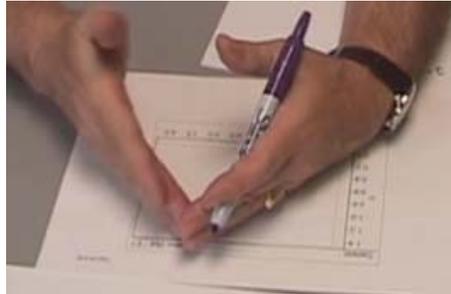
Figure 15. Examining the pattern of variability in bivariate data

Additional reasoning about elements related to the effects of sample size [DCP4] and characteristics of contextual data [DCP2] builds from reasoning about patterns.

> You've got this issue of only two points [points to rightmost two points] … we could actually talk about, uh, residual plots .… You'd like to have a reasonably high correlation, *r* value, and *r* squared. Um, and you really don't want to have this pattern [points to a drawing of a scatterplot with a fanning pattern]. (Dustin)

Reasoning about the effects of sample size is implicit in the "issue" of only two points, suggesting that the small sample size prohibits significant inferences about the relationship between variables from being made [DCP4(a)]. In addition to reasoning from the scatterplot, the teacher suggests reasoning from aggregate measures of correlation and the coefficient of determination [DCP2(b), DCP2(1c)] and from residual plots of deviations from hypothesized patterns [DCP2(1a)] in consideration of making a prediction from the data.

As with the example of relational reasoning about univariate data, indicators of reasoning from a data-centric perspective appear throughout this teacher's reasoning about the bivariate data, including reasoning about the coefficient of determination as a measure of the explained variance in measurements relative to total variance. Although some indicators and elements appear and reappear in passages beyond those presented, these passages include aspects of reasoning about variation for all four elements within the data-centric perspective and suggest a flowing and cohesive chain of reasoning. Passages from both teachers exemplify relational reasoning about variation within the data-centric perspective. In each case, reasoning about the combination of elements allows formation of preliminary conclusions about differences in scoring or the relationship between explanatory and response variables by reasoning strictly from the data.

Modeling perspective Relational reasoning about variation within the modeling perspective can be seen in reasoning about fitting models to univariate, bivariate, and multivariate distributions. Inferential reasoning—particularly for confidence intervals and formal significance tests—often indicates reasoning within the modeling perspective. Like the illustrations of relational reasoning within the design perspective, some indicators from the modeling perspective are reasoned about almost simultaneously. Other

times the indicators appear over time, similar to examples from the data-centric perspective.

Relational reasoning within the modeling perspective arises in reasoning about univariate distributions, such as those in the Consultant Task, and in reasoning about bivariate distributions, such as those in the Caliper Task. The former appears in a teacher's reasoning about what administrators should conclude. The teacher describes how expected variability in scores affects his perception of a relationship for a difference in means [MP1(b)].

> If this was a 9 point 7 and a 10 point 3 with not very much variation … I would say, hmm. These guys certainly seem to be disagreeing … if this was 9 point 7 and 10 point 3 and [they were] just swamped with variation …. I'd have to see the variation really to assess that difference. (Isaac)

Conclusions about the relationship depend on the variability in scores, an observation that suggests anticipation of some variation to reveal a variational disposition from the modeling perspective [MP1(b)]. When given measures of variation, the teacher expresses surprise at the magnitude of Consultant Two's standard deviation and appeals to a model of normality to aid in his reasoning [MP2(c)]. He presumes, "the scores are normal or mound shaped … a score scale with a shape like that would span 6 standard deviations" (Isaac). The normal model seems to be appropriate (Wilensky, 1997) for standardized test scores and suggests to the teacher that a distribution width of six standard deviations [MP2(c)] is problematic when the standard deviation is 20 and data vary between values of 0 and 15, inclusive. The teacher continues to reason as follows. A normal model approximates the pattern of variability in the dotplot of Consultant Two's scores [MP2(a)] to highlight aggregate features of the distribution [MP2(c)]. Inferential considerations for the relationship between scores include modeling the difference in means with a *t*-distribution to determine significance [MP3(a)]. Reasoning about the *t*-distribution also evokes reasoning about how sample size affects variability in calculating a test statistic [MP4(a)].

This teacher's reasoning includes indicators from each of the four elements, with some appearing almost simultaneously and others appearing over time. Overall, however, the teacher reasons about each element in a flowing manner and with a consistent focus on the relationship between scores to indicate relational reasoning about variation within the modeling perspective.

6.2. RELATIONAL REASONING IN THE SECOND CYCLE OF LEVELS

The preceding sections illustrate relational reasoning about variation within each of the design, data-centric, and modeling perspectives, which signifies relational reasoning in the first cycle of levels. *Robust understanding of variation* is indicated from additional relational reasoning across perspectives. Relational reasoning in this second cycle of levels appears in reasoning across perspectives for one or more elements and occurs through different combinations of indicators and elements due to limitations imposed by statistical questions, problems, and context.

Reasoning about design, data analysis, or inference is likely to be relational within design, data-centric, and modeling perspectives, respectively, in the first cycle of levels while spanning across perspectives in the second cycle of levels. Integrated reasoning within the design perspective and across perspectives for a variational disposition is one pattern of relational reasoning in the second cycle. A teacher's reasoning follows this pattern as he establishes details about the setting of the Consultant Task. The teacher indicates that the real question is "whether or not there's really a difference, or, just in the

random selection of the exams, um, that difference … would occur fairly often …. Without some kind of test for the differences … it might not be a significant difference" (Dustin). Suggesting that there may not be a significant difference indicates some expectation for variation while considering the effects of design on inference, a variational disposition from the design perspective [DP1(c)]. At the same time, allowing for variability in the context of inference—allowing "for the differences"—indicates a variational disposition from the modeling perspective [MP1(b)]. Allusion to random sampling and a view of a random sample as "a fair representation across the, across all the exams that were given" (Dustin) suggests that samples exhibit characteristics similar to a population without being necessarily identical and illustrates reasoning that balances sample variability and sample representativeness (Rubin, Bruce, & Tenney, 1990) in controlling variability [DP3(2a)]. Implicit acknowledgement of sample variability [DP2(a)] leads to describing situations that produce biased samples [DP3(2a)] to illustrate the importance of randomization in controlling variability [DP3(2a)]. The teacher's reasoning returns to drawing inferences about consultants' scoring when he asks for standard deviations, illustrating a variational disposition from the data-centric perspective [DCP1(b)]. In these initial considerations for the Consultant Task, the teacher's reasoning vacillates between design considerations and dispositional expressions to integrate reasoning about three elements within the design perspective and across three perspectives for the element of a variational disposition.

Reasoning across multiple perspectives for multiple elements is a second pattern of relational reasoning. The same teacher's extended reasoning about the Consultant Task illustrates this pattern. Prior to receiving the dotplots, the teacher reasons about the variability within both sets of scores—univariate data distributions—individually [DCP2(a)] before using the data to examine the variation between distributions [DCP3(d)]. He represents intervals of scores within some number of standard deviations from the mean for each set of scores [MP2(b)], as shown in Figure 16. He combines characteristics of the model [MP2(c)] with summary measures of data [DCP3(b)] to conclude that the amount of variability between distributions and the amount of variability within the second consultant's scores prohibit significant conclusions about differences in means but not differences in variances [MP3(a)]. Integrated reasoning for two elements (variability in data for contextual variables [2] and variability and relationships among data and variables [3]) from two perspectives (data-centric [DCP] and modeling [MP]) is seen in the teacher's reasoning with means and standard deviations to indicate relational reasoning in the second cycle of levels.
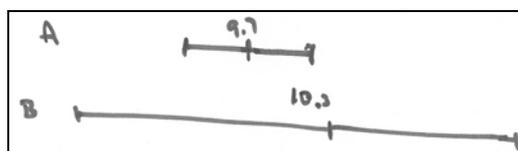

Figure 16. Model to compare scores using means and standard deviations

7. DISCUSSION

The framework for robust understanding of variation that is elucidated in this paper contributes to current empirical and expository literature in three ways. First, the framework extends the realm of reasoning about variation beyond statistical arenas heavily studied by researchers. Second, the framework illuminates reasoning in the formal mode and throughout the entire statistical problem-solving process. Third, the framework

reveals relationships among elements of reasoning about variation in place of indicators or elements that might be revealed in reasoning individually but that collectively fall short of reasoning indicative of robust understanding.

7.1. EXTENDED CHARACTERIZATION OF REASONING ABOUT VARIATION

The framework extends empirical and expository literature devoted to reasoning about and understanding variation by explicating aspects of reasoning almost entirely absent from previous research. Several indicators listed in Table 1 have received little to no attention in statistics education research. One of those indicators, for example, relates to identifying the pattern of variability in data or the expected pattern of variability for contextual variables by considering or creating distribution-free models or conducting simulations to explore relationships [MP2(d)]. Given recent focus on nonparametric inferential methods (Cobb, 2005; Holcomb, Chance, Rossman, Tietjen, & Cobb, 2010), the framework situates consideration of these methods within description of robust understandings of variation while also allowing for reasoning about parametric methods.

In addition to individual indicators, a set of specific indicators of reasoning about variation forming an element [DCP1] emerges from this study: anticipation of reasonable variation in data by considering the context of data [DCP1(a)], recognizing that data descriptions should include variability (and center) [DCP1(b)], and recognizing unreasonable variability in data [DCP1(c)]. Data are numbers with context (Moore, 1990), and researchers are focusing attention on how contextual knowledge affects data analyses (e.g., Dierdorp, Bakker, Eijkelhof, & van Maanen, 2011; Gil & Ben-Zvi, 2011; Langrall, Nisbet, Mooney, & Jansem, 2011). The work of these researchers addresses the variational disposition element by examining reasoning about anticipating and acknowledging variability from the design perspective (e.g., Petrosino et al., 2003) and the modeling perspective (e.g., Dierdorp et al.; Gil & Ben-Zvi) without attending to the data-centric perspective. The framework presented in this paper includes contextual considerations to anticipate unreasonable data variability [DCP1(a & c)] and thus extends research about reasoning with a variational disposition to the data-centric perspective.

Traditionally, research that investigates students' reasoning about variation within the data-centric perspective has focused on students describing (e.g., delMas & Liu, 2005; Lann & Falk, 2003), representing (e.g., Moritz, 2004), or comparing variation in data (e.g., Langrall et al., 2011). To the indicators of reasoning that emerge from prior work, the framework from this study adds anticipation of variability from the data-centric perspective to reveal a variational disposition by recognizing the need to characterize data using both variation and center [DCP1(b)] and using context to recognize reasonable and unreasonable variability in data [DCP1(a & c)]. This study adds to and enhances previously identified indicators and elements of reasoning about variation to offer a more complete characterization of reasoning indicative of robust understanding.

The framework for robust understanding of variation both encompasses and elaborates on outcomes of research detailing sophisticated reasoning about facets of variation. For example, a subset of indicators of reasoning about variation from the design, data-centric, and modeling perspectives shown in Table 1 corresponds with aspects of reasoning at the highest level of Reid and Reading's (2008) "consideration of variation" hierarchy. In one instance, the specific indicator of describing and measuring variability in data for contextual variables by calculating, using, or interpreting appropriate summary measures for variability in data [DCP2(c)] aligns with Reid and Reading's classification to "correctly describe variation" (p. 51). More generally, the

framework illustrates relationships among indicators, elements, and perspectives in reasoning about variation to exemplify reasoning as or more sophisticated than reasoning at the highest levels of hierarchies from previous research (e.g., Reading, 2004; Reading & Reid, 2004; Reading & Shaughnessy, 2004; Reid & Reading, 2006, 2010; Watson & Kelly, 2004; Watson, Kelly, Callingham, & Shaughnessy, 2003) while remaining consistent with statisticians' and statistics educators' expositions about what it means to understand statistical variation (e.g., Garfield & Ben-Zvi, 2005; Reading & Reid, 2010).

7.2. REASONING ABOUT VARIATION IN THE FORMAL MODE

Examples of teachers' reasoning presented throughout this paper illustrate reasoning about variation in the formal mode to characterize reasoning that is not tied to a particular context. Although the examples included in this paper might allude to a particular context, teachers' reasoning and prompted or unprompted justifications go beyond a single context. Additionally, reasoning from multiple perspectives arises in several tasks to avoid reasoning tied to a single context, as would characterize reasoning in the concrete-symbolic mode. Much previously published work investigates students' reasoning about variation in concrete-symbolic modes using specific contexts for tasks with a narrow focus such as reasoning about graphical representations (e.g., Meletiou & Lee, 2002), probability settings (e.g., spinners [Shaughnessy & Ciancetta, 2002]), or sampling (e.g., Watson & Kelly, 2002a, 2002b). Researchers call for studies to examine reasoning more sophisticated than contextually-tied, concrete-symbolic reasoning (e.g., Reading, 2002, 2004), and this study offers one response to this call by providing an image of reasoning about statistical variation in the formal mode.

The framework described in this paper reveals formal-mode reasoning about aspects of variation throughout the entire statistical problem-solving process. Researchers have proposed frameworks that more generally describe reasoning beyond individual concepts (e.g., Langrall & Mooney, 2002), reasoning throughout the problem-solving process (e.g., Garfield, 2002) or reasoning indicative of statistical literacy (e.g., Watson & Callingham, 2003). Although valuable for providing images of statistical literacy and reasoning in general, these frameworks have limitations. For example, the general framework proposed by Langrall and Mooney (2002) is limited to reasoning within the concrete-symbolic mode and focuses on mostly data-centric reasoning: "describing, organizing and reducing, representing, and analyzing and interpreting data" (p. 1). The general model proposed by Garfield (2002) resulted from generalizing a model of reasoning about samples and sampling distribution and thus offers characterizations of reasoning in broad, general terms. Useful for characterizing general reasoning throughout the problem-solving process, the framework falls short of providing a holistic, yet detailed, image of sophisticated statistical reasoning. The framework for statistical literacy posited by Watson and Callingham (2003) provides descriptions for and identifies characteristics of increasingly sophisticated levels of statistical literacy. Developed from large-scale assessment items designed for different purposes, the characterizations "reflect specific straightforward aspects of content" (p. 11) and hence do not allow for identification of subtle distinctions in reasoning. Although the framework described in this paper focuses on variation, the framework offers a detailed, holistic image of sophisticated reasoning about variation in the formal mode and throughout the problem-solving process and does so at a level of abstraction that complements descriptions of reasoning about variation in the concrete-symbolic mode from prior research.

7.3. INTEGRATED REASONING ABOUT VARIATION

The framework not only illuminates reasoning in the formal mode but also exemplifies elements that allow for integrated reasoning within and across perspectives. If the framework is used only to examine indicators in reasoning, then the framework does little to expand on the framework for reasoning about variability explicated by Garfield and Ben-Zvi (2005) or Reading and Reid (2010). The Garfield and Ben-Zvi framework describes key ideas for knowledge related to variability in seven areas, from intuitive ideas of variability to statistical thinking that incorporates consideration of variability. Reading and Reid add two more components related to recognizing sources of variation and reconciling expectation with variation from expectation. Garfield and Ben-Zvi note that deep understanding necessitates knowing a concept and knowing connections to the concept. In their words, "ideas related to variability must be constantly revisited along the statistics curriculum from different points of view, context and levels of abstraction, to create a complex web of interconnections among them" (p. 95). The framework described in this paper examines variability in multiple contexts and identifies a complex web of interconnections among multifaceted aspects of variation and interconnections between variation and related concepts. These connections take the form of integrated reasoning among indicators and elements within perspectives and integrated reasoning of indicators and elements across perspectives. The framework includes results synthesized from statistics education literature that focus on variation and offer descriptions and connections from additional empirical study to extend that literature.

8. LIMITATIONS AND IMPLICATIONS

Research suggests that students struggle to reason about the multifaceted concept of statistical variation in sophisticated ways (e.g., delMas et al., 2007; Reid & Reading, 2008). One goal of this study was to describe reasoning indicative of robust understanding of variation for researchers, instructors, and curriculum developers to form an image of the ultimate goal of instruction and to inform potential pathways from struggle to understanding. Results from this study have implications for the targeted audiences that should be interpreted with the study's limitations in mind.

8.1. LIMITATIONS AND IMPLICATIONS FOR RESEARCH

One obvious limitation of this study exists in its primary reliance on a single content interview with each teacher to evoke reasoning indicative of robust understandings of variation. The collection of tasks is intended to elicit reasoning about different aspects of variation from different perspectives through multiple opportunities for evidencing that reasoning. Nonetheless, the number of contexts is limited. Collectively, many aspects of reasoning about variation in introductory settings are considered in the framework; however, not every aspect of reasoning about variation can be captured in individual teacher's responses to the interview tasks. The possibility exists that different contexts or presentations might prompt reasoning not captured by the framework. Future research might bring to the surface aspects of reasoning that require changes to be made in the framework.

A further limitation of the framework stems from the extent to which teachers focused on content at the introductory level. Even though teachers exhibited sophisticated reasoning, there may be subtleties in reasoning about variation for content beyond the introductory level that are not captured by the framework. Further research with statistical

experts would need to be conducted to provide a definitive image of robust understandings of variation more generally—research, for example, similar to MacCullough's (2007) work that examines experts' understanding of the mean. Additional research with more diverse groups of individuals, including experts, is needed to test the framework's validity for illuminating reasoning indicative of robust understanding of variation.

8.2. INSTRUCTIONAL AND CURRICULAR IMPLICATIONS

Despite these limitations, the framework explicated in this paper reveals necessary distinctions about and connections among aspects of variation to inform how instruction can support students' development of robust understandings. A significant body of statistics education research reveals students' developing conceptions of variation (Shaughnessy, 2007). Consideration of students' views in conjunction with the framework for robust understanding yields insights for instruction. Consider, for example, students who view "variability as *change over time*" (Shaughnessy, p. 984). These students seek to find contextual explanations for time series data that deviate from patterns of variability. This search for causes or explanations appears to be relatively common for students (Reading & Shaughnessy, 2004). To build from these students' views of variation, instructors can engage students in identifying potential sources of variability and tie data analysis, sampling, and probabilistic explorations to context. Students can also consider natural, contextual, and human causes and explanations for observed variation. Seeking to attribute causation to observed relationships can naturally lead to consideration of ways to control variability. By extending discussions to consider sources of variation and ways to control variability from those sources, instructors can provide opportunities for students to integrate reasoning about variation within the design perspective and to connect reasoning about design elements to reasoning from data-centric and modeling perspectives in ways aligned with students' views of variation. Emphasis on connecting elements of design to data analysis is consistent with researchers' recommendations. For example, McClain and Cobb (2001) detail how discussing the data creation process supported their students' development of abilities to reason about data and suggest that the approach should be adopted more generally. The framework offered in this paper provides further direction regarding aspects of data creation that are important for supporting students' reasoning about variation and developing deeper understandings of variation.

The framework for robust understanding of variation may be used as a tool for considering how instruction and curriculum provide opportunities for developing reasoning consistent with robust understandings of variation. The framework may be used to analyze classroom discourse to determine whether discourse supports development of reasoning consistent with robust understandings. In this way, the framework provides a use beyond describing individual understanding; it may provide a foundation for determining relationships between learning experiences and understanding. The framework may also be used to determine whether textbooks or curriculum materials provide opportunities for students to reason about multiple facets of variation. The framework and the indicators and elements identified in the framework offer insights into places in statistics curricula where explicit attention can be given to variation during the design, exploratory data analysis, and inferential phases of statistical problem-solving and where connections among phases can be emphasized.

By explicating the framework for robust understanding of variation using teachers' reasoning about variation, this paper offers a positive image of teachers' understandings and an image of the desired goal for statistics teacher education. Expository literature

documents a widely held belief that many teachers lack knowledge in and experiences with statistics (e.g., Batanero, 2003; Ben-Zvi & Garfield, 2004; Franklin & Mewborn, 2006; Shaughnessy, 1992, 2007); however, few research studies have investigated teachers' understandings of statistical concepts. The framework provides insight into the complexity of understanding variation and what might be expected from teachers who teach or will teach statistics by considering the reasoning of teachers with robust understandings. This work contributes to and continues discussion about the crucial concept of variation as a further step towards advancing statistical understandings in the design of programs to prepare preservice and inservice teachers and their students.

ACKNOWLEDGEMENTS

REFERENCES

Bakker, A., & Gravemeijer, K. P. E. (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.

Batanero, C. (2003). Thematic working group 5, Stochastic thinking theme: Curricular issues and teacher education. In M. A. Mariotti (Ed.), *European research in mathematics education III: Proceedings of the Third Congress of the European Society for Research in Mathematics Education*. Bellaria, Italia: University of Pisa. [Online: www.dm.unipi.it/~didattica/CERME3/proceedings/Groups/TG5/TG5_introduction1_cerme3.pdf]

Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics, 45*(1)*, 35-65.

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. 3–15). Dordrecht, The Netherlands: Kluwer.

Ben-Zvi, D., Gil, E., & Apel, N. (2007, July). *What is hidden beyond the data? Helping young students to reason and argue about some wider universe.* Paper presented at Fifth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-5), University of Warwick, UK.

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.

Bullard, F. (2006, October 11). Washington Post article on penmanship [online ap-stat Electronic Discussion Group of The College Board comment]. Retrieved from http://mathforum.org/kb/thread.jspa?forumID=67&threadID=1467874&messageID=5239082#5239082

Callingham, R. (1997). Teachers' multimodal functioning in relation to the concept of average. *Mathematics Education Research Journal, 9*(2)*,* 205-224.

Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education, 10*(3).
[Online: http://www.amstat.org/publications/jse/v10n3/chance.html]

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 295–323). Dordrecht, The Netherlands: Kluwer.

Clark, J. M., Kraut, G., Mathews, D., & Wimbish, J. (2007). *The "fundamental theorem" of statistics: Classifying student understanding of basic statistical concepts.* Unpublished manuscript.
[Online: http://www1.hollins.edu/faculty/clarkjm/stat2c.pdf]

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.
[Online: http://www.causeweb.org/uscots/uscots05/plenary/]

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104*(9), 801–823.

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction, 21*(1), 1–78.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*(5), 382–393.

delMas, R. C. (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. 79–95). Dordrecht, The Netherlands: Kluwer.

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6*(2), 28–58.
[Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ6%282%29_delMas.pdf]

delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal, 4*(1), 55–82.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ4%281%29_delMas_Liu.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*(3)*,* 747–773.

Dierdorp, A., Bakker, A., Eijkelhof, H., & van Maanen, J. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning, 13*(1-2), 132–151.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education, 28*(1), 96–105.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheafer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework.* Alexandria, VA: American Statistical Association.
[Online: http://www.amstat.org/education/gaise/]

Franklin, C., & Mewborn, D. S. (2006). The statistical education of grades pre-K-12 teachers: A shared responsibility. In G. F. Burrill (Ed.), *Thinking and reasoning with data and chance: Sixty-eighth annual yearbook of the National Council of Teachers*

of Mathematics (pp. 335–344). Reston, VA: National Council of Teachers of Mathematics.

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*(2), 124–158.

Garfield, J. B. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education, 10*(3).
[Online: http://www.amstat.org/publications/jse/v10n3/garfield.html]

Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal, 4*(1), 92–99.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ4%281%29_Garfield_BenZvi.pdf]

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). Mahwah, NJ: Erlbaum.

Gil, E., & Ben-Zvi, D. (2011). Explanations and context in the emergence of students' informal inferential reasoning, *Mathematical Thinking and Learning, 13*(1&2), 87–108.

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

Groth, R. (2003). High school students' levels of thinking in regard to statistical study design. *Mathematics Education Research Journal, 15*(3)*,* 252–269.

Groth, R. E., & Bergner, J. A. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median, and mode. *Mathematical Thinking and Learning, 8*(1), 37–63.

Hammerman, J. K., & Rubin, A. (2004). Strategies for managing statistical complexity with new software tools. *Statistics Education Research Journal, 3*(2), 17–41.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ3%282%29_Hammerman_Rubin.pdf]

Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist, 27*(3), 337–364.

Holcomb, J., Chance, B., Rossman, A., Tietjen, E., & Cobb, G. (2010). Introducing concepts of statistical inference via randomization tests. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS-8)*, Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute.
[Online: www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_8D1_HOLCOMB.pdf]

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning, 12*(3), 217–230.

Langrall, C. W., & Mooney, E. S. (2002). The development of a framework characterizing middle school students' statistical thinking. In B. Philips (Ed.), *Developing a statistically literate society. Proceedings of the Sixth International Conference on Teaching Statistics*, Cape Town, South Africa. [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/1/6b3_lang.pdf]

Langrall, C., Nisbet, S., Mooney, E., & Jansem, S. (2011). The role of context expertise when comparing data. *Mathematical Thinking and Learning, 13*(1&2), 47–67.

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 (STRL-3)*. Lincoln, NE. Mount Pleasant: Central Michigan University.

MacCullough, D. L. (2007). *A study of experts' understanding of arithmetic mean.* (Unpublished doctoral dissertation). Pennsylvania State University, University Park, PA.

[Online: http://etda.libraries.psu.edu/theses/approved/WorldWideIndex/ETD-1789/]

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. [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.

[Online: http://www.stat.auckland.ac.nz/~iase/publications/1/10_18_ma.pdf]

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. 353–373). Dordrecht, The Netherlands: Kluwer.

Makar, K., & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27–54.

[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ4%281%29_Makar_Confrey.pdf]

Masnick, A. M., & Klahr, D. (2003). Error matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development, 4*(1), 67–98.

McClain, K., & Cobb, P. (2001). Supporting students' ability to reason about data. *Educational Studies in Mathematics, 45*(1-3), 103–129.

Meletiou, M., & Lee, C. (2002). Student understanding of histograms: A stumbling stone to the development of intuitions about variation. In B. Philips (Ed.), *Developing a statistically literate society. Proceedings of the Sixth International Conference on Teaching Statistics*, Cape Town, South Africa. [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.

[Online: http://www.stat.auckland.ac.nz/~iase/publications/1/10_19_me.pdf]

Meletiou-Mavrotheris, M. (2007). The formalist mathematical tradition as an obstacle to stochastical reasoning. In K. Francois & J. P. Van Bendegem (Eds.), *Philosophical dimensions in mathematics education* (pp. 131–155). New York: Springer.

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 (SRTL-3)*. Lincoln, NE. Mount Pleasant: Central Michigan University.

Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education, 26*(1), 20–39.

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. (1998). Statistics among the liberal arts. *Journal of the American Statistical Association, 93*(444), 1253–1259.

Moritz, J. (2004). Reasoning about covariation. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 227–255). Dordrecht, The Netherlands: Kluwer Academic Publishers.

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.

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, December). *Fundamental cycles in learning algebra: An analysis.* Paper presented at the 12th ICMI study conference on the future of the teaching and learning of Algebra, Melbourne.

Pegg, J., & Tall, D. (2005). The fundamental cycle of concept construction underlying various theoretical frameworks. *ZDM, 37*(6), 468–475.

Peters, S. A. (2009). *Developing an understanding of variation: AP Statistics teachers' perceptions and recollections of critical moments* (Unpublished doctoral dissertation). Pennsylvania State University, University Park, PA.
[Online: http://etda.libraries.psu.edu/theses/approved/WorldWideIndex/ETD-4200/]

Petrosino, A. J., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning, 5*(2&3), 131–156.

Porter, A. L. (2001). *Improving statistical education through the experience of reflective practice* (Unpublished doctoral dissertation). University of Wollongong, New South Wales, Australia.
[Online: www.stat.auckland.ac.nz/~iase/publications/dissertations/01.Porter.Dissertation.pdf]

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.
[Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ5%282%29_Prod_Pratt.pdf]

Reading, C. (2002). Profile for statistical understanding. In B. Philips (Ed.), *Developing a statistically literate society. Proceedings of the Sixth International Conference on Teaching Statistics*, Cape Town, South Africa. [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/1/1a4_read.pdf]

Reading, C. (2004). Student description of variation while working with weather data. *Statistics Education Research Journal, 3*(2), 84–105.
[Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ3%282%29_Reading.pdf]

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.
[Online: www.stat.auckland.ac.nz/~iase/publications/rt04/2.3_Reading&Reid.pdf]

Reading, C., & Reid, J. (2010). Reasoning about variation: Rethinking theoretical frameworks to inform practice. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS-8),* Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute.

[Online: www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_8E2_READING.pdf]

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.

Reid, J., & Reading, C. (2005). *Developing consideration of variation: Case studies from a tertiary introductory service statistics course.* Paper presented at the 55th session of the International Statistical Institute, Sydney.

Reid, J., & Reading, C. (2006). A hierarchy of tertiary students' consideration of variation. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics (ICOTS-7)*, Salvador, Brazil. [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/17/C122.pdf]

Reid, J., & Reading, C. (2008). Measuring the development of students' consideration of variation. *Statistics Education Research Journal, 7*(1), 40–59.
[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ7%281%29_Reid_Reading.pdf]

Reid, J., & Reading, C. (2010). Developing a framework for reasoning about explained and unexplained variation. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS-8)*, Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_C169_REID.pdf]

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.
[Online: http://www.stat.auckland.ac.nz/~iase/publications/18/BOOK1/A9-4.pdf]

Saldanha, L., & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*(3), 257–270.

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). New York: MacMillan.

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 20th annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 6–22). Waikato, New Zealand: Mathematics Education Research Group of Australasia.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (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, & J. Zilliox (Eds.), *Proceedings of the joint meeting of the 27th conference of the International Group for the Psychology of Mathematics Education and the 25th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 159–165). Honolulu, HI: University of Hawaii.

Shaughnessy, J. M., & Ciancetta, M. (2002). Students' understanding of variability in a probability environment. In B. Philips (Ed.), *Developing a statistically literate society. Proceedings of the Sixth International Conference on Teaching Statistics*,

Cape Town, South Africa. [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.

[Online: http://www.stat.auckland.ac.nz/~iase/publications/1/6a6_shau.pdf]

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.

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 (ICOTS-7),* Salvador, Brazil. [CD-ROM]. 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.

Snee, R. D. (1999). Discussion: Development and use of statistical thinking: A new era. *International Statistical Review, 67*(3), 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, East Lansing, MI.

[Online: www.stat.auckland.ac.nz/~iase/publications/dissertations/04.Sorto.Dissertation.pdf]

Truran, J. (1995). Some undergraduates' understanding of the meaning of a correlation coefficient. In W. Atweh & S. Flavel (Eds.), *Proceedings of the 18th annual conference of the Mathematics Education Research Group of Australasia* (pp. 524–529)*, Darwin, Australia: Mathematics Education Research Group of Australasia.

Watson, J. M. (2002). Lessons from variation research I: Student understanding. In M. Goos & T. Spencer (Eds.), *Mathematics–making waves. Proceedings of the 19th biennial conference of the Australian Association of Mathematics Teachers Inc.,* Brisbane (pp. 261–268). Adelaide, SA: Australian Association of Mathematics Teachers.

Watson, J. M., & Callingham, R. A. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal, 2*(2), 3–46.

[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ2%282%29_Watson_Callingham.pdf]

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*(2), 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*(3), 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. [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.

[Online: http://www.stat.auckland.ac.nz/~iase/publications/1/2a1_wats.pdf]

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. (2004). Expectation versus variation: Students' decision making in a chance environment. *Canadian Journal of Science, Mathematics, and Technology Education, 4*(3), 371–396.

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), 1–29.

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. 3, 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*(2)*,* 171–202.

Zieffler, A. S., & Garfield, J. B. (2009). Modeling the growth of students' covariational reasoning during an introductory statistics course. *Statistics Education Research Journal, 8*(1), 7–31.

[Online: www.stat.auckland.ac.nz/~iase/serj/SERJ8%281%29_Zieffler_Garfield.pdf]

SUSAN A. PETERS
University of Louisville
College of Education and Human Development
Louisville, KY 40292