PROSPECTIVE MIDDLE SCHOOL TEACHERS' KNOWLEDGE ABOUT DATA
ANALYSIS AND ITS APPLICATION TO TEACHING

By

Maria Alejandra Sorto

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Mathematics

2004

ABSTRACT

PROSPECTIVE MIDDLE SCHOOL TEACHERS' KNOWLEDGE ABOUT DATA
ANALYSIS AND ITS APPLICATION TO TEACHING

By

Maria Alejandra Sorto

The purpose of the study was to identify the important aspects of statistical

knowledge needed for teaching at the middle school level and to assess prospective

teachers' conceptions and misconceptions of statistics related to teaching data analysis.

An analytic study of the current literature, including state and national standards, was

conducted to identify the important aspects of statistical knowledge for teaching. A

written assessment instrument was developed and administered to a sample of 42

prospective middle school teachers. The purpose of the instrument was to gather data in

order to describe teachers' conceptions for teaching data analysis and statistics. A subset

of the sample (n = 7) was interviewed to provide deeper insight into their conceptions and

to assure reliability of the instrument.

Results show that state and national standards differ greatly on their expectations

of what students and teachers should know about data analysis and statistics. The

variation is also large for the emphasis or importance given to the content. The average

emphasis of all the documents reviewed is given to the selection and proper use of

graphical representations of data, and measures of center and spread. Important aspects

of knowledge applied to teaching are proper selection and use of teaching strategies and

inferring students' understanding from their work and discourse.

Prospective teachers that participated in this study performed better at the level of

pure statistical knowledge than at the level of application of this knowledge to teaching.

In particular, they showed abilities on reading, interpreting, and constructing graphical representations, and computing measures of center and spread. Difficulties were shown in judging students' comments and identifying students' mistakes.

To my dear husband, Alex and my adorable daughter Isabel.

ACKNOWLEDGMENTS

This dissertation has been the product of many people and institutions that in one way or another have made great contributions to its completion. First, I would like to thank my major advisor, Dr. Sharon Senk, for her guidance, help, and great effort to make this a quality work. Also, I would like to thank the members of my committee: Joan Garfield for her statistical education contributions and for serving as a great model and inspiration; Dr. Glenda Lappan for her very thoughtful comments and for providing the necessary vision; Dr. Joan Ferrini-Mundy for her broad prospective contributions and motivation, and Dr. Melfi for his encouraging comments and statistical perspective.

Second, I would like to thank the faculty and staff of the Mathematics Department; in particular I would like to thank Dr. Zhou and Dr. Ulrich for their time and effort to make me learn advanced mathematics; and especially Barbara Miller for all the things, big and little, she has done to make this possible. Also, I would like to thank the faculty and staff of the Statistics Department for treating me as one of their own.

Third, I am grateful to my colleagues and friends. To Sarah Sword and Andy Jones for their words of comfort and suggestions on draft versions of this work. To Romelia and Irvin Widders for serving as my family during my years in East Lansing.

Finally, thanks to my family. To my husband, Alex, for all his support in many ways. For being a father and a mother to our daughter in my numerous absences, for his comfort when things were difficult, for the many useful discussions about making this work meaningful, and for his love for me and respect for what I do. To my daughter, Isabel, for allowing me to spend time not playing with her or watching her teeth grow.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Images in this dissertation are presented in color

CHAPTER 1

INTRODUCTION

As statistics becomes more prevalent in the K-12 school mathematics curriculum and nearly ubiquitous in everyday discourse, the need for a statistically literate population becomes imperative. Implementation of K-12 school mathematics curriculum aligned with reform movements has challenged teachers and teacher educators in many ways. New content, such as statistics, is one of these challenges. Many experts agree that teachers' own deep and substantial knowledge of mathematics content is a key factor to provide quality mathematics education (Ball, Lubienski, & Mewborn (2001); NRC, 2001b). The inclusion of statistics topics across the school curriculum from kindergarten to Grade 12, as suggested by *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics [NCTM], 2000), provides the rationale to investigate how this new content relates to teachers' own knowledge of the subject.

The purpose of this study is to address two main questions fundamental to the preparation of middle grades mathematics teachers. First, what are the important aspects of content knowledge for teaching statistics? Secondly, how can we assess the extent to which prospective teachers possess this knowledge? Since these questions are inextricably linked to context and content, this study focuses in particular on the teaching of statistics, often called data analysis, at the middle school level. This chapter places this study within the context of current movements in mathematics education. It starts by

outlining the problems of the content preparation of teachers. Next, the case of statistics

education is described as an example of new content aligned with reform movements. As

new content, statistics poses a big challenge for teachers. In the next three sections, the

rationale, significance, and purpose of this study are stated. Research questions and

assumptions underlying this study follow.

Mathematics for Teaching

Knowing and Learning

Teaching mathematics is a practice that involves the knowledge of what, how and

whom to teach. The "what" is usually determined by a curriculum guide or textbooks

and teachers supposedly learn the content by taking mathematics classes. The "how" is

suggested in teachers' editions and is presumed to be learned in methods and education

classes. Although in practice all these components are intertwined, teachers learn them in

separate academic departments. Schools of education are usually responsible for

teaching courses in methods of teaching and educational psychology, whereas the subject

matter is the responsibility of academic departments (Ball, 2000; Lagemann, 1996).

Research on teachers knowledge (Ball & Bass, 2000; Berenson, Friel, & Bright,

1993; Bright, Friel, & Berenson, 1993; Even, 1993; Even &Tirosh, 1995; Heaton, Prawat,

& Remillard, 1992; Lloyd & Wilson, 1998; Ma, 1999; Putman, et. al., 1992; Tirosh,

2000; Russell, Goldsmith, Weingerg, & Mokros, 1990; Vacc & Bright, 1999) has shown

that many teachers have misconceptions about mathematics, have difficulty explaining in

simple terms an abstract idea and cannot correct or anticipate students' mistakes, even

though they have had the required classes to become teachers. It is clear that something is missing in their preparation to become teachers.

Thus, mathematics education researchers, mathematicians, and teacher educators are trying to answer in a more systematic manner the questions "What do mathematics teachers need to know to teach well?" and "How can teachers develop the knowledge of mathematics they need to teach well?"

Data Analysis and Statistics: New Content Promoted by Reform

Historical Prospective

The introduction of statistics in school mathematics started at the beginning of the 20th century when the Mathematical Association of America (MAA) in 1916 appointed the National Committee on Mathematical Requirements. The committee's final report published in 1923 (known as "The 1923 Report") advocated a general mathematics program for grades 7-12 which included integrated study of arithmetic, informal geometry, elementary algebra, graphs and descriptive statistics (Jones, 1970). In particular, the report suggested the study of measures of center as part of either tenth- or eleventh-grade courses.

Although "The 1923 Report" suggested an integrated approach for grades 7-12, this curriculum was only implemented at the junior high school level. This was due partly to the philosophy of mathematics at that level which is exploratory and partly to materials and textbooks developed at this time. Relative to secondary school, two major curriculum reports were published in 1940. One was, *Mathematics in General Education* by the Progressive Education Association (PEA) committee, and the other was a product

of the Joint Commission of the MAA and the National Council of Teachers of Mathematics (NCTM). The PEA committee selected broad categories of mathematical behavior including one on Data, stating "The teacher must help students become aware of differing kinds of data and their characteristics such as accuracy and relevancy. The student should acquire the ability to collect and record data, understand the measurement process, and be familiar with the construction and use of tables and graphs." (Jones, 1970, p. 226). Unfortunately, these recommendations were not implemented. A national war crisis intervened and people felt that "new and formerly 'pure' mathematics was needed by the technicians, engineers, and scientists of an expanding wartime technology" (Jones, 1970, p. 233).

After the war in 1945, the secondary school's college preparatory program based the objectives on The Harvard Report in which the key word to describe the role of mathematics was "appreciation". The report suggested a course in the senior year which comprised a survey of elementary trigonometry, statistics, and precision of measurements and use of graphs. At the same time the Commission on Post-War Plans of the NCTM reports suggested statistics as one of the twenty-nine key concepts for junior high school.

Even though clear efforts were undertaken to improve the secondary program, a survey conducted by the Educational Testing Service in 1954 indicated that there were problems. Hence, in 1955 the Committee on Examinations of the College Entrance Examination Board (CEEB) appointed a commission formed by college mathematicians, high school teachers, and college teachers of mathematics education. The suggestions of the commission published in 1959 were minor with respect to change in content but emphasized important changes in instruction and teaching. As in the previous reports,

the commission recommended descriptive statistics in grade 9 and the study of

probability with statistical application as an optional course in grade 12.   For the first

time attention was paid to the teaching of probability and statistics.

> The probability and statistics course the commission proposed represented an area
> that had never been studied seriously at the high school level. Consequently, the
> commission felt its recommendation should be accompanied by a demonstration
> of its feasibility; members of the commission prepared a textbook for this course
> and taught it in experimental classes. (Jones, 1970, p. 265)

Suggestions about including the study of statistics and probability in school

mathematics in the late 50s were reflected by the inclusion of the content in curriculum

materials developed by the School Mathematics Study Group (SMSG). The SMSG

consisted of college teachers of mathematics, high school teachers and supervisors, and

representatives of corporations such as Rand and Bell Telephone Laboratories. The study

group created a book for a one semester course on probability and statistics for twelfth

grade. The materials were tried out in schools, edited, and revised in 1958 and 1959.

By the mid 1960s the secondary school program was well established with many

exemplary experimental programs, textbooks and new materials and the reform efforts

turned their attention toward the elementary school level. In the summer of 1962,

mathematicians and representatives of the NSF met in Cambridge to discuss the state of

mathematics in elementary and secondary schools. The following year they published

Goals for School Mathematics; one of the key ideas was that "some 'feeling' for

probability and statistics was considered important for all students" (Jones, 1970, p. 291).

The Cambridge Report also included curriculum plans for grade K-2 and grades

3-6 and two proposals for grades 7-12. The suggestions of this report focus more on the

intellectual ability of children and coherence among grade band levels. Critique of the report lead to debates and controversy.

In the late 70s the National Council of Supervisors of Mathematics (NCSM), an organization composed of mathematics leaders at district, state and university levels, published its Position Paper on Basic Mathematical Skills (1977), which defined "basic skills" as including not only computation but also estimation, geometry, problem solving, computer literacy, and statistics and probability. More specifically, the report suggests that students should know how to read, interpret, and construct tables, charts, and graphs. This report, and others, prompted NCTM to appoint a committee to develop recommendations for school mathematics for the '80s. The product, *An Agenda for Action* (1980), was one of the earliest position statements from NCTM and a definitive step toward reform. It set problem solving as the curricular focus, and recommended that the definition of "basic skills" be broadened to include such mathematical skills as estimation and logical reasoning, and promoted the use of calculators and computers in the classroom at all grade levels. Another publication that awakened the general public to crisis in the United States was *A Nation at Risk* (1983). Commissioned by the National Commission for Excellence in Education (NCEE), this report lays out the critical status of students' performance and gives recommendations about content, expectations, time, and teaching. In particular, the report suggests that "The teaching of *mathematics* in high school should equip graduates to: (a) understand geometric and algebraic concepts; (b) understand elementary probability and statistics; (c) apply mathematics in everyday situations; and (d) estimate, approximate, measure, and test the accuracy of their calculations. In addition to the traditional sequence of studies available for college-bound

students, new, equally demanding mathematics curricula need to be developed for those who do not plan to continue their formal education immediately" (Retrieved June 28, 2004, from http://www.ed.gov/pubs/NatAtRisk/recomm.html).

In 1986, the Board of Directors of the NCTM established a Commission on Standards for School Mathematics to help improve the quality of school mathematics. The commission published in 1989 *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989) and later *Principles and Standards for School Mathematics* (NCTM, 2000) and placed, for the first time, statistics and probability on an equal footing with numeration, measurement, algebra, and geometry in their importance in Kindergarten through Grade 12. These standards suggest that instructional programs should enable all students to formulate questions that can be addressed with data, collect, organize, and display relevant data; select and use appropriate statistical methods to analyze data; develop and evaluate inferences and predictions that are based on data; and to understand and apply basic concepts of probability.

As before, the new content and teaching philosophy suggested by the reform movement led to the production of curriculum materials, which included data analysis and probability. For example, all of the new National Science Funcation (NSF) funded comprehensive middle school curricula – Connected Mathematics Project (CMP) (Lappan, Fey, Fitzgerald, Friel, & Phillips, 2002), Mathematics in Context (http://www.wmich.edu/cpmp), and MathScape: Seeing and Thinking Mathematically (http://www2.edc.org/MathscapeSTM) include a statistics and probability strand. At this point, as the new curricula are being implemented there is a need for professional

development and teacher preparation in statistics, since teachers may have not studied the subject for teaching in grades K-12 (CBMS, 2001, p. 114).

At the end of 2003, the American Statistical Association [ASA] and the University of Georgia held a planning conference on the statistics education of future teachers as a response to *The Mathematical Education of Teachers* (CBMS, 2001) report published in fall 2001. The participants were a group statisticians and mathematics educators which came together to start a conversation on current models of teacher education, research on how students learn, and exposition of current activities used for preparing future teachers. The group, which is known by the acronym Math/Stat TEAMS (Math/Stat Teacher Education: Assessment, Methods, and Strategies), is scheduled to meet again in 2005 to continue planning for setting an agenda on the integration of statistical content and pedagogy into teacher preparation.

Rationale

In statistics education, teachers' knowledge is of special interest since the reform and new curricula are challenging teachers not just with new teaching approaches but with new content as well. While there is consensus about their importance as basic statistical concepts in the school mathematics curriculum, few studies have been conducted on learning and teaching statistical concepts. Although in the last decade, significant advances have been made regarding students' conceptions of measures of center; we still do not know as much about how teachers learn important concepts of statistics and data analysis. This is particularly true for the concepts of variability and distribution. Despite this lack of research, new curricula emphasizing statistics have

become available for use at the K-12 levels in this country (e.g. Quantitative Literacy Series (Landwehr & Watkins, 1986), the University of Chicago School Mathematics Project (Senk et al., 1998), Data-Driven Mathematics Series (Burrill & Hopfensperger, 1997), and Connected Mathematics Project (Lappan, Fey, Fitzgerald, Friel, & Phillips, 2002). Also, in the United Kingdom, the Schools Council Project in Statistical Education has published materials for secondary students (Garfield, 1988).

At this point we only know that teachers as adults may have the same difficulties found in college students (Mevarech, 1983; Pollatsek, Lima, & Well, 1981). Even though it is important to explore teachers' conceptions of topics he or she teaches, we need to explore them in teaching contexts. Teachers need to understand concepts not just to know them and apply them in real world problems, but to use that knowledge in the context of practice to help students learn (Ball, 2000; Ball et al., 2001). More research is needed on knowledge for teaching statistics and data analysis.

The approach taken here attempts to measure certain aspects of prospective teachers' subject matter knowledge of statistics and knowledge of statistics for teaching. Several of the measurement items are posed in the context of teaching to examine teachers' decision making and knowledge of statistics. Much, if not all, the formal training teachers receive in statistics is in university classes geared, not to teachers of the subject, but instead to practitioners and to consumers of statistics. Do problems which arise in the context of teaching unveil incomplete understanding and misconceptions different from those typically assessed in these university statistics courses? When exposed to situations where they need to teach and help others learn, do prospective teacher's call upon their statistical knowledge, rely solely upon concepts and lessons

learned from methods and education classes or successfully combine the two sources?

Describing teachers' knowledge and any differences requires the creation of a

measurement tool and its administration. This is a primary goal of this dissertation.

As the importance of statistics in the middle school curriculum is a fairly new

phenomenon, it is crucial to be clear about what statistical content teachers are currently

required to teach.   In particular, what are the important concepts or big ideas that they

need to know? And what are the teaching problems that arise when they are teaching it?

Furthermore given constraints of time, money and manpower, which concepts and

problems can be feasibly measured? So, before trying to assess teachers' knowledge, it is

necessary first to find out what aspects and topics can and should be assessed.

Hence, this study will contribute to the literature on teacher preparation and

statistics education in several ways. First, it will identify the important aspects of content

knowledge for teaching statistics and data analysis at middle grades level. Second, it will

produce a reliable instrument for the assessment of statistical knowledge for teaching,

and it will measure and describe the knowledge some prospective teachers have with

respect to some of these aspects.

<div align="center">Research Questions</div>

The central questions addressed by this study are:

1. What are the important aspects of *content knowledge for teaching* data analysis and

    statistics at the middle school level? More specifically, what are the important

    statistical *content topics* taught in middle school for which teachers need to be

    prepared?; What are the *cognitive demands* (such as memorize, perform procedures

and solve non-routine problems) that are related to the content?; What are the important aspects of *knowledge for teaching* that relate to the aspects above?

2. What are the conceptions and misconceptions prospective middle school teachers have with respect to these important aspects of knowledge for teaching data analysis and statistics?

Data Sources and Assumptions

Although *content knowledge for teaching* is a construct still under development (Ball et al., 2001), for this study, knowledge for teaching will be understood as the application of content to teaching contexts. The identification of the aspects of content knowledge for teaching to be studied will be based on integrating different kinds of work: the content suggested by state and national standards at the student and teacher level; the statistical content in the middle school mathematics curriculum; research and theoretical work on learning and understanding data in particular and other statistical concepts in general; and research and theoretical work on teachers' content knowledge as used for teaching.

A written instrument designed to measure some of these aspects will given to senior prospective middle grade teachers from selected universities and followed by face-to-face interviews with selected respondents.

Significance

The vast majority of statistics presented in the media deals with the basic statistical concepts of shape, center and spread. Furthermore, clear understanding of these concepts is fundamental to any further study of statistics. Hence there is a need to

prepare teachers who are capable of knowing and explaining these concepts to their students.

To address this need, programs (e.g. ASA-NCTM Quantitative Literacy Project), have been developed and more are underway (e.g. Explorations in Statistics: The Learning Math Project) to train in-service and pre-service middle and high school teachers. Some universities (e.g. San Diego State University, Kennesaw State University and University of Missori-Columbia) offer statistics courses that are especially targeted to future statistics teachers. These efforts, however, are being carried out in the near absence of research on the conceptions and misconceptions that preservice teachers and other college students bring to the statistics classroom. Even less is known about what knowledge and conceptions teachers transfer from their time as a student to the classroom as teacher.

This study will contribute to the research effort on teacher preparation and statistics education in several ways. First, it will develop a systematic method to organize and identify the important aspects of content knowledge for teaching statistics and data analysis at middle grades level. Second, it will produce a reliable instrument for the assessment of statistical knowledge for teaching, and it will measure and describe the knowledge some prospective teachers have with respect to some of these aspects.

The Structure of this Dissertation

Chapter 2 provides the theoretical basis for this study. The first section of this chapter summarizes the recommendations about statistical content from Kindergarten to

College level. It describes the conceptions and misconceptions related to data analysis and statistics for K-12 students, college students, and pre-service and in-service teachers as reported in the literature. The last section of this chapter describes the different frameworks that exist for learning and teaching statistics and for investigating teachers' knowledge.

In Chapter 3 the first research question is considered. A variety of documents are analyzed to determine the important aspects of content knowledge for teaching data analysis and statistics at the middle school level. The first three sections review the content standards and goals for middle school students and teachers. A specialized instrument called a content matrix and its map are used to compare and contrast the standards demanded by the different documents at the two levels. The following section describes the analysis of standards for knowledge applied to teaching. The different frameworks used in the documents reviewed are described and the most important aspects identified. Finally the last section provides a summary of the results.

Chapter 4 describes the procedures used for the development and administration of a written instrument and follow up interviews used to assess content knowledge for teaching data analysis and statistics. The first section discusses the choice of items, pilot testing and reliability of the written instrument. The second section lists the procedures used to administer the written instrument including a description of the participants selected. The next section explains the process used to score and analyze the results of the written instrument. The final section gives the purpose of and method of analysis used for the follow up interviews.

The principal results of the written instrument and follow up interviews are presented in Chapter 5. The two sections report the item by item performance and overall performance, respectively.

Finally, Chapter 6 provides a summary and recommendations, and discusses limitations of this study. The first section lists major findings of the study along with its limitations, the second section discusses their significance and implications and the final section provides recommendations for future work.

CHAPTER 2

REVIEW OF LITERATURE AND THEORETICAL FRAMEWORK

This chapter is organized around the research questions. First the

recommendations about the statistical content students and teachers should know is

summarized to set the grounds for identifying the important aspects of content

knowledge. Second, research on students' and teachers' conceptions and misconceptions

are described to help understand how these are connected to content knowledge and

identify aspects of statistical knowledge for teaching. The chapter ends with a section on

theoretical frameworks on learning, teaching, and assessing statistical concepts and a

framework to measure teachers' knowledge. These help develop the instrument to

measure the desired construct of statistical knowledge for teaching.

Data Analysis and Statistical Content

Basic concepts in statistics such as measures of central tendency (mean, median,

mode) and dispersion (range, variance, standard deviation) appear everywhere. In

everyday life, newspapers, and work, statistics are used to inform (or mislead) the public

on a broad range of issues. These basic statistical concepts, traditionally called

descriptive statistics, are included in a broader concept called data analysis. Shaughnessy

(1996) points out that "the current meaning of data analysis emphasizes organizing, describing, representing, and analyzing data, with a heavy reliance on visual displays such as diagrams, graphs, charts and plots. … Conceptually, data analysis looks for patterns, centers, clusters, gaps, spreads, and variations in data" (p. 205).

Burrill (1998) also adds, "Understanding statistics and probability enables people to reason from and make conclusions based on data, judge the quality of other people's conclusions, recognize the degree of uncertainty in any endeavor, and quantify that uncertainty".

<div align="center">Statistical Content in School Mathematics</div>

Although research in teaching and learning statistical concepts is still in its infancy, its basic concepts are now being introduced in pre-college levels guided by the *Curriculum and Evaluation Standards* (NCTM, 1989) and the *Principles and Standards for School Mathematics* (NCTM, 2000); the *Guidelines for the Teaching of Statistics* (ASA, 1991); and the *Benchmarks for Science Literacy* (American Association for the Advancement of Science [AAAS] , 1993).

The study of basic statistical concepts, as recommended by most curriculum guides, should start in elementary grades allowing students to work with some data, but it is in middle grades that these understandings begin to be developed in depth. The expectations of the NCTM are that at the end of high school, students should have a full range of data-analytic skills and be comfortable designing an appropriate study for a question of interest to them, collecting data, and summarizing the results.

The *Principles and Standards for School Mathematics* (NCTM, 2000) recommends that students should begin their study of Data Analysis in preschool and

progressively build their skills until 9th grade. In Grades Pre-K-2, students begin by describing parts of the data and the set of data as a whole to determine what the data show. In Grades 3 – 5 they should continue describing the shape and important features of a set of data and comparing data sets, but with an emphasis on how the data are distributed; using measures of center, focusing on the median, and understanding what each does and does not indicate about the data set; and comparing different representations of the same data and evaluating how well each representation shows important aspects of the data. Finally in Grades 6 – 9 students complete their study of describing data by finding, using, and interpreting measures of center and spread; discussing and understanding the correspondence between data sets and their graphical representations, especially histograms, stem-and-leaf plots, box plots, and scatterplots.

These standards envision the learning of basic statistical concepts embedded in the exploration of real world data. Students must make sense of the data before they simply compute the algorithms. Furthermore, the learning of basic concepts goes along with the use of appropriate graphical representation, inferences and predictions about the data, and design of statistical experiments.

Besides the national standards mentioned here, content standards/curriculum frameworks/objectives exist for each state in the country. Other reports and documents suggest what students and teachers should know and able to do in statistics. Some of them are examined and analyzed in Chapter 3 to identify the statistical content in a more systematic way.

Statistical content at the college level has a broader scope than the content taught at K-12 level. Colleges and universities teach statistics as a general education requirement which usually consists of an introductory non-calculus based course; as part of an undergraduate degree in statistics or mathematics; and as a research tool for undergraduate and graduate programs in other disciplines. Recently, some institutions have created a new course in statistics designed specially for future teachers as well.

The content taught in these courses varies according to their purpose and level. Recommendations and guidelines for the non-calculus based introductory course are described by Cobb (1992) in the report of the Join Curriculum Committee of the ASA and the Mathematical Association of America [MAA] entitled *Heeding the Call for Change*. The report suggests three major recommendations: 1) Emphasize statistical thinking; 2) More data and concepts: Less theory, fewer recipes; and 3) Foster active learning. These recommendations are being revised and updated by the Guidelines for Assessment and Instruction in Statistics Education [GAISE] Project of the ASA. The latest draft of the GAISE College Report (J. Garfield, personal communication, June 29, 2004) builds on the previous recommendations and suggests the following: 1) Emphasize statistical literacy and develop statistical thinking; 2) Use real data; 3) Stress conceptual understanding rather than mere knowledge of procedures; 4) Foster active learning; 5) Use technology to develop conceptual understanding and analyze data; and 6) Use assessments to improve and evaluate learning. The draft also includes goals for students regarding what it means to be statistically educated and suggestions for how to implement them.

Recommendations for undergraduate statistics and mathematics majors are given by the ASA Curriculum Guidelines for Undergraduate Programs in Statistical Science (http://www.amstat.org/education/index.cfm?fuseaction=Curriculum_Guidelines) which suggest statistical, mathematical, computational, non-mathematical, and substantive area topics. They also recommend the approach to teaching these topics which is similar to the guidelines mentioned previously:

- Emphasize real data and authentic applications.
- Present data in a context that is both meaningful to students and indicative of the science behind the data.
- Include experience with statistical computing.
- Encourage synthesis of theory, methods, and applications.
- Offer frequent opportunities to develop communication skills.

As for the content recommended for future teachers, the report by the Conference Board of the Mathematical Science (CBMS) entitled *The Mathematical Education of Teachers* (CBMS, 2001) suggests that future teachers, in particular middle grade teachers, should know and be able to design simple investigations and collect data to answer specific questions; understand and use a variety of ways to display data; explore and interpret data by observing patterns and departures from patterns in data displays, with particular emphasis on shape, center, and spread; draw conclusions with measures of uncertainty; and know something about current uses of statistics in many fields.

The recommendations described in this section provide a basis for the identification of the content in data analysis and statistics teachers need to know. A more systematic analysis of other documents such as national and state standards at the student

and teacher level is necessary to identify this content. Chapter 3 addresses this issue in detail.

Conceptions and Misconceptions of Data Analysis and Statistics

Studies of student understanding of basic statistical concepts are relatively new and most of them focus on measures of center (mean, median and mode). There is very little research on measures of dispersion, and the few which do exist have only addressed the range and standard deviation. This section describes findings of research on students' conceptions and misconceptions at the different levels: K-12 students, college students, and pre-service and in-service teachers.

Kindergarten – Grade 12 Students

Measures of center

Strauss and Bichler (1988) investigated the understanding of properties of the mean with 80 children of ages 8, 10, 12, and 14. The purpose of their study was to determine the development of children's understanding of the mean and to assess the effect of the type of task presented to students (continuous, discrete, story, concrete and numerical). Strauss and Bichler found that about half of 8 year-olds and almost all 10-, 12-, and 14-year-olds could solve a task involving the property of the mean being located between the extreme values. For the property about the sum of the deviations about the mean being zero, it was found that at all ages few children offered correct judgements. An example of a task that measured this property follows:

> Children brought cookies to a party they were having.
> Some children brought many and some brought few.
> The children who brought many gave some to those
> who brought few until everyone had the same number

of cookies. Was the number of cookies given by
those who brought many the same as the number of
cookies received by those who brought few? Was it
more? Less? Why do you think so? (Strauss &
Bichler, 1988, p. 69)

The most dominant justification was "that we could not know the answer to the

question because we did not know how much material there was or how many children

received it"(p. 72). The most difficult property for children to understand was the

average as representative of the values being averaged. Given a list of data points

children could not come up with a number that would be representative of all the data.

Some young children picked the biggest number claiming that it was the biggest that best

represented the whole data set.

In summary, this study concluded that for some properties of the mean there are

development differences and that the material used or the medium had no effect.

However, a replication study by Leon and Zawojewski (1993) with fourth grade, eight

grade and college students with a paper-and-pencil test (as opposed to interviews) found

significant differences between students' responses to items in story format and

numerical format.

Although the two studies give some understanding about children's

conceptualization of the mean, neither of them provides insight in how children

understand and construct indicators of center, or what the average tells us about the

population it summarizes. In a qualitative study by Goodchild (1988), 13 and 14-year-

old students were interviewed to explore their understanding when they are confronted

with the word *average* in the context of everyday situation. When students were asked

what the sentence printed on a matchbox "Average contents 35 matches" meant to them,

they had different interpretations. The word average was viewed as a representative

number, as a measure of location, and as an expected value. The words "around",

"about", "roughly", "not exactly", and "close to" where indicators of measure of location

and were used by 15 out of 17 students (Goodchild, 1988). Four of them indicated the

idea of expectation with the use of words such as "normally" and "usual amount".

Goodchild wanted to explore these perceptions about the mean more deeply and asked

the students to hypothesize the distribution of the contents of 100 boxes. The

distributions the children constructed suggested that the students see the contents of the

boxes as variable. This particular task gave more insight into children's perception of

variability, which will be discussed in a later section of this chapter.

Mokros and Russell (1995) also looked at children's own construction and

interpretation of the mean. They interviewed fourth, sixth, and eight graders using open-

ended problems. An example of one of the problems used was to put price stickers on

pictures of nine bags of potato chips so the "typical or usual or average" price of the

chips would be $1.38. Many of the fourth graders associated the average with the mode,

making all or most of the values the same as the mean. One fourth grader who used the

modal strategy argued that the typical value was 15 cents (1.38 divided by 9) because it

occurs nine times and was the result of (mis)using the mean algorithm (p.28). Older

students (mainly eight graders) used the mean as a "middle" point, placing one value at

$1.38 and creating a symmetrical distribution around that value. "Symmetry had a strong

attraction for these students" (Mokros & Russell, 1995, p.32). However, as Mokros and

Russell pointed out, students' strategies were efficient and elegant in dealing with

symmetrical distributions but when they introduced a constraint that did not allow them

to make a perfectly symmetrical distribution (e.g. forbidding the use of the value $1.38),

all of them, except one, were stumped.

Mokros and Russell (1995) noted that some of the oldest group of children tried to

apply the property of the sum of the deviations about the mean is zero. The children

thought about this property informally using the idea of "balance" or that "a higher value

must be balanced by a lower one" (p.32). However, when they tried to apply this idea to

construct the data, none of them fully understood the way in which the data on either side

of the mean must balance. Instead, they used a phenomenon called by Mokros and

Russell as *balancing totals,* where students create data so that the *sum* of the data points

below the mean is equivalent to those above the mean. Mokros and Russell interpret this

misconception by saying "the procedure of balancing totals on either side of the average

is modeled on a notion of balance represented by a pan balance or by an equation:

Whatever is added to one side must be added to the other. Viewing the mean as the

fulcrum of a pan balance leads students to pay attention to the *value* of each data point

only, rather to the *distance* between the mean and the piece of data" (p.34).

Mokros and Russell also encountered students who relied almost exclusively on

the algorithm for finding the mean, and none of them used it effectively. "For them the

average means a series of steps involving addition and division; it is not a mathematical

object" (p.35).

The last study about students' understanding of the mean reviewed in this section

is related to student visual representation of the distribution and comparison of data sets.

Comparison of data sets is the beginning of statistical inference and it is of particular

interest because in order to compare data sets students start by using the basic concepts of

spread, center and shape. Gal, Rothschild, and Wagner (1989) gave elementary school

children pairs of data sets represented by a line or dot plot with different conditions:

distributions in which means are very different, distributions in means are close or equal,

and distributions in which have different sizes. Children were asked to decide whether

one of the groups was "better" or the same. Their responses where matched with the

actual comparison of the mean of the groups.

Virtually all the children answered correctly the problems where the means of the

groups are very different, even the one that had an outlier. The children had more

difficulty when asked to compare distributions that overlap significantly, especially with

the distributions that had the same mean and mode but different range (see Figure 2.1[1]).

```
        Group A          *                  Group B          *
                        ***                                  *
                      ******                                 *
                   _____                              ***
                                                           ***
                                                        _____
```

Figure 2.1. Distributions Presented to Children to Compare "Which Group is Better"

The results of Gal, Rothschild, and Wagner (1989) agree with those of Mokros

and Russell about young children focusing exclusively on the mode and deciding in favor

of the group that had the "tallest" column, but without consideration of the actual value of

the modal column. But the most difficult problems were the ones with different group

sizes, "only about one third of the 3th-graders, and two thirds of the 6th-grades, gave any

---

[1] Adapted from "Which Group is Better?: The Development of Statistical Reasoning in Elementary School Children," by I. Gal, K. Rothschild, and D. Warner, 1989, paper presented at the meeting of the Society for Research in Child Development, Kansas City, MO, p. 7.

indication that group-size information was taken into account in forming a decision" (p. 3).

Besides looking at the responses of the students, Gal and colleagues looked at the strategies used. There were students who summarized the data by estimating the mean, or looking where the "bulk of data" lay in each group. Others concentrated only on some part of the data (e.g. the "tallest" column) or trying to "balance" high and low scores within a group without using that knowledge to compare the groups. Watson and Moritz (1999) extended this study to 88 students in grades 3 to 9 to explore their developmental understanding. Their findings suggest that comparison of two equal size data sets with a visual approach is appropriate for grade 3. For older students it is possible to use the comparison of groups to develop higher order statistical thinking.

Less attention has been given to the other measures of location, the median and the mode. Carvalho (1998) cited in Batanero (2000) observed the following computational errors in calculating the mode, the median and the mean:

- Mode: Take the highest absolute frequency.

- Median: Failing to order the data before calculating the median; take the central value of the absolute frequencies ordered increasingly; take the mode instead of the median.

- Mean: Take the mean of the frequency values; not to take into account the absolute frequency of each value when calculating the mean.

Cobo and Batanero (2000) also point out that the algorithm of the median is a complex one, as it is not uniquely defined. It depends on whether the number of data points is even or odd, and on whether the data is presented in grouped or ungrouped form.

Measures of spread

Goodchild (1988), in his attempt to explore deeply students' understanding of expectation, asked how much difference from 35 they would expect in one matchbox when the average across matchboxes is 35. The answers range from one or two to five (i.e. $35 \pm 1, 2, 5$). He asked them the same question for 100 boxes and the majority of students thought it would be 10 or more (i.e. $350 \pm 10, 15$). Goodchild says that "one of the reason the pupils do not make allowances for the swamping affect of a large number of boxes may be because they see all the boxes having the same number of matches in them, and 'average contents' being used only as a guide to that number, rather than as a descriptor of the outcome as a stochastic process" (p. 79). Thus, to explore this possibility further, he asked them to hypothesize a distribution for the content of 100 boxes. Students' distributions actually showed that they see the contents as variable but without any particular *form*. Goodchild mentions that older students were able to give a symmetric bell-shaped distribution. What does this tell us about students' understanding of the variance? The answer is still pending.

<center>*College Students*</center>

Measures of center

Although college students can easily compute the mean of a group of numbers by the algorithm "add-them-up-and-divide", results from studies by Pollatsek, Lima, and Well (1981) and Mevarech (1983) show that a large proportion of them do not understand the concept of weighted mean when calculating the average of averages. Furthermore, computing the simple mean was the only method they had available (not just the easiest

or most obvious) to attack the problem. One of the problems presented by Pollatsek and his colleagues to students was the following:

> There are ten people in an elevator, four women and six men. The average weight of the women is 120 pounds, and the average weight of the men is 180 pounds. What is the average of the weights of the ten people in the elevator?

Although only four students were interviewed for this particular problem, two of them calculated the unweighted mean of 120 and 180 (150 pounds). Mavarech (1983) extended the work of Pollatsek et al. (1981) by investigating students' misconceptions of other properties of the mean. She tested 103 freshman and sophomore college students majoring in education, all with at least one course in statistics completed. She found that students mistakenly attribute group structure properties like associativity, identity and closure to the operation of computing means. For example, only 40% of the students recognized the following misconception that concerns the "identity element":

> A score of zero (0) was added to a set of 5 scores (52, 68, 74, 86, and 90) with a mean equal to 74. What is the mean of the new set?
> Solution: The mean will not be changed because adding zero to the sum does not change the sum.

As the college students studied by Pollatsek et al. (1981) these students thought it possible to "average averages" (closure) by "add-them-up-and-divide" algorithm. These results are especially troubling for teachers, because they occurred after the students had an elementary course on descriptive statistics.

In summary, college students have difficulties solving weighted and simple average problems and these problems can be attributed to the fact that students think of the operations of the average as binary operations satisfying the four laws of an additive

group. Furthermore, conceptual understanding did not seem to be modified by exposure to formal instruction in statistics. These two studies were conducted with college students and more research was needed, at that time, in identifying specific aspects of conceptual knowledge such as properties and interpretations.

As mentioned before, the literature about the learning of measures of dispersion is extremely scarce. Even the few studies that try to get at how students understand variability are within other contexts and little can be learned from them.

Mevarech (1983) besides looking at the conceptual understanding of the mean in terms of its non-additive group properties, looked at the concept of variance.   She found that all students could calculate and recognize the formula for variance but only a few students possessed the conceptual knowledge to solve the problems. That is, students thought that in order to calculate the overall variance of different groups they needed to calculate the variance of the variances, or that adding zero to the data would not change the value of the variance. Unfortunately, Mevarech does not provide examples of the tasks used to measure these misconceptions nor does she provide much elaboration.

Measures of spread

Loosen, Lioen, and Lacante (1985) investigated the interpretation of the words variation, dispersion, and spread among 154 college freshmen who had not received any instruction in dispersion. They showed two different sets of blocks A and B. The length of the blocks in set A were 10, 20, 30, 40, 50, and 60 cm. and the length of the blocks in set B were 10, 10, 10, 60, 60, and 60 cm. When they asked the students which of the two sets presented more variability, they got the following responses: 50% of them thought that the set A was more variable, 36% that set B was more variable, and 14% that both

had the same variability. Loosen and colleagues interpret these responses as evidence that the intuitive concept of variability is equivalent to "not similar", that is, how do the values vary with respect to each other, as opposed to varying with respect to a fix point. In that sense, set A is in fact more variable, however the standard deviation is bigger for set B.

Pre-service and In-service Teachers

Although limited, there exists some research on teachers' knowledge of statistics and the need for more research has been expressed. Russell et al. (1990) point out that as researchers become more interested in how students learn basic statistical concepts, they become curious about the teachers' misconceptions and understandings. Furthermore, investigating the conceptions and approaches teachers bring into the classroom, investigators gain insight into what teachers are likely to view as important and the complexities they face in translating their own knowledge about describing and summarizing data.

One approach, described in Ball, Lubienski, and Mewborn (2001), focuses solely on teachers and their knowledge of statistics. That is, researchers have interviewed in-service and pre-service teachers about conceptual knowledge on topics like measures of center and graph comprehension without placing them in the context of the classroom.

Russell, Goldsmith, Weinberg, and Mokros (1990) studied teachers' knowledge of the mean and dispersion by giving teachers the same tasks they gave the students (see summary of Mokros and Russell (1995) in the section of K-12 students of this chapter). Eight teachers (2 elementary, 4 middle, and 2 mathematics coordinators for K-5) were interviewed. For the Potato Chip Problem, the task was to put prices on nine bags so that

the "typical or usual or average" price would be $1.38. Five of the eight teachers used

approaches which involved "middle" or "midpoint" while seven of them used the

algorithm at some point (usually to check if the less algorithmic approach was "right").

Different from the students, teachers used these two approaches in combination (Russell

et al., 1990).

Here is some of the language used by one of the teachers:

"Kind of like the mid-point of the middle or the average"…"I would say there had
to be a number of these above that figure and then also probably a comparable
number below the figure" (p. 4)

Teachers were successful at constructing distributions and they apparently

understood of the relationship between the data and the mean. However, when teachers

were confronted with larger data sets, which were more spread out and/or less

symmetrical, difficulties arouse. One teacher tried to choose a point of balance such as

pieces of data on one side of that point balanced by a piece of data symmetrically placed

on the other side of that point, ignoring pieces of data of value zero. Another teacher

who successfully had applied the strategy of balancing deviations in the Potato Chip

Problem tried to create a non-symmetrical distribution with large number of data points

using the same strategy used by some students, "balancing totals" (Makros & Russell,

1995). She explains her definition of balance:

Evelyn: Balance [means] that if you add something to this side of the
$1.50 you have an equal addition to the other, I mean, the equation
idea.

Interviewer: So, for instance, if you added, let's say you added six
[tiles] here at $1.00, then how would you think about…

Evelyn: Then I'd have to put something over here [greater than the
mean given] that would equal $6.00.

30

Teachers' solutions differed from students' solutions in various ways and dimensions. Teachers were more flexible in their use of two or more strategies in combination and used the arithmetic mean almost exclusively as their definition of average.

Berenson, Friel, and Bright (1993) assessed the understanding of the "center or middle of the data" and "typical" of 55 elementary teachers. The assessment instrument consisted of a non-symmetrical line plot (number of raisins in a box) and a histogram (length of cast in inches). Teachers were asked to a) determine the center or middle of the data; b) determine what is typical; and c) predict assuming the data was a representative sample. Their findings show that teachers tend to look at the center of the range or the center of the horizontal scale as the center of the data. Most teachers (like the students) picked the mode as "typical", even though the distribution was skewed. Few differences were found in the interpretation of these concepts between the two representations, but the extra information in the histogram produced alternative conceptions such as the average of the vertical scale and the size of the sample as typical.

In a comparative study among science and mathematics pre-service teachers Gfeller, Niess, and Lederman (1999) looked at the use of multiple representations in solving arithmetic mean problems. Teachers were given an instrument with 10 items (various contexts) and they were asked to solve each problem in two different ways. An example of one item is shown in Figure 2.2[2].

---

[2] From "Preservice Teachers' Use of Multiple Representations In Solving Arithmetic Mean Problems," by M. Gfeller, M. Niess, and N. Lederman, 1999, *School Science and Mathematics*, 99, p. 256.

Below is a chart depicting a data set and its mean. The data point for week 3 is missing. Place a dot on the graph that would correspond to week 3 so that the mean of all the data is 4 inches and explain how you determined that.
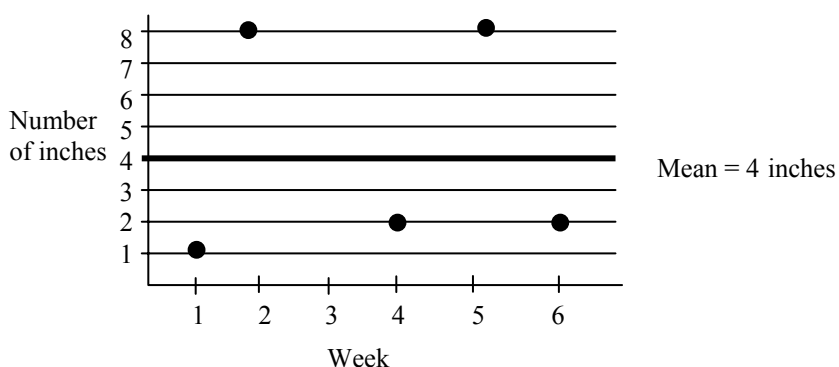
Figure 2.2. Item Given to Science and Mathematics Pre-service Teachers.

The findings of this study are consistent with previous ones. The method most often used equally by both groups was the computational algorithm, however mathematics pre-service teachers balanced deviation more often than science teachers.

As part of the implementation of a curriculum, Friel and Bright (1998) designed a project[3] to help elementary teachers in their professional development in statistics education. A variety of assessment instruments were developed, including a statistics content survey and a pedagogical survey. Major findings in the content survey were that despite instruction, "teachers continued to be quite confused about the median as a measure of center" (p.106); also, teachers demonstrated some confusion about reading data from a bar graph (looking for data values instead of frequencies on the y-axis). As for the pedagogical survey, teachers' perceptions about statistical concepts changed from an "isolated content" prospective to a "process" perspective. For example, instead of organizing the statistical concepts via simple but isolated activities such as graphing and organizing data, teacher viewed the statistical concepts in terms of more integrated

processes such as interpret data. Also, teachers showed notions of data representation in a variety of ways, extended discussions, follow-up questioning and question formulation.

Theoretical Frameworks

A variety of theoretical frameworks have been developed by researchers in the field of learning, teaching, and assessing statistics. Some researchers have proposed the theory as a result of their experience in the field (e.g. Garfield, 2002; Moore, 1997; Shaughnessy, 1992) and others have accompanied these theories with empirical results (e.g. Friel, Bright, Frierson, & Kader, 1997). Similarly, in the field of teachers' knowledge researchers have developed and tested different frameworks for understanding teachers' knowledge (e.g. Lappan, 2000; Fennema & Franke, 1992; Shulman, 1986) and frameworks to measure that knowledge (e.g. Ball, Lubienski, & Mewborn, 2001). In this section, these empirical results and frameworks are summarized.

Learning Statistics

In this subsection two frameworks of basic statistical concept development are described as guides to organize the research on learning statistics, as well as determine the knowledge of students' cognition about these concepts teachers must have. Shaughnessy (1992) suggests a model characterizing stochastic conceptions based on research results and his own practice. He identifies four types of conceptions: Non-statistical, Naïve-statistical, Emergent-statistical, and Pragmatic-statistical. These four types of conceptions can be explained in terms of students' understanding of the mean and its relation to variability.

---

[3] Gideon, Joan, Ed. (1997). *Professional Development Manual Teach-Stat for Teachers Statistics: A Key to Better Mathematics.* Palo Alto, CA: Dale Seymour Publications.

1. *Non-statistical*. When people do not operate in a statistical setting and use the mean as a representative of data with no variation. For example, the misconceptions that the mean must be one of data values or that the mean is represented by the mode.

2. *Naive-statistical*. When people understand that the mean represents data that varies and it is the balancing point, however they do not understand how the balancing occurs. For example, the misconception of "balance totals" or that the mean is represented only by the median.

3. *Emergent-statistical*. Interpret the mean as a mathematical balance with small and symmetrical data only. For example, misconceptions about taking average of averages for unequal groups and difficulties with data with large variation.

4. *Pragmatic-statistical*. In-depth understanding of the mean and its relation to variability in any context.

These type of conceptions are not necessary linear or mutually exclusive (Shaughnessy, 1992). A person does not need necessarily to be first a "naïve statistician" in order to become an "emergent statistician". Besides, people can operate in several of these conceptions depending on the settings and the nature of the task.

If we agree with Shaughnessy that most students who take beginning courses in statistics are usually in one of the first two categories, non-statistical or naive-statistical, the process of teaching and learning should aim for the last two categories. However, he points out that "the latter two conceptual stages will not occur without

carefully guided learning experiences under the tutelage of a well-trained teacher who is mathematically and statistically competent as well as sensitive to the types of beliefs and misconceptions that students have about stochastics" (p. 486).

To better understand students' conceptions and misconceptions, the Godino and Batanero (1994) framework on understanding a mathematical object can also be applied to measures of center.

- *Common errors on procedural skills*: e.g. not ordering the data before computing the median

- *Misconceptions due to notations, displays or words used to represent the concept*: confusing the sample mean and population mean

- *Difficulties in understanding and justifying the properties*: what is the effect on the mean and median of adding a large number to the data set.

- *Difficulties using the concept in relation to others*: determining the center or spread given a histogram or box plot.

Teaching Statistics

The teaching reform movement in mathematics and science has influenced the teaching of statistics and probability. However, Moore and Cobb (2000) point out the major differences between mathematics and statistics, stating that the practice of statistics is not strictly mathematical but rather is characterized by a dialogue between data and models. They stress the importance of real data and less mathematical and probability theory, as well as interpretation and communication. "Statistics combines computational

activity in a meaningful setting with the exercise of judgment in choosing methods and interpreting results" (Moore, 1990, p.96)

David S. Moore (1997) presents the following summary of research-based reformed diagnosis and prescription.

- *Goals*: Higher-order thinking, problem solving, flexible skills applicable to unfamiliar settings.

- *The old model*: Students learn by absorbing information; a good teacher transfers information clearly and at the right rate.

- *The new model*: Students learn through their own activities; a good teacher encourages and guides their learning.

- *What helps learning*: Group work in and out the classroom; explaining and communicating; frequent rapid feedback; work on problem formulation and open-ended problems.

Moore (1997) elaborates on the first and second point saying that instruction should concentrate on interpretation of graphics, strategies for effective exploration of data, basic diagnostics as preliminaries to inference. Technology is also an important part of helping students learn. See Shaughnessy (1996) for a review of research on computer software for teaching statistics. Many agree that the appropriate use of technology empowers students to do data analysis that is interactive and exploratory, using visualization and simulations to understand statistical concepts.

Although the above claims are supported mostly by research in probability and randomness (see Shaughnessy (1992) for literature review) there are some studies that focus on the teaching of basic statistical concepts. For example, Mevarech (1983)

studied the effects of exposing college students to the instruction of Mastery Learning

Strategies (MLS)[4] in overcoming statistical misconceptions (in particular the

misconceptions about treating the set of means as an additive group). Instruction for the

experimental group (N=75) differed from the instruction of the control group (lecture-

discussion strategy, N=64) only with respect to the provision of feedback and the

implementation of corrective activities.

Results showed that students in the experimental group achieved significantly

higher scores than the control group and that 70 percent of the students in the

experimental group demonstrated mastery (grade of B or higher) compared with only 40

percent in the control group. These results do not necessarily show that students

overcome their misconceptions, but Mevarech explains that the posttest to measure

achievement required more than computational knowledge, "students working with these

formative tests have to understand the concepts, apply the rules and evaluate solutions"

(p. 426).

In another more qualitative study, George (1995) compared the nature and extent

of the procedural and conceptual understanding of the mean developed by two groups of

students exposed to different forms of instruction. Six seventh-grade students

participated in the study; two of them were from a school which employed the *Visual

Mathematics* curriculum[5] and the remaining four received instruction on the numerical

"add-them-up and divide" algorithm. After giving each student five different tasks

George concluded that the students that learn the numerical algorithm were confident and

---

[4] Mastery Learning Strategies is defined by Bloom (1976) to consist of a) unit mastery requirements, b) provision of feedback, and c) corrective activities.
[5] In this curriculum, students learn the visual "leveling-off" method for finding the average of a given set of numbers.

successful in finding the mean when a complete set of data was given, regardless of size of the set and number sizes.   By comparison, students from the Visual Mathematics "revealed greater flexibility in moving back and forth between the numbers in the set and the mean" (p. 7).

Bright and Friel (1998) also did a study on the impact of an instructional unit developed specifically to highlight the connections between pairs of graphs (e.g. bar graphs for ungrouped and group data, line plot and bar graph, stem plots and histograms). They tested and interviewed students in grades 6, 7, and 8 before and after the unit and found that students don't necessarily make translations between representations easily and quickly and that students should have the opportunity to compare multiple representations of the same data set to encourage recognition of the similarities and differences in what is communicated by each representation. Although this study did not focus on the concept of mean or variance but on the overall understanding of graphical representation, its results show how different representations of data influence how students interpret properties of the data.

Statistical ideas have their own substance and modes of reasoning, hence statistics has its own pedagogy. What do teachers need to know about statistical pedagogy to help students learn? Moore (1997) suggests a "synergy" among content-pedagogy-technology. "Content and pedagogy – our understanding of what students should learn and of effective ways to help them learn- should drive our instruction. Technology should serve content and pedagogy. Yet technology has change content and allows new forms of effective pedagogy. The most effective teachers will have a substantial knowledge of pedagogy and technology, as well as comprehensive knowledge about and

experience applying the content they present" (p.134). Moore suggests the following

framework called "Synergy in Statistical Education":

- Content $\Leftrightarrow$ Pedagogy

    Data analysis $\Leftrightarrow$ Hands-on work

    Statistics in practice $\Leftrightarrow$ Communicate, cooperate

    Concepts $\Leftrightarrow$ More explanation, less proof

- Pedagogy $\Leftrightarrow$ Technology

    Visualization (Multiple representations) $\Leftrightarrow$ Automate graphics

    Problem solving $\Leftrightarrow$ Automate calculations

    Active learning $\Leftrightarrow$ Multimedia

- Technology $\Leftrightarrow$ Content

    Computing $\Leftrightarrow$ Data analysis, diagnostics, bootstrap,…

    Automation $\Leftrightarrow$ More and larger concepts

    Simulation $\Leftrightarrow$ Alternative to proofs

Moore (1997) expands on each of the components but of special interest in the

pedagogy of statistics is the use of data visualization and technology. Tufte (1983) as

cited in Shaughnessy (1997) suggests that excellence in statistical graphics consists of

complex ideas communicated with clarity, precision, and efficiency. "Students can see

an enormous variety of data in the media, which vary in appropriateness and which

sometimes are intended to deceive…Such distortions can provide points of departure for

data-handling topics" (p. 218). Computer software has long been available for statistical

analysis, however, its role in teaching and learning is still evolving. As in teaching

mathematics, as the capabilities of technology increase it is important to consider the

most appropriate use of technology in facilitating students' learning of statistics and consider potential disadvantages of using computing in statistics courses (Rubin, 1991).

Assessing Statistics

Assessment tools and frameworks in statistics, as in many fields, have a variety of purposes. Some measure student performance with the goal of assigning a grade, others are designed to measure students' knowledge and understanding with the goal of learning their conceptions and misconceptions to improve instruction. Whatever the purpose is, the assessment tool or instrument should focus on providing sufficient information about student learning and should be aligned with curricular goals (Garfield, 2003).

At the K- 8 level, Friel, Bright, Frierson and Kader (1997) suggest a framework for assessing knowledge and learning in statistics, in particular, graphical representations. The model consists of designing tasks or items that are 1) meaningful to students at the proper age; 2) ask different kinds of questions focusing on the process of data reduction; and/or 3) ask different kinds of questions to reflect the different levels of graph comprehension (read the data, read between the data, and read beyond the data). Friel et al., (1997) conclude that their work in the area of data representation "has provided a model of one way to build both understanding based on consideration of the development of graph knowledge and of strategies of assessment that may be used to support and evaluate this development" (p. 62).

At the college level, Garfield (2003) and Lui (1998) provide a valid and reliable instrument to measure the construct of *statistical reasoning*. The instrument consist of 20 multiple-choice questions carefully designed to identify correct reasoning skills and misconceptions in statistics and probability concepts. Items for the assessment were

based on the nature of reasoning skills in statistics and the identification of misconceptions found in research of student learning.

Another framework for instruction and assessment that builds on the work of Garfield's is the one suggested by delMas (2002). He suggests considering three instructional domains, basic literacy, reasoning, and thinking. *Basic literacy* refers to identification or recognition, computation, construction of graphs; *reasoning* refers to explaining why or how results were produced or why a conclusion is justified; and *thinking* refers to the application of students' understanding to real world problems, to critique, evaluate, and generalize.

<center>Teachers' Knowledge</center>

Scholars during past years have proposed several frameworks and models of teachers' knowledge. They all provide insight into defining teacher knowledge and new ways of thinking about it. The approaches are not inconsistent but rather they build upon each other. In order to identify useful conceptions of teacher knowledge taken into account by the various aspects of the models proposed, a review of the consistencies, inconsistencies and relationships follows.

Shulman (1986) proposed a framework that analyzes teachers' knowledge by considering different categories. These categories are: subject-matter knowledge, pedagogical content knowledge, and curricular knowledge. *Subject-matter* knowledge is the "amount and organization of the knowledge per se in the mind of the teacher" (p. 9*)*. *Pedagogical content* knowledge includes "the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and

<center>41</center>

demonstrations – in a word, the ways of representing and formulating the subject that make it comprehensive to others." He also includes "an understanding of what makes the learning of specific topics easy or difficult, the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning of those most frequently taught topics and lessons" (p. 9). *Curricular* knowledge is the "set of characteristics that serve as both the indications and contraindications for the use of particular curriculum or program materials in particular circumstances" (p. 10).

Peterson (1988) builds on Shulman's framework arguing that it is not only important for teachers to know how students think and how to facilitate growth in student thinking but also self-awareness of her or his own thinking in mathematics. She claims that knowledge of content will be useless in structuring the classroom so that students can learn unless teachers understand their own cognitive processes.

Lloyd and Wilson (1998) also extend Shulman's types of knowledge by defining *content conceptions* "to encompass the range of conceptions that a teacher might hold about a particular topic or concept that is taught" (p. 250). The word "conceptions" for Lloyd and Wilson refer not only to knowledge but also to beliefs, understandings, preferences, and views of a particular topic. They also assume that teachers' instructional decisions and actions are closely linked to their conceptions of mathematics, teaching and the cognition of their students. Shulman, Peterson; and Floyd and Wilson describe many of the same components of teacher knowledge, emphasizing the importance of the content, its organization, and how it should be studied. In addition, Shulman and Peterson make a strong case for the importance of considering knowledge of the learner's cognition about subject matter.

42

In an attempt to built a cognitive model of teacher knowledge, Leinhardt, Greeno, Putman, Stein and Baxter (cited in Fennema & Franke, 1992) propose that the skill of teaching is determined by at least two fundamental and related systems of knowledge: subject matter (content knowledge) and lesson structure (practical knowledge). For Leinhardt and her colleagues, knowledge of lesson structure consists of organizing the lesson with an overall goal, some procedures or activities to achieve the goal, and some routines that enable the class to function smoothly. Knowledge of the subject matter includes not only mathematical knowledge but knowledge of curricular activities, methods of presentation, and assessment procedures. Leinhardt et al. tested the model on expert and novice teachers and found that the expert teacher's knowledge tended to be more organized, used richer systems of representations and presented more detailed conceptual and procedural knowledge. The only critique of the model, according to Fennema and Franke (1992), is that it lacks attention to individuals and the actual mathematics they are being asked to learn.

Another framework that can be used to understand teacher learning is that of situated knowledge, "the mathematics that teachers learn must be learned in a context that is much broader than traditional in-school learning so that the teachers' knowledge is more similar to what we have called the nature of mathematics" (p. 160). Ball and Cohen (1999) agree that teachers' knowledge, in the way it has been defined, "could be used only in complex interactions in the unpredictable situations that we call classrooms" (p. 10). Recommendations of learning experiences in context for pre- and in service teachers are reviewed by Putnam and Borko (2000). They include the used of integrated multimedia environments, written cases, student work and videotapes of lessons.

In an attempt to include all the important components of teachers' knowledge pointed out by the above scholars, Fennema and Franke (1992) propose a new model for its examination and discussion, centered on teachers' knowledge as it occurs in the context of the classroom (see Figure 2.3[6]). The model includes knowledge of the content of mathematics, knowledge of pedagogy, knowledge of students' cognition, and teachers' beliefs. All of these components agree with the ones previously analyzed, except the knowledge of mathematics.
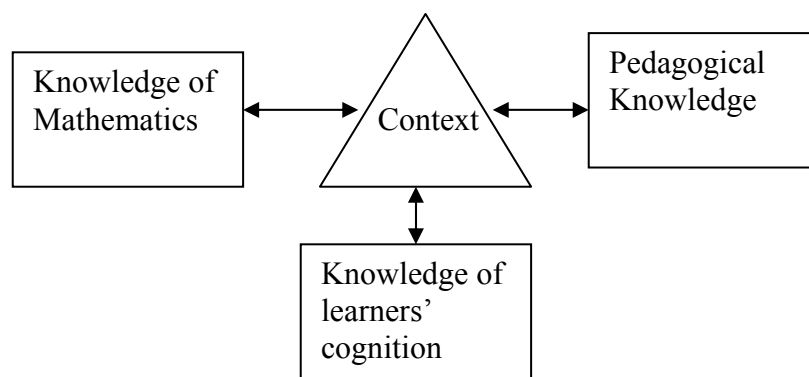
Figure 2.3. Teachers' Knowledge Developing in Context.

For Fennema and Franke, knowledge of mathematics includes knowledge of concepts, procedures within the domain in which they teach, "it also includes knowledge of the concepts underlying the procedures, the interrelatedness of these concepts, and how these concepts and procedures are used in various types of problem solving" (p. 162).

What differentiates this model from previous ones is the center triangle indicating the teachers' knowledge and beliefs in context or as situated. "Within a given context, teachers' knowledge of content interacts with knowledge of pedagogy and students'

---

[6] From "Teachers' Knowledge and its Impact," by E. Fennema and M. Franke, 1992, *Handbook of Research on Mathematics Teaching and Learning*, p. 162.

cognition and combines with beliefs to create a unique set of knowledge that drives

classroom behavior." (p. 162).

Although, Fennema and Franke make it clear that the study of the components of

teacher knowledge out of context or in isolation will not reflect the dynamic nature of

teacher knowledge, part of their model have been used by some scholars (Swafford,

Jones, Thornton, Stump, & Miller, 1999) in isolation. Swafford and her colleagues

designed a professional development program that included content courses, seminars on

pedagogical practice and research seminars on student cognition as separate components.

The study of teachers' knowledge components in isolation from each other and in

teacher education programs has caused researchers to rethink the conceptualization of

such knowledge. Lappan (2000) views the components (or domains) of teacher

knowledge in a nonlinear fashion. She proposes the interaction of knowledge domains

illustrated by the Venn diagram in Figure 2.4[7], and claims that "effective teaching occurs

in the intersection of these domains of knowledge" (p. 321).

Content for
Teaching   Pedagogy

Learning   Evaluation
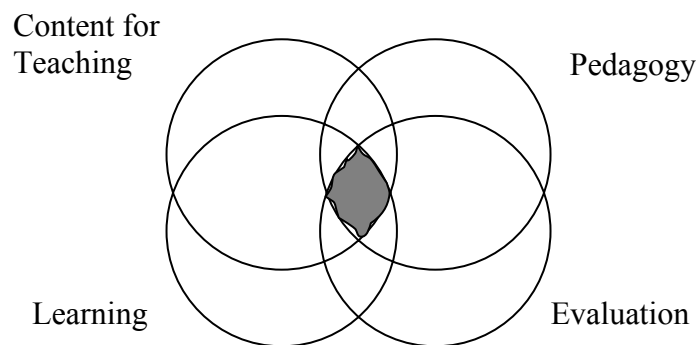
Figure 2.4. Interaction of Knowledge Domains.

---

[7] From "A Vision of Learning to Teach for the 21st Century," by Glenda Lappan, 2000, *School Science and Mathematics*, 100, p. 321.

The domain *Content for Teaching* in this model includes not just the content of mathematics but knowledge of the mathematics taught in preK-12 programs and how the ideas at different level relate. The *Pedagogy* domain includes "compelling examples of tasks that can be used to elicit particular kinds of thinking and investigation on the part of the learner" (p.322). Knowledge of student development as it is related to learning and teaching mathematics is included in the *Learning* domain. It also includes knowledge of and experience in listening to what students can make sense of and where they have problems. Although not in the diagram, Lappan (2000) also mentions as an important domain, a vertical vision of curriculum that includes the lesson, unit, school year and pre-K-12 level. Finally, Lappan adds one more component to the teacher knowledge that is the Evaluation domain, which includes assessment of students' knowledge and evaluation of programs.

As the conceptualization of teacher knowledge becomes more complex, the challenge is how to blend all the components or domains to prepare teachers to help all students learn. Ball (2000) proposes three problems that need to be solved in order to bridge the gap between content and pedagogy: 1) How to determine which content matters for teaching; 2) How subject matter must be understood to be usable in teaching; and 3) How to create opportunities to learn to use subject matter knowledge in the context of practice.

For problem number one, Ball (2000) suggests that scholars " identify core activities of teaching, such as figuring out what students know; choosing and managing representation of ideas; appraising, selecting, and modifying textbooks; and deciding among alternative courses of action, and analyze the subject matter knowledge and

insight entailed in these activities" (p.244). The second problem on how to use this knowledge is the capacity to deconstruct or unpack one's own knowledge into a less polished and final form, which is related to Shulman's definition of pedagogical content knowledge. The third problem is related to where and how to use this knowledge, that is in the context of practice. Ball suggests that scholars "design and explore opportunities to learn content that are situated in the contexts in which subject matter is used, a core activity of teaching." (p. 246). Outlining these problems, Ball mentions some of the same components or domains that other scholars do, however, she looks at it in at least two different ways. One, is that knowledge for teaching is more rooted in content knowledge; and two, she provides practical solutions for teaching and learning such knowledge which helps better understand its conceptualization.

Measuring Content Knowledge for Teaching

It is one thing to define a construct and another to identify it or to measure it. Ball, Lubienski, and Mewborn (2001) give a review of the literature on measuring teacher's knowledge using a historical prospective on teachers' mathematical knowledge. The authors explain that the identification of the mathematical knowledge is related to the way it has been measured. For example, the research that focuses on the assumption that knowledge of mathematics is essential for teaching; measures it by counting the amount of mathematics courses or workshops taken, the degree or certification earned. The research that focuses on the nature of teachers' knowledge assumes that content knowledge is intertwined with other aspects such as how students learn, and what they find difficult. This approach is basically trying to measure, in part, the construct of *pedagogical content knowledge* by interviewing teachers – or prospective teachers – to

probe their knowledge about a specific mathematical topic. Ball, Lubienski, and

Mewborn (2001) describe this method:

> Researchers working in this approach often use methods that probe teachers'
> knowledge and that situated the questions in an around questions that might arise
> in teaching. Such questions respond to a student's confusion, explain a
> mathematical procedure whose meaning is buried inside rules of thumb, or
> consider the connections among ideas. Those questions create conditions where
> teachers have to make explicit their understanding of the mathematical ideas and
> procedures behind the questions. (p. 444)

Although this approach helps understand the nature of teachers' knowledge it

does not produce a quantitative measure and thus it can not be associated to any other

factor that may influence student learning. Hence, this approach to measuring teachers'

knowledge is more attractive to researchers than to policymakers. Another limitation on

this approach, because of its qualitative nature, is that researchers can only focus on a

small number of teachers and/or on a specific topic. Most of the studies of this type have

been conducted with preservice elementary teachers in the areas of numbers and

operations, functions, geometry and measurement, and proof (for detailed review on these

areas see Ball, Lubienski, and Mewborn (2001)).

Stylianides and Ball (2004) suggest an integrative framework for studying the

mathematical knowledge for teaching which accounts for different data sources and sites

of inquiry. The authors propose six different approaches: analyzing experts'

perspectives, teacher's mathematics curricula, teachers' mathematical knowledge,

students' mathematics curricula, students' mathematical knowledge, and school

mathematics practice (see Figure 2.5[8]). The organization of these approaches is a

---

[8] From "Studying the Mathematical Knowledge Needed for Teaching: The Case of Teachers' Knowledge of Reasoning and Proof", paper presented at the 2004 Annual Meeting of the American Educational Research Association, San Diego, CA.

network of dynamic relationships that bring together practice, policy, and research. The

way Stylianides and Ball envision this system is that the first five approaches are in

coordination and alignment and that the sixth approach, school mathematical practice,

would confirm the utility of these contributions. The study of knowledge for teaching

could also start from examining school practice and see how it plays out in the actual

work of teaching. The dotted segments in Figure 2.5 suggest the existence of relations

among these elements and the dotted arrows suggest that these elements are influenced

by outside factors such as cognitive psychology, the structure of mathematical discipline,

learning and pedagogical principles among others.
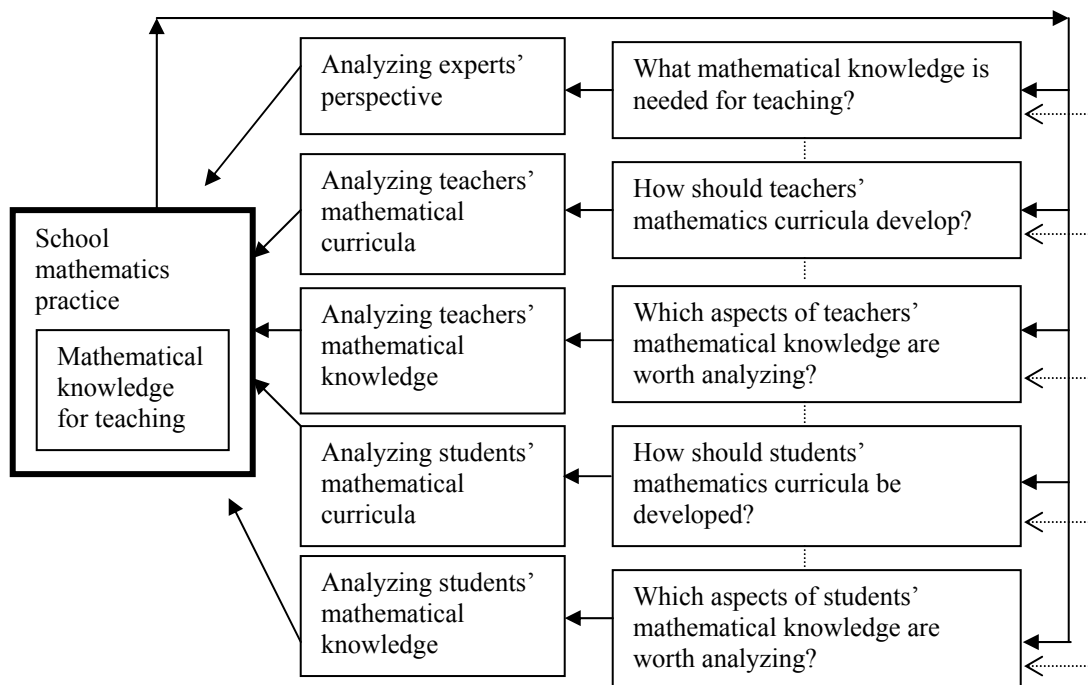
Figure 2.5. Framework for Investigating Mathematical Knowledge for Teaching

Studying the development of teachers' knowledge by looking at the different

frameworks suggested by researchers, it is clear that it has become more complex as

more is learned about the field. Investigating statistical knowledge for teaching taking

into account all of the components suggested by the latest suggested framework are beyond the scope of this dissertation. However, the first research questions mentioned in chapter 1 combines two of the components of this framework, what statistical knowledge is needed for teaching? And which aspects of teachers' statistical knowledge are worth analyzing? The starting point for this study will be analyzing experts' perspectives, and analyzing mathematical curriculum for students and teachers. From here, the component of analyzing teachers' statistical knowledge is investigated to see what can be learned about prospective teachers' knowledge for teaching data analysis and statistics.

Summary

In order to identify what knowledge is necessary to teach statistics and assess whether the teachers, in fact, possess this required knowledge, important components discussed in this chapter; content, learning, and teaching, and their interaction in particular contexts must be taken into account. Although some studies of teacher's knowledge in statistics exist, at present no studies on teachers of statistics that consider the interaction of the different components have been published.

All of the studies mentioned above focus their investigation on teachers' understanding of measures of center. As in the students' case, not much has been investigated about teachers' knowledge of measures of dispersion or other components of the process of statistical investigation, although some conclusions could be drawn from teachers' understanding of distributions. Furthermore, these studies have concentrated on the conceptual knowledge and understanding of basic statistical concepts or on how to teach these concepts to students, i.e. on content or pedagogy separately. No research appears to exist on teachers' understanding of student cognition. Are teachers, when

placed in a given context, able to combine their knowledge of statistics, pedagogy, and student cognition to create an effect learning environment? In other words do they posses the statistical knowledge for teaching? In order to investigate teacher content knowledge about a particular topic and to see its effects in instructional practices and students' achievement, it is necessary to develop a reliable and valid way to measure it.

CHAPTER 3

ASPECTS OF KNOWLEDGE FOR TEACHING DATA ANALYSIS AND

STATISTICS

This chapter addresses the first research question – what are the important aspects
of content knowledge for teaching data analysis and statistics at the middle school level?
More specifically, it seeks to identify the important aspects of content (i.e. "the big
ideas") at the student and teacher level and to identify the important aspects of knowledge
for teaching.

Several approaches are taken to investigate this question. First, the analysis of
policy documents, (e.g. national and state standards, books, and reports) is a reflection of
experts' perspectives giving insight on what it is that they value or consider important.
Second, the analysis of students' mathematical curricula (e. g. mathematics textbooks and
teacher's guide) is a reflection of what teachers are supposed to teach with an intended
curriculum. This second approach identifies what statistical knowledge teachers need to
have in order to implement these curricula in their classrooms.

Specifically, the data sources for statistical content knowledge at the student level
are content standards at the middle grades level from ten states, *Principles and Standards
for School Mathematics* (NCTM, 2000), and the Mathematical and Problem-Solving
Goals of the units on data analysis from the Grade 6 and 8 textbooks produced by the

Connected Mathematics Project (Lappan, Fey, Fitzgerald, Friel, and Phillips, 2002) Grade 6 and 8 textbooks.

The *Connected Mathematic Project* (CMP) textbooks constitute one of the recent middle grades curricula supported by the National Science Foundation.   A set of two units on Data Analysis and Probability from the *Connected Mathematics* Teacher's Guide (Lappan et al, 2002) where examined. The analysis was conducted on a list of mathematical and problem-solving goals published in each unit on the preliminary pages (*Data About Us*, page 1a; and *Samples and Populations*, page 1l) and the "big ideas" published in the introduction of the textbooks. Each goal served as the unit of analysis and the "big ideas" served as a reference to determine the intent of the goals. Although the textbooks provide many other sources for analysis such as the "summary of the investigations" and "mathematical highlights"; the lists of mathematical goals provided the most closed form of the intent of the content for the textbooks and facilitated the process of coding.

The ten states included in the analysis are the states that require middle school certification for teachers; they are Connecticut, Florida, Georgia, Kentucky, Missouri, North Carolina, Ohio, Oregon, Virginia, and West Virginia (http://www.enc.org/professional/standards/state/).

For statistical content at the teacher level the data sources are: the topics covered in the PRAXIS II Middle School Mathematics (ftp://ftp.ets.org/pub/tandl/0069.pdf), and *The Mathematical Education of Teachers* (CBMS, 2001).

For content knowledge applied to teaching, the data sources are *Knowing and Learning Mathematics for Teaching* (NRC, 2000), *Adding It Up* (NRC, 2001a), *Middle*

childhood through early adolescence/Mathematics Standards (National Board for

Professional Teaching Standards, 1998 or http://www.nbpts.org/pdf/mcea_math.pdf),

Professional Standards for the Accreditation of Schools, Colleges, and Departments of

Education (National Council for Accreditation of Teacher Education, 2002), National

Middle School Association Middle Level Teacher Preparation Standards

(http://www.nmsa.org), professional standards from four states, *Connected Mathematics*

Teacher's Guide (Lappan, Fey, Fitzgerald, Friel, & Phillips, 2002), and *The Mathematical*

Education of Teachers (CBMS, 2001).

The first section of the chapter describes and summarizes the data analysis and

statistical knowledge *students* are expected to know. The second section describes and

summarizes the statistical knowledge *teachers* are expected to know. The third section

describes and summarizes the *content knowledge applied to teaching* evident in teaching

documents and suggested by recent research.

Data Analysis and Statistics for Students in Middle Grades

National and state standards vary in form and content. In order to identify

common content and the level of emphasis suggested by national and state standards in a

systematic way, an instrument that identifies common language was needed.

Although there are several frameworks that organize content, including Bloom's

taxonomy (Bloom, B., Englehart, Furst, Hill, & Krathwohl, 1956) and TIMSS assessment

framework (Valverde et al, 2002), they are not suitable for this research.   Bloom's

taxonomy is not specific to mathematics, and the TIMSS assessment framework is not

designed to compare or find commonality among different documents. The framework

that better suits the purpose here is the *content matrix* developed by Porter, Kirst, Osthoff,

Smithson, and Schneider (1993). This instrument was adapted to meet the needs of this study.

<center>Background on Measuring Content</center>

The conceptual development of the content matrix started when Porter and others (1993) needed to measure what was taught in high school mathematics and science classes in six states. In this study, Reform Up Close, descriptors of high school mathematics and science were organized into three dimensions: topic coverage, cognitive demand, and mode of presentation. *Topic coverage* consisted of mathematics topics such as ratio, volume, and relation between operations. *Cognitive demand* included nine descriptors including memorize, perform routine and non-routine procedures, and conjecture. Among the seven *modes of presentation* were exposition, pictorial models, and equations/formulas. A content topic was defined as the intersection of all three dimensions. Porter refers to this "language" as a rich and systematic way to describe instructional content. Teachers or observers were trained to use codes for the descriptors for daily logs and observation protocols.

Although this conceptualization gave Porter and others a language and coding scheme to compare teachers' reports and observation reports of a given lesson, it had its limitations. The survey instrument could not be constructed in a simple way to include all three dimensions and not to impose a burden to the teacher. So, Porter and Smithson (2001) proposed a simpler conceptualization using only two dimensions – content category and cognitive demand – displayed in a matrix format. Furthermore, they reduced the number of categories of cognitive demand from nine to five.

Deciding which topics, cognitive demand, and labels to use is key to the internal validity of the instrument. Porter and Smithson (2001) claim that the language used in

their study is reform-neutral, with the hope that "the language should be translatable into reform language distinctions so comparison to state and other standards is possible" (p. 8).

The *Topic coverage* dimension was organized to establish a comprehensive list. For mathematics, they suggested a list of topics for elementary, middle, and high school organized by area such as numbers, operations, measurement, algebra, and data analysis and statistics. For middle school data analysis and statistics they suggest the following list of topics:

> Bar graph, histogram
> Pie charts, circle graphs
> Pictographs
> Line graphs
> Stem and leaf plots
> Scatter plots
> Box plots
> Mean, median, mode
> Line of best fit
> Quartiles, percentiles

Porter and Smithson are aware that this is not the only way to organize content, especially if the instrument is used for other purposes. They suggest an alternative to organize content using instead "big ideas" such as change or shape as the framework, which is the one desired for this study. However, they claim, that in practice, content is still organized along the traditional lines of algebra, geometry, statistics, etc. not in terms of big ideas and that this approach can be taken when practice is reformed to better reflect these big ideas.

The *cognitive demand* dimension was defined as the cognitive activities that engage students in the content topics; therefore they are behaviorally defined. The five cognitive demands used by Porter and Smithson (2001), and Porter (2002) are: memorize

facts, definitions, formulas; perform procedures/solve routine problems; communicate understanding of concepts; solve nonroutine problems/make connections; and conjecture, generalize, prove. For *memorize facts, definitions, formulas*, classroom activities focus on recalling traditional math skills and knowledge. For example, recall the formula for the mean. For *perform procedures/solve routine problems*, classroom activities focus on demonstrating basic skills, selecting and applying various computational methods. In statistics, this is interpreted as creating graphs, computing measures of center or spread, and collecting data for analysis. For *communicate understanding of concepts*, classroom activities focus on students sharing their mathematical understanding in both oral and written form. In statistics these activities include choosing the appropriate graph and measure of center and spread to answer questions about the data, as well as matching verbal descriptions of the data to distributions and their graphs. For *solve nonroutine problems/make connections* classroom activities focus on students applying mathematical knowledge to solve unfamiliar problems or seeing relationships between topics within mathematics and to other content areas. Examples of nonroutine problems include analyzing a complicated data set, recognizing patterns in data, and using statistics to explore real world problems. Finally, for *conjecture, generalize, prove* students focus on making and justifying conjectures. In the context of statistics at the middle school level, this translates mainly to informal inference and prediction from data about real world problems. Table 3.1 gives the language associated with each category.

The content of instruction is then described at the intersection between topics and cognitive demand, based on data gathered from teacher surveys. Porter asked the teachers to indicate, for the past school year (a) the amount of time devoted to each topic and then, for each topic, (b) the relative emphasis given to each student expectation.

Table 3.1

Language Frequently Associated with Cognitive Demands

A	B	C	D	E
Memorize facts, definitions, formulas	Perform procedures/solve routine problems	Communicate understanding of concepts	Solve nonroutine problems/make connections	Conjecture, generalize, prove
• Recognize • Identify • Recall • Recite • Name • Tell	• Do computations • Make observations • Take measurements • Compare • Develop fluency	• Communicate mathematical ideas • Use representations to model mathematical ideas • Explain findings and results from statistical analyses • Explain reasoning • Describe • Select	• Apply and adapt a variety of appropriate strategies to solve nonroutine problems • Apply mathematics in context outside of mathematics • Analyze data, recognize patterns • Explore • Judge	• Complete proofs • Make and investigate mathematical conjectures • Infer from data and predict • Determine the truth of a mathematical pattern or proposition

Note. From "Measuring the Content of Instruction: Uses in Research and Practice," by Andrew C. Porter, 2002, Educational Researcher, 31, p. 13.

Identifying Data Analysis and Statistics Content

Instrument

In his more recent publication Porter (2002), suggests that the content matrix can be used to go beyond describing content instruction in the classroom. He gives examples of the use of the instrument to measure alignment of assessments with content standards, alignment of instruction with assessments, and as a tool to create powerful graphical displays to describe the content emphasized (and not emphasized) in state standards. It is for the latter that the instrument will be put in this study.

Since the only area of interest for this research is the content in Data Analysis and Statistics in Middle grades, a modified context matrix was created, based on the list

suggested by Porter and Smithson (2001) for Data Analysis/Statistics topics in Middle School presented above. Porter and Smithson's list of topics is the product of the Reform Up Close study, a Consortium for Policy Research in Education project and follow up studies. Since the articulation of their latest approach is still in the beginning stages of development, a modified version of Porter's content matrix toward the organization of topics by "big ideas" was used. For example, Porter and Smithson (2001) list "bar graph, histogram" as a topic. For the modified matrix, a re-organization of the graphs and plots was made to reflect the big ideas in data analysis of graphical representation of categorical and numerical data. Graphs and plots are organized according to the type of data they describe. For example, graphs and plots that are used to organize categorical data are grouped together. Similarly, the graphs and plots to summarize numerical data and bivariate data appear as separate topics.   Porter & Smithson (2001) suggest measures of center but leave out measures of spread such as range and interquartile range. Therefore a new topic was added to take spread into account. The big idea of describing the shape of distributions was also added to account for other features that characterize distributions of data such as skewness, outliers, clusters, and gaps. Finally, since some of the standards encountered address the process and design of statistical investigation, a separate topic focusing on the formulation of questions, collection of data and design of studies was added. Following is the modified content matrix used in this research.

As for the list of categories of the cognitive demand dimension, the ones suggested by Porter (2002), with the proper interpretation, can be seen as a more refined version of a framework suggested by statistical educators (Garfield, 2002 and delMas, 2002). Garfield and delMas suggest three categories of learning outcomes: Statistical Literacy, Statistical Reasoning, and Statistical Thinking. Literacy refers to recognition or

computation; reasoning refers to explaining why and how a specific process works; and thinking refers to applying statistics to a context, critiquing, generalizing.   A link can be made between these three learning outcomes and Porter's five categories of cognitive demand. For example, the description of Porter's category of Memorize and Perform routine procedures is similar to Garfield and delMas' Statistical Literacy category which is mainly associated with creating graphs or plots, finding measures of center and spread. Communicate understanding is similar to the Statistical Reasoning category associated with the appropriate use and selection of graphs and measures.   For the purpose of achieving a finer analysis of the content of state and national standards and assessment, all five categories from Porter's (2000) work were used in this section. In later sections where it was decided that coarser grid is more appropriate, however, only the three levels suggested by Garfield and delMas are used.

Table 3.2   Data analysis and Statistics Content Matrix

Levels of Cognitive Demand	Big Ideas and corresponding Topics						
	Formulation of questions, designing studies, collect data	Categorical data representation	Numerical data representation	Bivariate data representation	Shapes of Distributions	Measures of center	Measures of spread
		Bar/pie graphs, pictographs, tables	Stem-and-leaf plots, histograms, box plots	Scatter plots, line graphs, regression line	Shapes of data distributions, skweness, gaps, outliers, clusters	Median, Mean, Mode	Range, IQR, Standard Deviation
A Memorize							
B Perform procedures							
C Communicate understanding							
D Solve nonroutine problems							
E Conjecture, generalize, prove							

Data Analysis and Probability
Standard 1:
The student understands and uses the tools of data analysis for managing information. (MA.E.1.3)
1. collects, organizes, and displays data in a variety of forms, including tables, line graphs, charts, bar graphs, to determine how different ways of presenting data can lead to different interpretations.
2. understands and applies the concepts of range and central tendency (mean, median, and mode).
3. analyzes real-world data by applying appropriate formulas for measures of central tendency and organizing data in a quality display, using appropriate technology, including calculators and computers.

Figure 3.1. Florida's Standard on Data Analysis and Probability for Grades 6-8.

Content is then described at the intersection between topics and cognitive

demands, based on data gathered from content standards and assessment.

Content Analysis

The documents analyzed at the student level were content standards for middle

grades students from ten states, Mathematical and Problem-Solving Goals of the

Connected Mathematics (Lappan, Fey, Fitzgerald, Friel, & Phillips, 2002) Grade 6 and 8

textbooks and the Data Analysis standards for Grade 6-8 (NCTM, 2000). The states

considered were those that require a middle school certification, and they are:

Connecticut, Florida, Georgia, Kentucky, Missouri, North Carolina, Ohio, Oregon,

Virginia, and West Virginia. For the complete set of standards see Appendix B.

Content standards vary in their structure and level of specificity from state to state.

Some states present their standards in a general statement followed by a set of more

specific standards. For example, Standard 1 from the state of Florida is shown in Figure

3.1(www.firn.edu/doe/curric/prek12/frame2.htm). Others do not present any general

statement and only list specific standards. For example, see Figure 3.2 from the state of

Virginia (www.pen.k12.va.us).

Probability and Statistics

6.18   The student, given a problem situation, will collect, analyze, display, and interpret data in a variety of graphical methods, including line, bar, and circle graphs and stem-and-leaf and box-and-whisker plots. Circle graphs will be limited to halves, fourths, and eighths.

6.19   The student will describe the mean, median, and mode as measures of central tendency and determine their meaning for a set of data.

6.20   The student will determine and interpret the probability of an event occurring from a given sample space.

Figure 3.2. Three Virginia standards on Probability and Statistics for Grade 6.

Because standards needed to receive a code or codes, it was decided to make the

unit of analysis the specific standard. In the case that a general standard is presented, like

in the case of Florida, it was used to help identify the intent of topics and cognitive

demands of the specific standards to follow. Each unit of analysis received a code or

codes according to the topics and cognitive demands observed. Codes consisted of a digit

from 1 to 7 corresponding to the seven topics (see Table 3.2) and a letter from A to E

corresponding to each cognitive demand.

Mapping the Standards to Cells on the Matrix

After becoming familiar with the instrument – the content matrix- and the units of

analysis – specific standards, the first thing to determined was which topics on the content

matrix were the best match for the specific standard. Identifying the topics in the matrix

that were appropriate was done parsimoniously. The first principle was, "stick to the

language of the standard". Coders were instructed not to overgeneralize or to interpret

intended topics that were not noted in the standard. Second, if the standard applied to more than one topic, all of those topics were coded. For example, consider the following standard "select, create, and use appropriate graphical representation of data, including histograms, box plots, and scatterplots". This refers to two topics of the matrix, "Numerical data representation" and "Bivariate data representation", but not the topic "Categorical data representation".

Similarly, matching the cognitive demand(s) on the content matrix to the already chosen topic was also done parsimoniously.   Here too, the principle "stick to the language of the standard" was applied. For example, if the standard was "describe the shape of the data using range, outliers, and measures of center, including the mean, median, and mode", the correct cognitive demand would be "communicating understanding". The raters did not assume that in order to "describe the shape of the data" the student would also be able to find or select measures of spread and center ("performing routine problems") or to formulate conjectures or infer beyond the data. In the case where the standard did not specify whether the student should be able to solve routine or non-routine problems and it was not possible to determine whether the problem was routine or non-routine from the context, then both cognitive demands were selected.

For example, the Florida standard "The student collects, organizes, and displays data in a variety of forms, including tables, line graphs, charts, bar graphs, to determine how different ways of presenting data can lead to different interpretations. " received all of the following codes: 1B, 2B, 2C, 4B, and 4C. Although the standard does not mention some of the graphical representations listed in topic 2, like pictographs and some listed in topic 4, like scatter plots, there is at least a representation of categorical and bivariate displays listed in topic 2 and topic 4. In terms of the cognitive demand, the "collects,

64

organizes, and displays data" corresponds to category B: "perform routine procedures"

and the last part about different interpretations corresponds to the category C:

"communicating understanding".

A spreadsheet for each state was created to record the codes. Columns were

created for each specific standard. For example, Figure 3.3 shows three out of the four

specific standards for the state of Florida, which correspond to the specific standards

shown in Figure 3.1. They are labeled FL1, FL2, and FL3 and form the columns of the

spreadsheet with the corresponding codes assigned. Note how each specific standard

received more than one code.

	A	B	C	D	E	F	G	
1		State:		Florida				
2		Grade:		Grades 6-8				
3								
4		Standard:		Standard:		Standard:		
5		FL1		FL2		FL3		
6		Topic	Demand	Topic	Demand	Topic	Demand	Tc
7	1	1	B	6	C	2	D	
8	2	2	B	7	C	4	D	
9	3	4	B			6	D	
10	4	2	C					
11	5	4	C					
12	6							

FL.xls
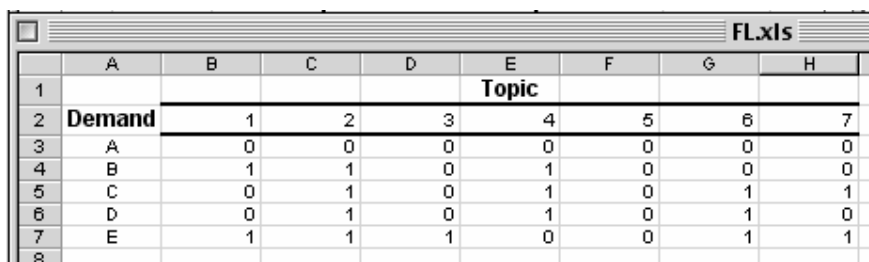
Figure 3.3. Excel Spread Sheet to Record Codes for Three Specific Florida Standards.

To assure reliability of the coding process, one rater was trained to code the

specific standards. The trained rater and the researcher coded the specific standards

independently. The researcher adjudicated when there was disagreement between them.

A total of 67 specific student standards were coded and since each specific standard could

receive one or more codes, there were a total of 171 codes recorded.

65

The next step in the analysis was to create a visual representation of the data

analysis from each document reviewed. Content maps were used to provide a

representation of the content in a given set of standards using a surface area chart which

results in a graphic similar to contour or topographical maps, except that these graphs

better suite the categorical nature of the data.

The graphs were created using a simple Mathematica software code. To

demonstrate how this was done consider again the example of Florida standards. First, a

frequency of the codes for all standards was recorded in a two-dimensional matrix (see

Figure 3.4), the columns correspond to the seven topics and the rows to the five cognitive

demands. These frequency values are what would constitute the measurement cell of the

content graph. In each cell of the matrix, the number of times a code was observed was

recorded.   That is, the code A1 was observed zero times, the code B1 was observed one

time, and so on. Note that the sum of the values of each cell should correspond to the

total number of codes assigned to the standards of Florida, 15 codes in this case. Also

note that the highest frequency of codes is one, which corresponds to about 6.7%.

						FL.xls		
	A	B	C	D	E	F	G	H
1					Topic			
2	Demand	1	2	3	4	5	6	7
3	A	0	0	0	0	0	0	0
4	B	1	1	0	1	0	0	0
5	C	0	1	0	1	0	1	1
6	D	0	1	0	1	0	1	0
7	E	1	1	1	0	0	1	1
8								

Figure 3.4. Frequency Matrix for Codes Observed in Florida Standards.

Then the matrix was copied into a Mathematica file and defined as "fl". A small

function defined with the specific arguments created the desired colored graph:

```
fl = {{0, 0, 0, 0, 0, 0, 0}, {1, 1, 0, 1, 0, 0, 0}, {0, 1, 0, 1, 0, 1, 1},{0, 1, 0, 1, 0, 1,
0},{1, 1, 1, 0, 0, 1, 1}}
graph[ mat_]:= Module[{},
        rects = Table[Rectangle[{i,j}, {i+1, j+1}] , {i, 0, 6}, {j, 0, 4}];
        Map[Graphics, Flatten[MapThread[{Hue[colorfun[#1], brightfun[#1],1],
        #2}&, {mat, rects} , 2],1 ]]]
```

The grid of colored rectangles in the graph identifies the seven topics (indicated

by columns) and five cognitive demands (indicated by rows). The intersection of each

topic and category of cognitive demand represents a measurement cell and corresponds to

a particular cell of the content matrix. The image of the map is simply a computer-

generated graphic from Mathematica based on the frequency values for each cell. The

color of a cell indicates the percentage of the content observed by the raters for a given

topic and category of cognitive demand. The darker the cell the more frequently the

content (code) was observed across the standards. The resulting content map for the

Florida standards is shown in Figure 3.5.
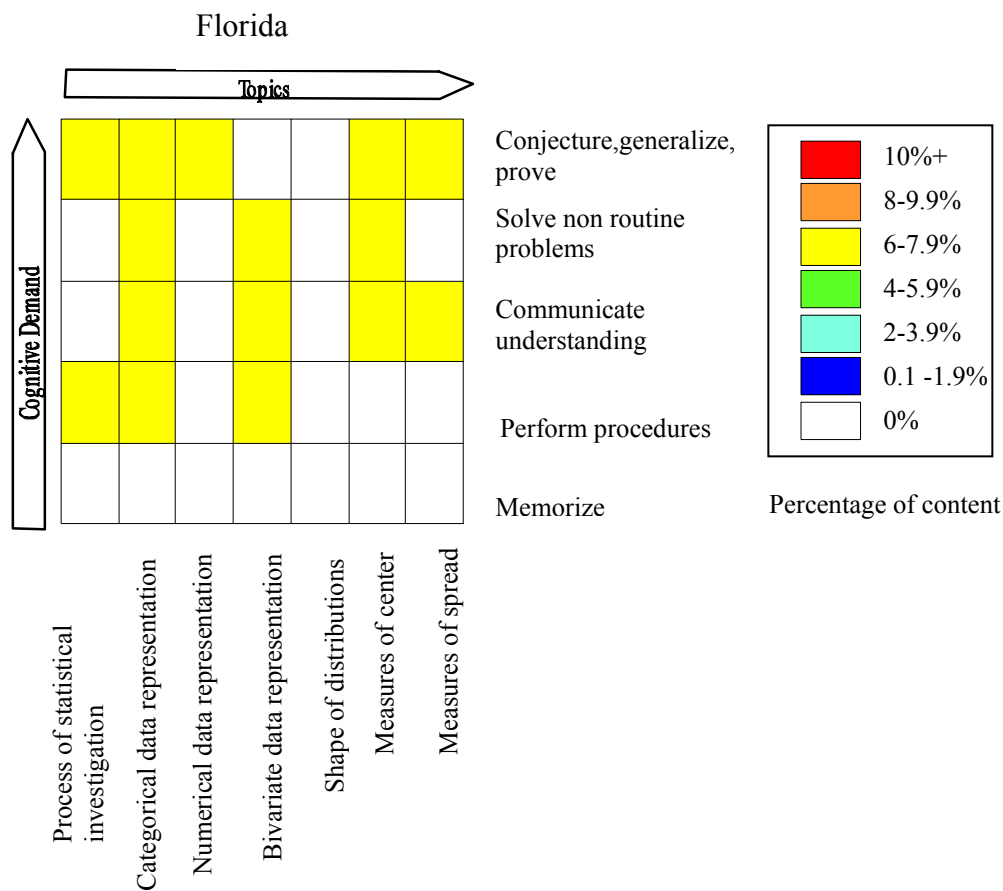
Florida

Figure 3.5. Content Map for Florida Standards Grades 6-8.

Content matrices and maps were created for each of the ten states, the NCTM (2000) standards and the CMP Mathematical and Problem-Solving Goals. To allow for easy comparison, the same color scheme and scale was used for all maps.   The scale, shown in the legend of each content map, was chosen to highlight the important features of the map for each document as well as the map summarizing all of the documents reviewed. Note that for the majority of documents the frequency observed for each cell observed is either 1 or 0. Hence the two colors represent the presence or absence of that

content in the document. Table 3.3 gives the total number of specific standards coded,

total number of codes assigned, highest frequency of codes observed and its

corresponding percentage for each set of state standards, the NCTM standards and the

CMP Mathematical and Problem-Solving Goals.

Table 3.3

Number of standards analyzed and codes assigned

State	Number of specific standards	Number of codes assigned	Highest frequency of codes observed	Highest percent of content
Connecticut	4	10	1	$1/10=10\%$
Florida	4	15	1	$1/15\approx7\%$
Georgia	7	17	1	$1/17\approx6\%$
Kentucky	9	21	2	$2/21\approx10\%$
Missouri	4	17	2	$2/17\approx12\%$
North Carolina	6	10	1	$1/10=10\%$
Ohio	5	11	1	$1/11\approx9\%$
Oregon	8	15	1	$1/15\approx7\%$
Virginia	7	18	1	$1/18\approx6\%$
West Virginia	6	16	1	$1/16\approx6\%$
NCTM	7	21	2	$2/21\approx10\%$
CMP	9	36	3	$3/36\approx8\%$
Total	76	207	12	$12/207\approx6\%$

Analysis of Content Maps

Figure 3.6 thru Figure 3.10 show content maps for each set of state standards and

Figure 3.11 shows the content map for *Principles and Standards for School Mathematics*

(NCTM, 2000). State standards vary considerably on the topics, cognitive demands and

the emphasis given to those. For example, the states of Connecticut, Florida (see Figure

3.6), and West Virginia (see Figure 3.10) all show emphasis on the topics of measures of

center and categorical data representation, but Florida has these topics at the level of

solve non-routine problems while Connecticut and West Virginia have them at the level

of perform procedures.

Virginia (see Figure 3.10) has one of the most comprehensive content maps. It covers all the topics but one, shapes of distributions; and it is pretty uniform on the emphasis of level of cognitive demand. The content map for Oregon (Figure 3.9) does not show clusters around any particular topic; instead it puts emphasis on all but categorical representation of data. The map for Georgia (Figure 3.7) covers all topics on data representation at pretty much all levels of cognitive demand, but leaves out the concept of shape of distribution and includes finding and using measures of center and spread.

Figure 3.11 shows that the *Principles and Standards for School Mathematics* (NCTM, 2000) includes all topics except categorical data analysis (which is included in the standards for Grades 3 – 5) and the explicit language of shapes of distributions; and includes all levels of cognitive demand. Nevertheless, these standards also show higher percent of content on the interpretation of numerical and bivariate data representation and on the top level of cognitive demand: conjecture, generalize and prove.

The content described in the *Connected Mathematics* Teacher's Guide (Lappan et al, 2002) is also different from eleven sets of standards for students reviewed earlier. The most emphasis is given to the process of statistical investigation and to inference about that data using shapes of distributions, measures of center and spread. This is a reflection on the emphasis of studying statistics as an exploration process of investigation. These textbooks include topics related to the design of studies (survey), random samples, and the comparison of sampling distributions to make conclusions about the population.

There is only one commonality among all 12 sets of standards examined. None covers any topic at the level of memorization.

Connecticut

Topics

Cognitive Demand

Conjecture, generalize, prove

Solve non routine problems

Communicate understanding

Perform procedures

Memorize

	10%+
	8-9.9%
	6-7.9%
	4-5.9%
	2-3.9%
	0.1 -1.9%
	0%

Percentage of content

Florida

Topics

Cognitive Demand

Conjecture, generalize, prove

Solve non routine problems

Communicate understanding

Perform procedures

Memorize

	10%+
	8-9.9%
	6-7.9%
	4-5.9%
	2-3.9%
	0.1 -1.9%
	0%

Percentage of content

Process of statistical investigation
Categorical data representation
Numerical data representation
Bivariate data representation
Shape of distributions
Measures of center
Measures of spread

Figure 3.6. Content Maps for the Connecticut and Florida Standards.

71
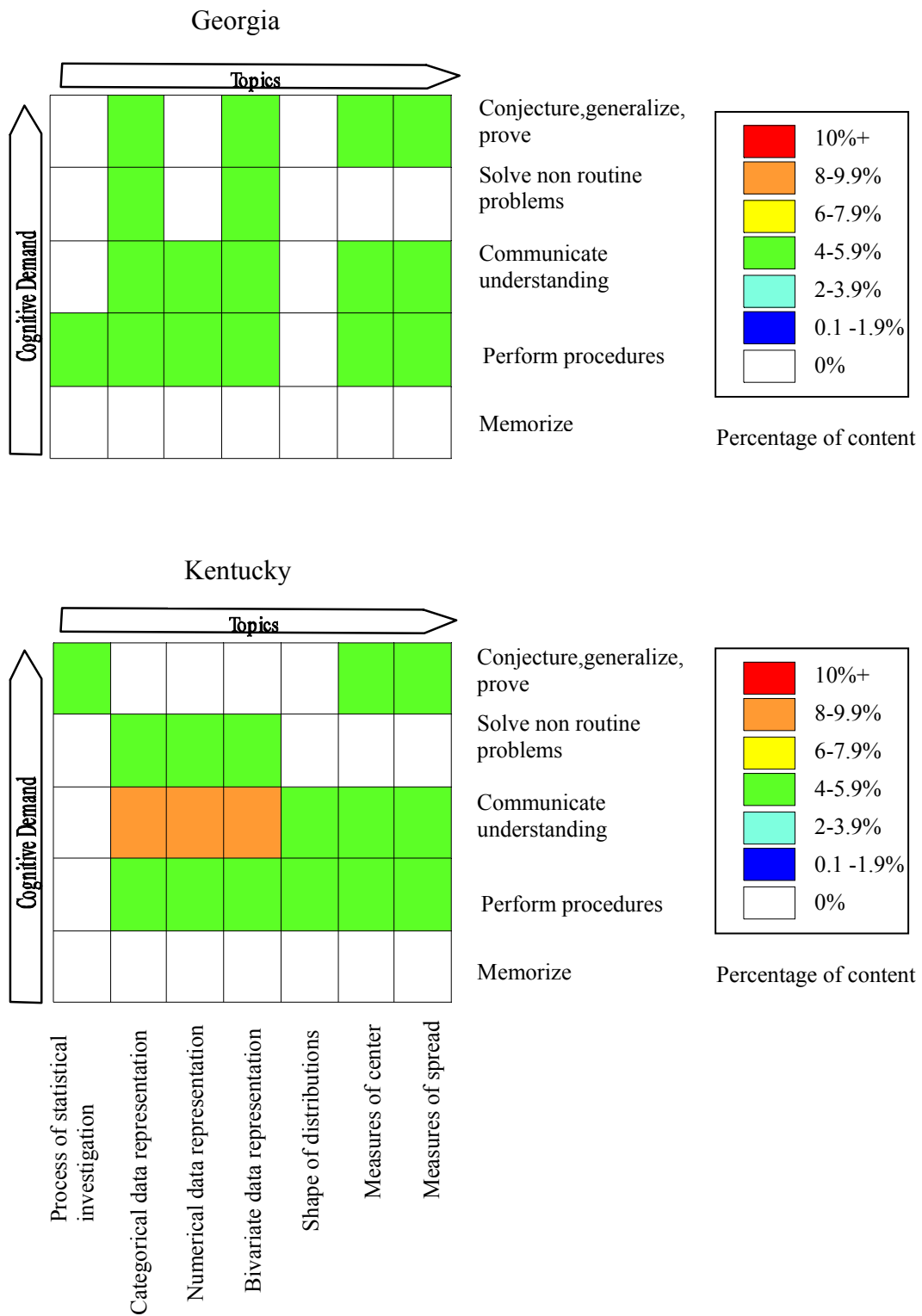
Georgia

Kentucky

Figure 3.7. Content Maps for the Georgia and Kentucky Standards.

Missouri
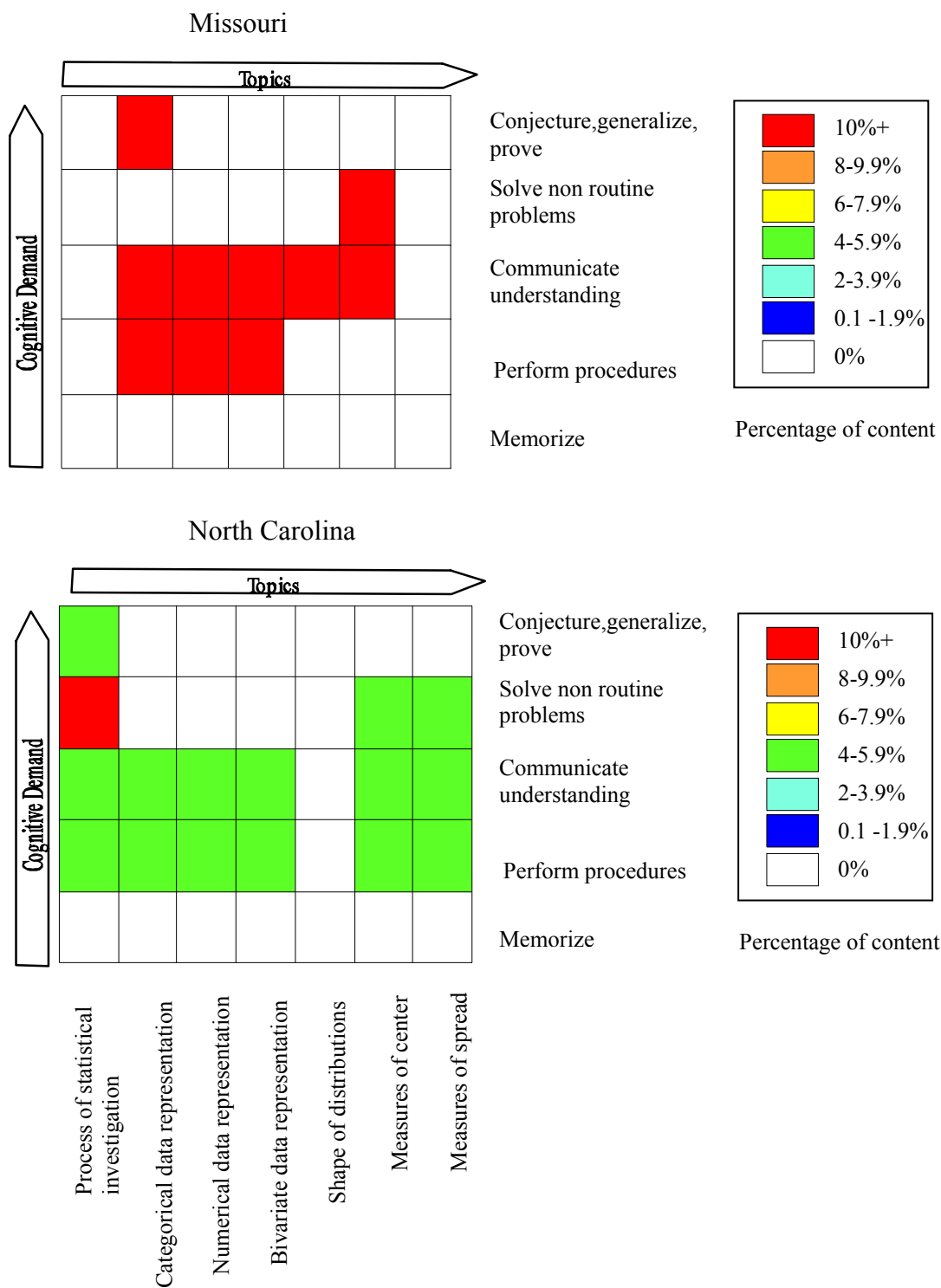
North Carolina

Figure 3.8. Content Maps for the Missouri and North Carolina Standards.
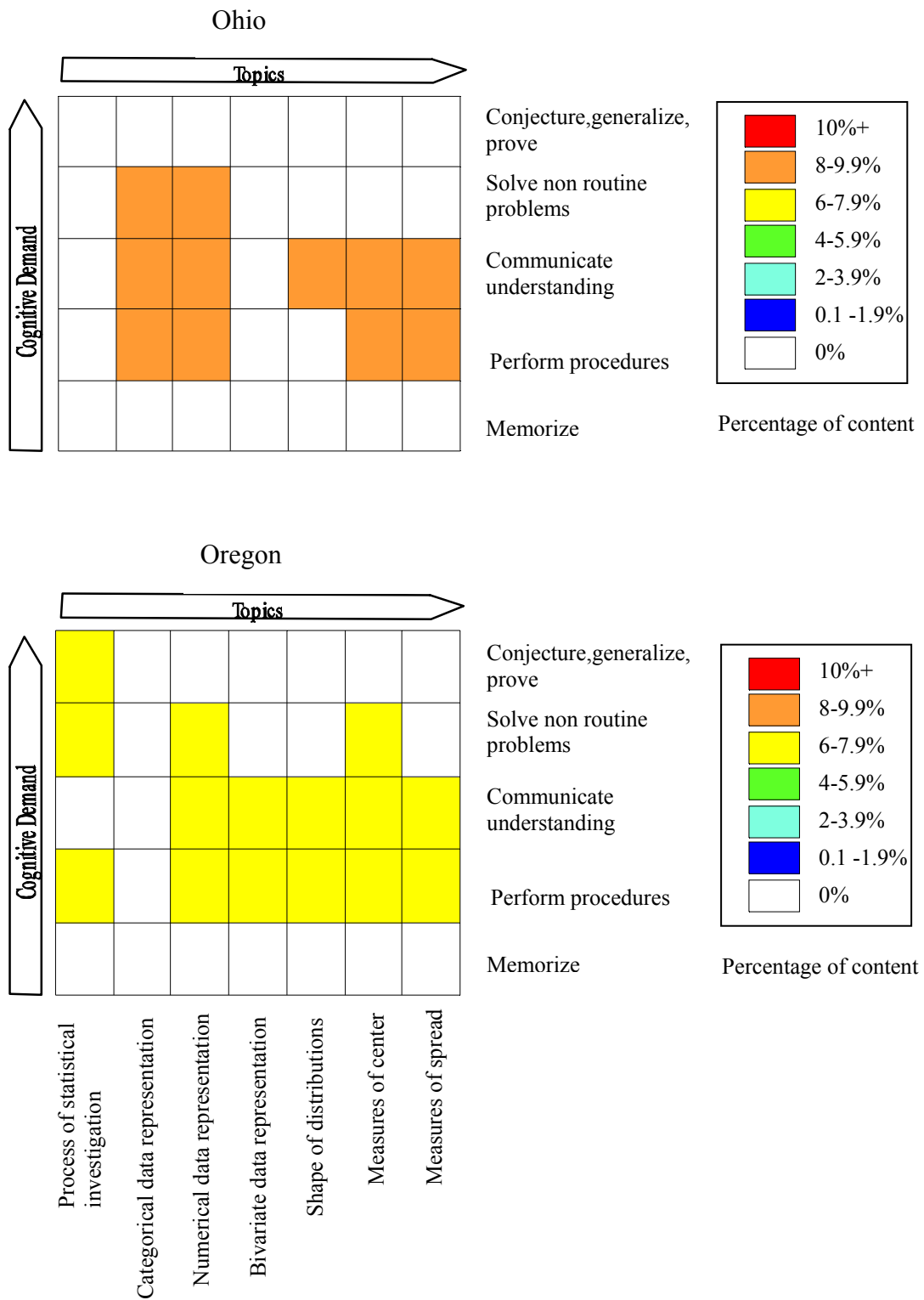
Figure 3.9. Content Maps for Ohio and Oregon Standards.
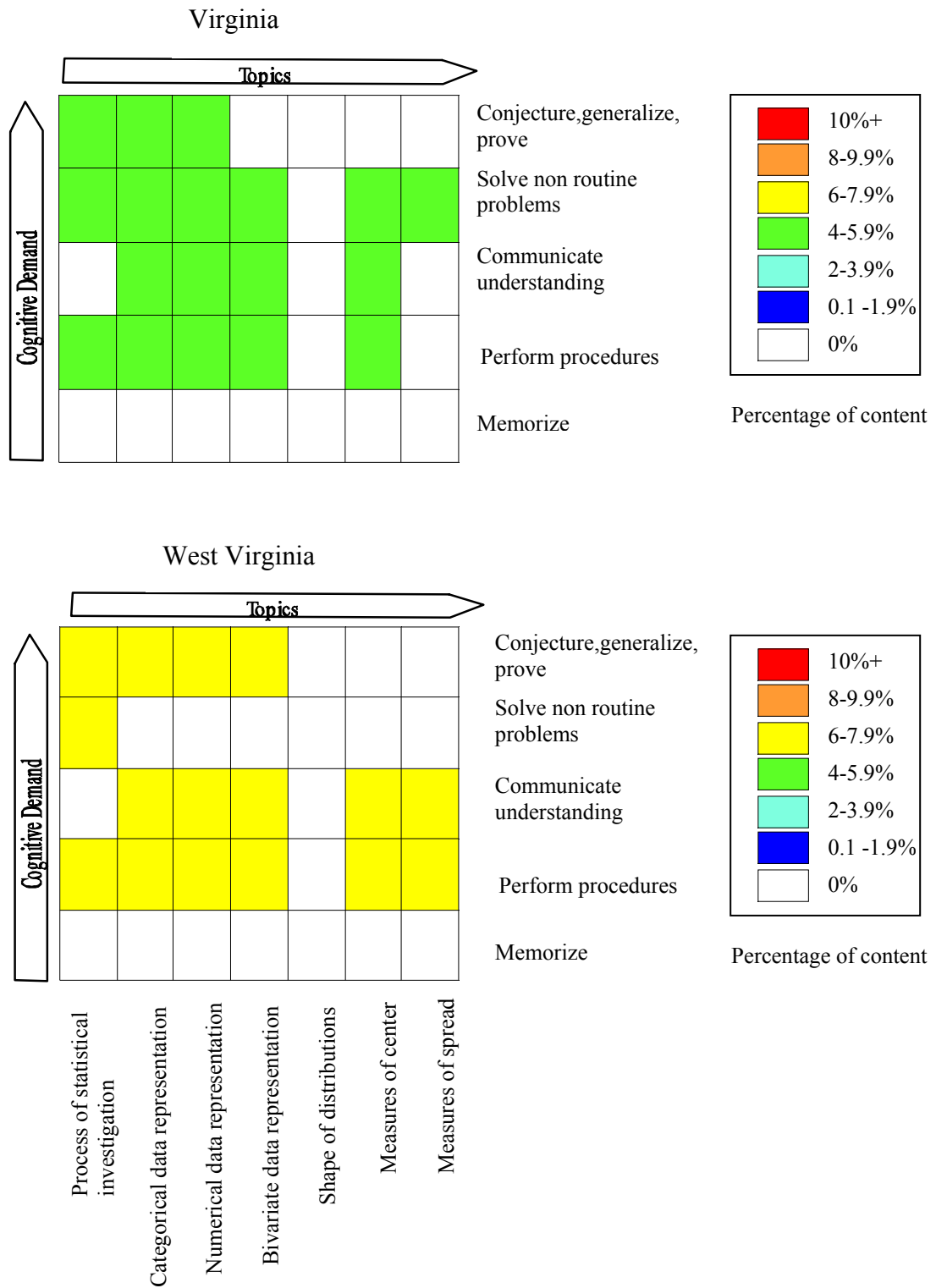
Figure 3.10. Content Maps for Virginia and West Virginia Standards.
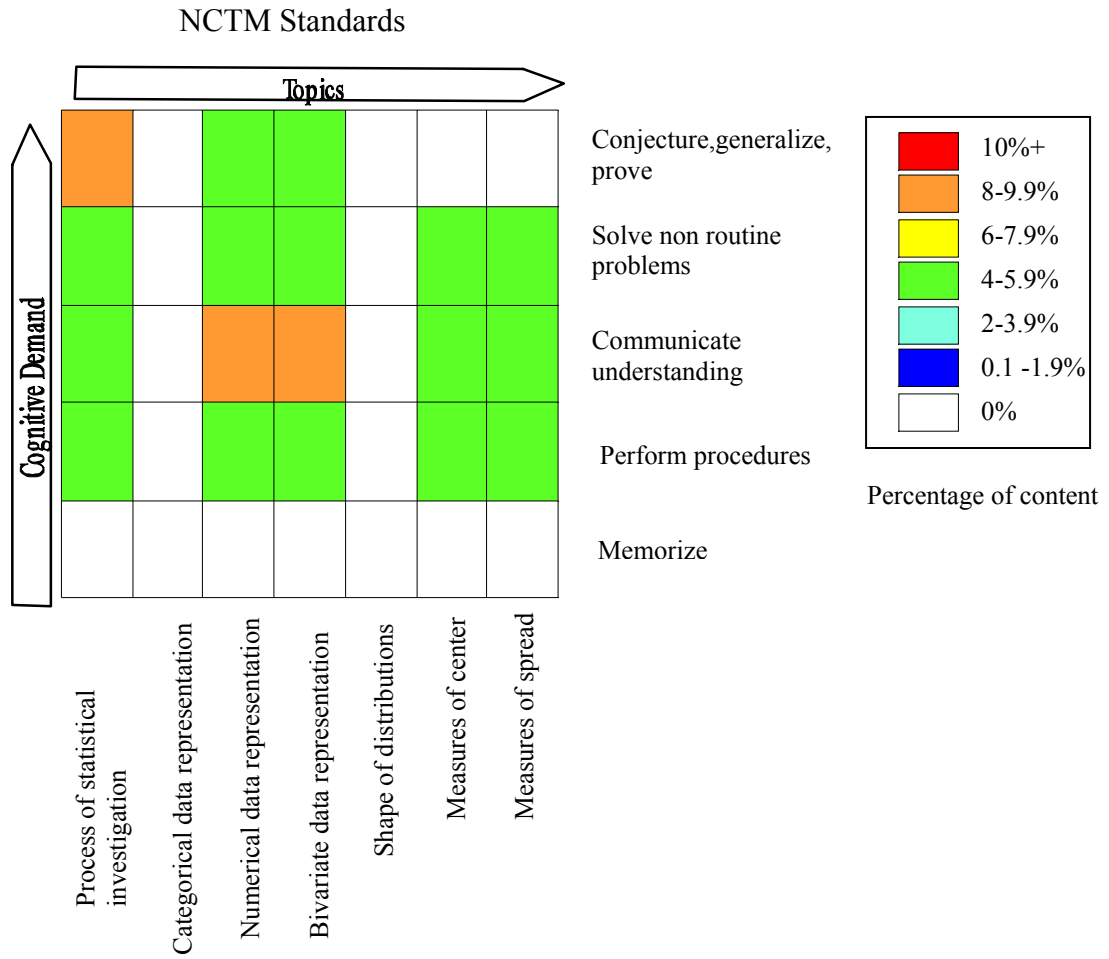
NCTM Standards

Figure 3.11. Content maps for *Principles and Standards for School Mathematics* (NCTM, 2000)
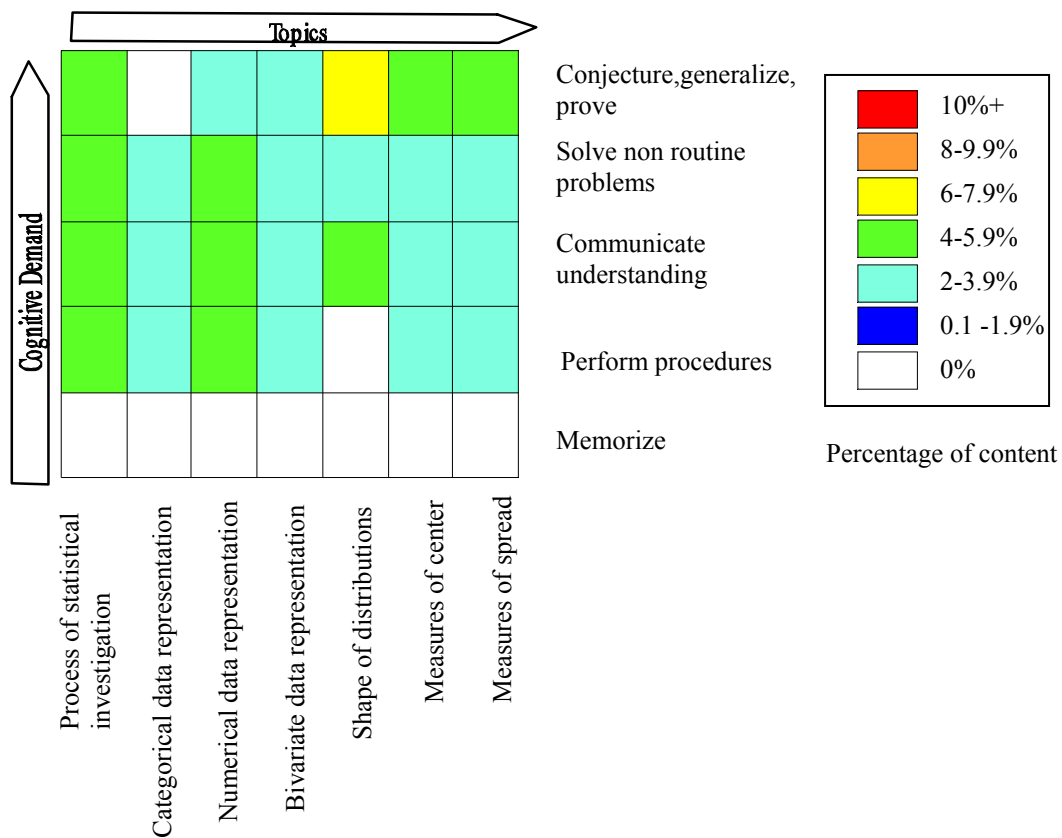
Figure 3.12. Content Map for the Mathematical and Problem-Solving Goals in *Connected Mathematics* Teacher's Guide Grade 6 and 8 Textbooks.

When all matrices are put together the content map of Figure 3.13 is created. This map allows one to "see" this set of 12 documents as a whole and identify the statistical content that students in middle grades are expected to know. The map indicates that the least covered topics are shapes of distribution and the process of statistical investigation. Instead, emphasis is placed on representation of data, particularly numerical representation and measure of center and spread at the level of communicate understanding. That is, 5.8% or more of the content in middle grades described in these documents is dedicated to communicating understanding with graphical displays of

numerical data. Whereas between 5.1% and 5.8% of the content is dedicated to

communicating understanding with measures of center. In general, the main focus of the

standards is at this middle cognitive demand. Only for the topic of process of statistical

education is the highest level of conjecture, generalize and prove emphasized.
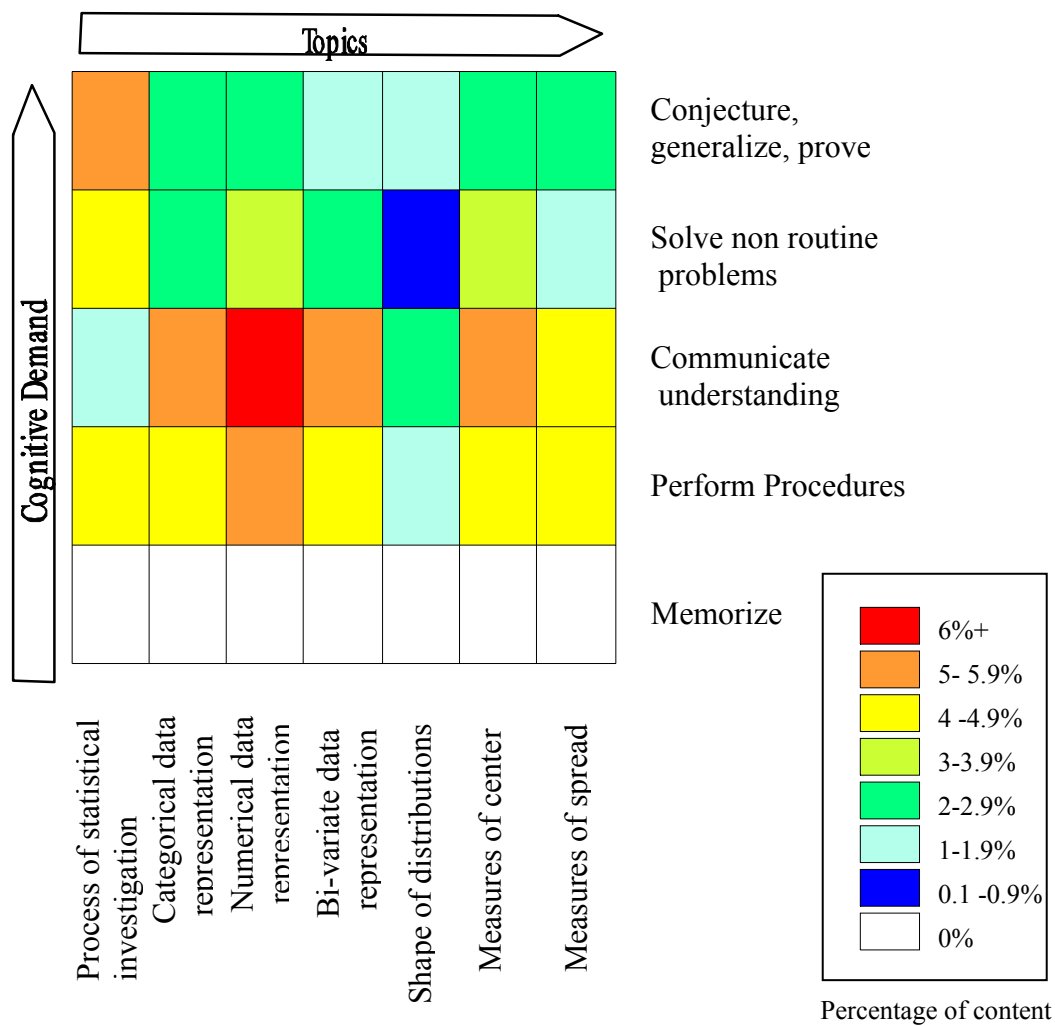
Figure 3.13. Content map for the ten states , the Mathematical and Problem-Solving
Goals in *Connected Mathematics* Teacher's Guide Grade 6 and 8 Textbooks and
Principles and Standards for School Mathematics (NCTM, 2000)

Data Analysis and Statistics for Middle Grades Teachers

We now focus our attention on the content the teachers, as opposed to students, are expected to know in the area of data analysis and statistics. For teachers, content standards are not as developed as they are at the student level. Only four of the ten states (Florida, Georgia, Missouri, and North Carolina) analyzed had professional standards for teachers which were available at the time this research was conducted. Furthermore the standards for these states are very broad and not content specific. For example, North Carolina's content pedagogy standard states in part "The teacher understands the central concepts, tools of inquiry, and the structures of the discipline …". However, for teachers to get accreditation to teach mathematics at the middle grades level in the ten states under consideration, they need to pass a standardized content test called PRAXIS II: Middle School Mathematics (0069). The content covered in this examination, sets the base of what teachers need to know. The actual test was not available for analysis, but the Educational Testing Service (ETS) publishes an on-line booklet called *Tests at a Glance* (ftp://ftp.ets.org/pub/tandl/0069.pdf). *Tests at a Glance* includes content outlines, sample questions in each content area with a rationale for the best answers, and test-taking strategies. The topics covered are organized by content area and the description includes level of cognitive demand required. The content standards on the Data, Probability and Statistical Concepts in the on-line booklet for test 0069 were analyzed for this research.

Recently, the Conference Board of the Mathematical Science (CBMS) published a report called *The Mathematical Education of Teachers* (CBMS, 2001), which was designed to be a resource for those involved in the education of mathematics teachers. The recommendations in this report are organized by elementary, middle, and high school level and by areas of content. The second set of standards for teachers analyzed for this

research were the specific statements in the summary of data analysis and statistics

content (page 113) under the section "The Mathematical Content Needed by Prospective

Teachers".

Content Analysis and Maps

Content analysis for these documents was preformed similarly to the analysis for

students' documents. The same content matrix (Table 3.2) and language was used even

though some of the documents included more advanced topics[1].   Table 3.4 shows the

number of specific units of analysis examined and the number of codes assigned to each

document. The table also shows the highest frequency of codes observed and the

corresponding highest percent of content, which correspond to the highest level of

shading on the content maps.

Table 3.4

Number of specific statement analyzed and codes assigned in three sets of standards for

teachers.

Document	Number of specific statements	Number of codes assigned	Highest frequency of codes observed	Highest percent of content
Topics covered in PRAXIS II Middle School Mathematics	3	13	1	7.7%
The Mathematical Education of Teachers (CBMS, 2001)	8	14	1	7.1%
Total	11	27	2	7.4%

Figure 3.14 shows the content maps for the PRAXIS II test and *The Mathematical*

Education of Teachers (CBMS, 2001). These two documents show two different patterns

---

[1] In order to facilitate comparisons between the standards for teachers and students some advanced topics
were excluded from this analysis. Advanced topics which were considered well beyond the middle grades

corresponding to their nature; one is a multiple choice assessment and the other is a report on recommendations. The PRAXIS II content pattern is quite different from the content suggested for students in Figure 3.13. The most striking difference is for the topic numerical data representation. For the PRAXIS II the cognitive demand of perform procedures and solve nonroutine problems are emphasized over communicate understanding while the reverse is true for the student level content map in Figure 3.13. In contrast, *The Mathematical Education of Teachers* (CBMS, 2001) shows its highest percent of content across all topics, most of them at the middle level of cognitive demand. Note that the emphasis for data representation is not on performing procedures (i.e. construction of graphs) but rather on communicating understanding (interpreting graphs). In addition to the topics described in the content matrix, the MET suggests that prospective middle school teachers should know how to design simple investigations including random sampling or random assignment to treatments; sampling distributions, margin of error, confidence intervals, and expected values.

As was done with the content for students, the two content matrices were put together to form a single content map (Figure 3.15) to identify the commonality between the two documents. The map shows emphasis on categorical data representation, bivariate data representation and measures of center and spread.

---

level and were not coded included expected values, random assignment to treatments, sampling distribution and procedures of formal statistical inference such as confidence intervals and hypothesis testing.

Topics Covered in PRAXIS II
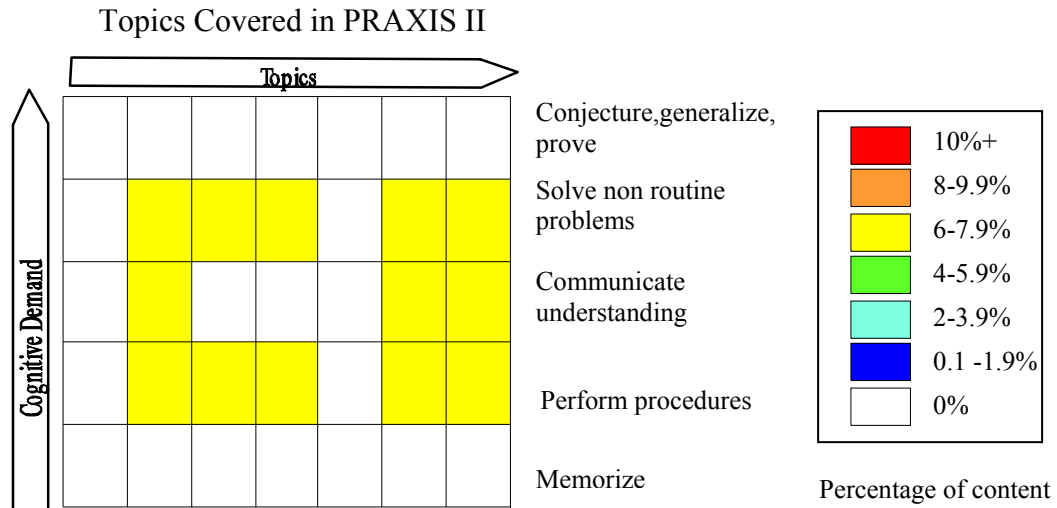
The Mathematical Education of Teachers (CBMS, 2001)
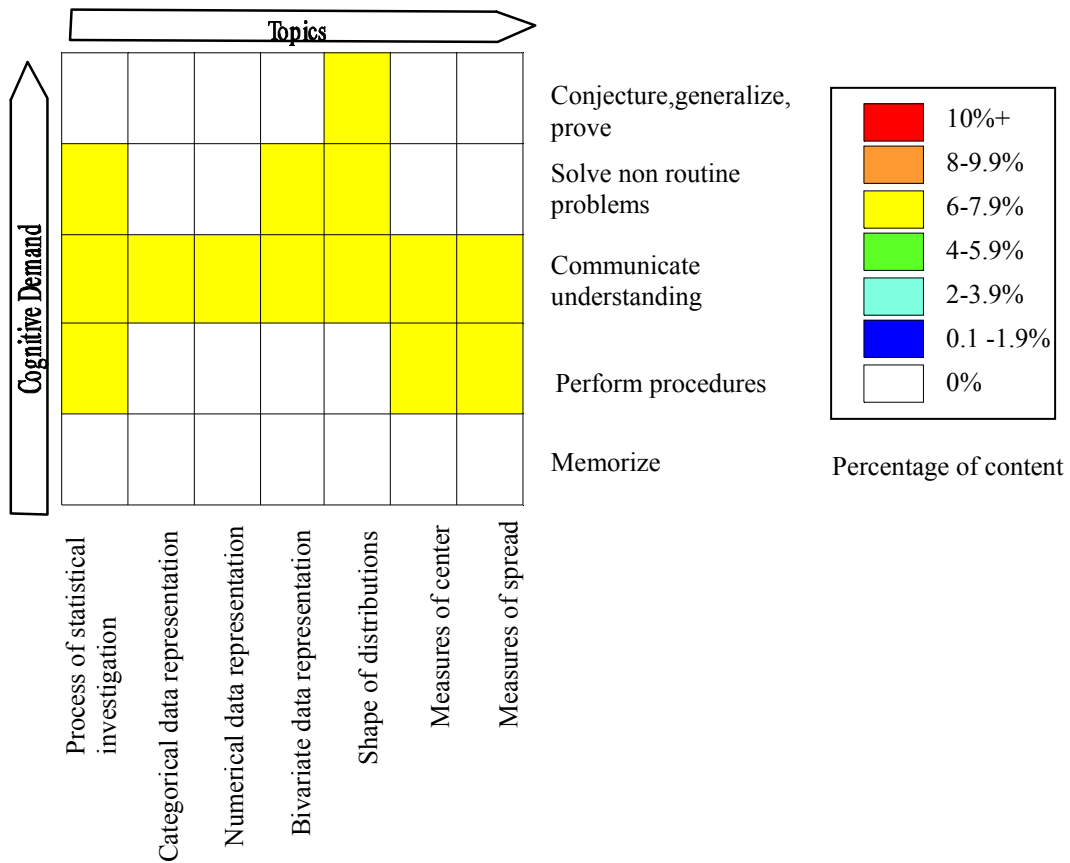
Figure 3.14. Content Maps for *The Mathematical Education of Teachers* (CBMS, 2001) and the Topics Covered in PRAXIS II Middle School Mathematics Test.

Topics

Cognitive Demand

Conjecture, generalize, prove

Solve non routine problems

Communicate understanding

Perform procedures

Memorize

Process of statistical investigation

Categorical data representation

Numerical data representation

Bivariate data representation

Shape of distributions

Measures of center

Measures of spread

	10%+
	8-9.9%
	6-7.9%
	4-5.9%
	2-3.9%
	0.1 -1.9%
	0%

Percentage of content
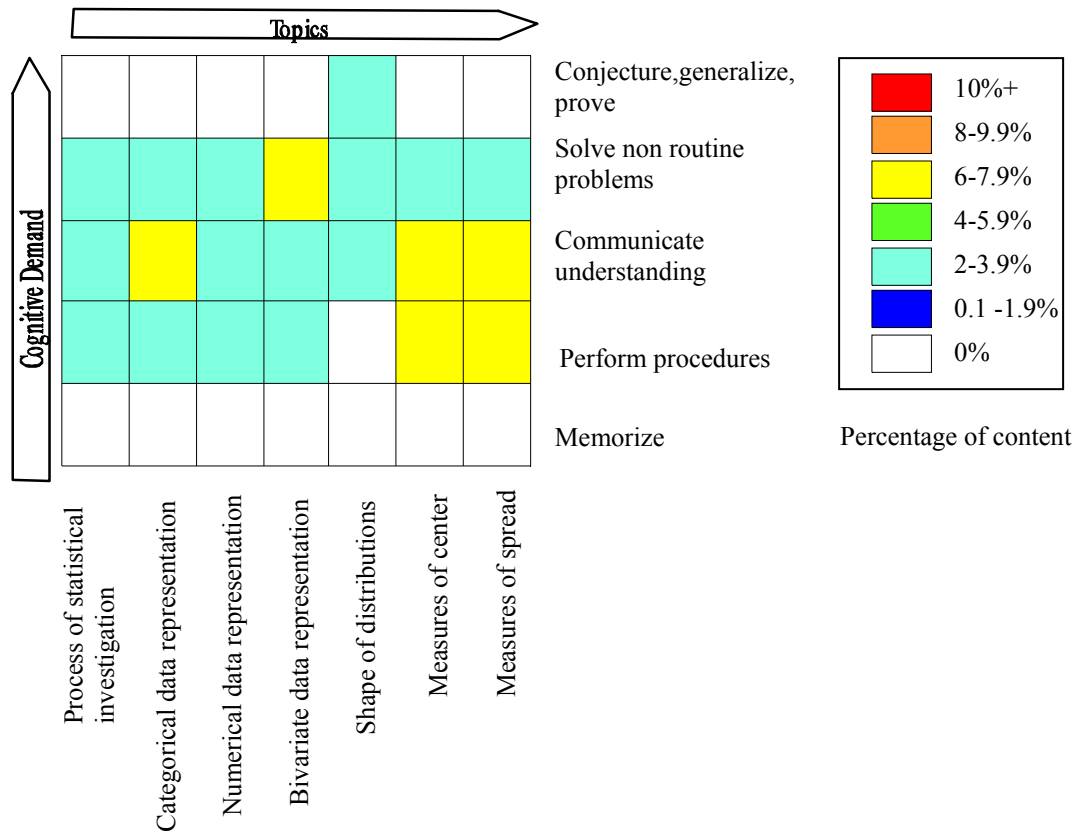
Figure 3.15. Combined Content Map for *The Mathematical Education of Teachers* (CBMS, 2001) and the Topics Covered in PRAXIS II Middle School Mathematics Test.

Relationship between Content for Students and Teachers

As we compare the two sets of documents, one at the student level and the other at the teacher level, we observe some differences in coverage and emphasis. On the one hand, Figure 3.12 shows clearly two clusters around communicating understanding of numerical data representation and measures of center. On the other hand, the map of content for teachers (Figure 3.15) shows clusters around communicating understanding of categorical data representation and measures of center and spread, as well as computing measures of center and spread and using bivariate data representations to solve nonroutine problems. However, it is important to recall that the teacher level map in Figure 3.15 is based only on two documents, while that student level map is based on 12 documents. Thus caution should be taken not to overgeneralize these findings.

While the content matrix and its corresponding map are useful to visualize the breadth and emphasis of coverage of the various standards analyzed, it is also desirable to have a list of specific tasks students and teachers are expected to perform. This is especially true if the ultimate goal is the creation of an instrument to assess if teachers in fact possess the knowledge demanded of them. To this end, Table 3.5 includes the most important tasks found in the documents reviewed in this study. The columns represent levels of cognitive demand using the framework developed by Garfield (2002) and delMas (2002).

Table 3.5

Summary of Tasks of Content Knowledge for Data Analysis and Statistics

	Statistical Literacy	Statistical Reasoning	Statistical Thinking
Knowledge of Data Analysis and Statistics	• Identify categorical and numerical data. • Create and read information presented in data displays. • Find and compute mean, median and mode. • Find and compute range. • Identify clusters, gaps, outliers, symmetry, modality, and skewness.	• Formulate questions that can be addressed through data collection. • Understand what constitutes a random sample. • Understand how surveys are undertaken and how experiments are designed. • Interpret and integrate information presented in data displays • Interpret what measures of center and spread tell about the data. • Identify misuse of cause-and-effect interpretations of correlations.	• Make decisions on what and how to measure. • Extend, predict, or infer from information presented in data displays to answer implicit questions. • Use measure of center to make predictions and inferences from data about the group to which the data pertains. • Use the spread and shape of a data set to make judgments about the accuracy and reliability of the data and make inferences from data.

Statistical Content Knowledge Applied to Teaching

Identifying the content knowledge that is needed to teach data analysis and statistics in middle grades is a much more difficult task than identifying the "pure" content knowledge for several reasons. One, authors of theoretical and empirical literature of the past and present propose divergent elements to be considered for such knowledge. Second, the development of a construct of the nature of teachers' knowledge and its organization is still in the works (www.soe.umich.edu) with clear evidence that the construct has multi-dimensionality characteristics. Third, the variability in and sometimes lack of complete structure of documents which suggest teachers' knowledge make systematic comparison difficult. Finally, as the literature and researchers point out (Ball, Lubienski, and Mewborn, 2001) this knowledge is better manifested or embedded in the actual practice of teaching. However, a comprehensive observational study of classroom practice or in-depth interviews with a large sample of in-service teachers is beyond the scope of this present study.

Nevertheless, in this section, an attempt is made to take a look at written documents related to teachers' knowledge in order to describe their similarities and contrasts and finally to summarize the important aspects. As a proxy for examining actual (or modeling ideal) classroom practice, the Teacher's guide of the *Connected Mathematics* (Lappan et al, 2002) data analysis and statistics unit "Data About Us" was also studied. The hope is that the discussion here will serve as an important step in the development of content area-based teaching standards on par with those that currently

exist for content knowledge for students, and perhaps more importantly a framework in which to base teacher preparation programs and assessment instruments.

Teachers' Knowledge as Suggested by Documents

For content knowledge applied to teaching, the data sources examined were *Knowing and Learning Mathematics for Teaching* (NRC, 2000), *Adding It Up* (NCR, 2001a), *Middle childhood through early adolescence/Mathematics Standards* (National Board for Professional Teaching Standards, 1998), *Professional Standards for the Accreditation of Schools, Colleges, and Departments of Education* (National Council for Accreditation of Teacher Education, 2002), National Middle School Association Middle Level Teacher Preparation Standards (http://www.nmsa.org), four state professional standards (Florida, Georgia, Missouri, and North Carolina), Teacher's Guide Grade 6 Statistics Unit: Data About Us (Lappan et al, 2002), and *The Mathematical Education of Teachers* (CBMS, 2001). An attempt to examine all ten state professional standards was made, however only four were accessible at the time this research was conducted

Classification of Documents by its Characteristics

Documents can be classified by their structure into three different categories based upon the level of specificity of the recommendations contained. In the first category, with the most general recommendations, are professional teaching standards. These documents are written to suggest knowledge for teachers in general terms without specifying any particular subject matter or, with the exception of the National Middle School Association Middle Level Preparation Standards (http://www.nmsa.org), even the grade level. Documents that fall into this category are the *Professional Standards for the*

Accreditation of Schools, Colleges, and Departments of Education (National Council for Accreditation of Teacher Education, 2002), National Middle School Association Middle Level Teacher Preparation Standards (http://www.nmsa.org), and the states' professional standards.

In the second category, the documents suggest teachers' knowledge for mathematics specifically, but do not focus on any content area within mathematics. Documents falling into this category are *Knowing and Learning Mathematics for Teaching* (NRC, 2001b), *Adding It Up* (NRC, 2001a), and the *Middle childhood through early adolescence/Mathematics Standards* (National Board for Professional Teaching Standards, 1998).

Finally, the third category is made of two documents: *The Mathematical Education of Teachers* (CBMS, 2001) which gives separate recommendations based on specific content area and grade level; and the *Connected Mathematics* Teacher's Guide Grade 6 Statistics Unit: Data About Us (Lappan et al, 2002) in which the application of statistical knowledge is implicitly suggested in the development of lessons. Note that these two documents were also analyzed for the statistical content needed for teachers in a previous section. In this section the focus is on the content for teaching and the analysis is more descriptive and in depth.

General Aspects of Teachers' Knowledge

General aspects refer to those aspects suggested for all teachers regardless of their area of specialization and grade level. Professional teaching standards, national and state, shows that these aspects are organized by domains of knowledge following a similar framework as that suggested by Lappan (2000). They all suggest that teachers and/or prospective teachers should have in-depth knowledge and understanding of the subject

matter they plan to teach, knowledge of pedagogy, knowledge of students as learners, and knowledge of assessment. Although the structure among all standards is similar, the language used in these documents is not at all uniform. For example, some standards prefer the language "knowledge of teaching practice" or "art of teaching" for the pedagogical component. Others like to use better "pedagogical content knowledge" as defined by Shulman (1986) and include the categories of content, pedagogy and knowledge of learners. Furthermore, a few include technology and planning as separate domains of knowledge.

The two documents related to national teacher standards: *Professional Standards for the Accreditation of Schools, Colleges, and Departments of Education* (National Council for Accreditation of Teacher Education, 2002) and the National Middle School Association Middle Level Teacher Preparation Standards (http://www.nmsa.org), as a whole, put the emphasis of the pedagogical or pedagogical content knowledge on the appropriate use and selection of teaching strategies. In terms of students as learners, the emphasis is on child developmental processes and ways of learning.

As for state standards, they all have different patterns in terms of the emphasis of the domains of knowledge. For example, professional standards from Florida refer more to knowledge about students' developmental learning and explanations but have no mention of assessment. In contrast, Georgia professional standards do not mention any topic related to knowledge of students as learners and the emphasis is on content and pedagogical content knowledge. Professional standards from the state of Missouri cover all domains of knowledge but very superficially. In particular, they mention very little about students as learners, student work or classroom discourse. Finally, North Carolina standards follow more the pattern of national documents providing coverage of all

domains with emphasis on student development, accessing student thinking though written work and classroom interaction.

These documents suggest one kind of general organization of teachers' knowledge by defining four domains of knowledge with specific aspects each: 1) Depth and breadth of subject matter knowledge teachers plan to teach which includes demonstration of understanding of central concepts; 2) knowledge of pedagogy which includes the appropriate use and selection of teaching strategies or methods, the ability to explain and present important concepts in different ways and contexts, and the use and selection of materials, and technology; 3) knowledge of students as learners which includes the use of students' prior knowledge, design instruction appropriate for social, cognitive, and emotional; access student thinking through classroom discourse and written work; and 4) knowledge of assessment, which includes selection, development, and use of informal and formal assessment strategies for the purpose of measuring achievement and to adjust instruction.

Although a framework is provided by the national and state teachers' standards and some aspects of knowledge for teaching are identified, they do not provide clear expectations in terms of cognitive demands for teachers or what this knowledge looks like in teaching practices. Hence, the guidelines they provide are too general to be useful for the development of assessment instruments or teacher preparation programs. A look at documents related to teaching mathematics in particular is necessary to have a clearer view of this knowledge.

Mathematical Aspects of Teacher's Knowledge

These aspects refer specifically to teachers of mathematics, not necessarily to a specific grade level. The first document considered is a book based on the Proceedings of the Mathematics Teacher Preparation Content Workshop, held on March, 1999 at the National Academy of Sciences. The workshop was designed to set the stages for defining and identifying the mathematical knowledge teachers need to know to teach mathematics well. The published document is entitled *Knowing and Learning Mathematics for Teaching* (NRC, 2001b). The document addresses many different issues related to mathematical knowledge for teaching; one of interest for this study is the organization of knowledge for teaching around what they call "tasks of teaching practice". The discussion was motivated around the fact that often teaching is seen as presenting material to students, but of course teaching includes many more small and large tasks such as figuring out what students know, composing good questions, assessing and revising textbook lessons, and so on. The key question was what are some of these recurrent tasks of teaching that require the use of mathematics?

The participants of this workshop listed many recurrent tasks and discussed which required the use of mathematics. They organized the tasks into six groups: 1) managing class discussion, 2) establishing a classroom culture for mathematical reasoning, 3) designing and selecting tasks, 4) analyzing student thinking and work, 5) planning instruction, and 6) assessing student learning.

Managing class discussion, involves selecting the language/terminology to use to explain an idea or procedure, to pose a task, or to relate to students' explanations and observations; anticipating misconceptions, deciding when to give feedback and what type of feedback to give to students, deciding when to acknowledge "good" mathematical

thinking or an explanation and when to remain nonjudgmental, deciding how to build on what students say, deciding which student solutions or strategies to focus on in whole-group discussions, and assisting students by providing hints to move them along in their thinking.

Establishing a classroom culture for mathematical reasoning involves sharing or developing criteria with students for their work, developing definitions as a group, examining and critiquing student ideas and work as small groups or a whole class, and discussing expectations with students for their mathematical explanations.

Designing and selecting tasks for students involves aspects of pedagogical content knowledge such as selecting the language to use to describe a task, making tasks accessible to a range of learners, selecting a context for a task, evaluating mathematical tasks through a child's eyes to determine the "hard" parts, sequencing the use of mathematical tasks, remodeling mathematics tasks, and selecting mathematical tasks that will yield the best results for student learning.

Analyzing student thinking and work involves interpreting student explanations and making sense of what they are saying, determining the mathematical validity of a student strategy, solution, or conjecture; determining a student's prior knowledge of a mathematical idea; figuring out what students know and do not know, as well as what conceptual knowledge connections are missing or are fragile; and examining student strategies and solutions to determine which are more elegant and sophisticated requiring that teachers have a sense of the range of potential strategies and solutions.

Planning instruction relates to deciding what mathematical topics to teach, composing good questions, making long-range and short-range plans, assessing and revising textbook or resource book lessons, designing lessons, selecting mathematical

models and manipulative to use, and making decisions regarding the amount of time to spend on a topic, lesson, or activity.

Finally, *assessing student learning*, is similar to the previously defined knowledge of assessment seen in the first category of documents. Teachers need to design formal and informal assessments, set criteria to make judgments about student work and analyze and use information from assessments to guide student learning.

Authors of these lists of teaching tasks are aware that some tasks may fit into more than one category but they used their judgment to place each tasks to the best fit. With this type of organization one is forced to look at the work of teachers rather than just examining their mathematical content knowledge. This is particularly useful for designing assessment instruments; it provides the contexts to develop application problems where teacher's work is connected with their mathematical knowledge. One of the pitfalls of a long list like this one is that educators may think that every topic in mathematics can be connected with every task.

The other document that gives recommendations about teaching mathematics is *Adding It Up* (NRC, 2001a), which addresses the question of "What does it take to teach for mathematical proficiency?" In doing so, the authors discuss the kinds of knowledge needed to develop proficiency in mathematics and problems involved in this task. Similar to the framework observed on the state professional standards for teachers, this document lists three kinds of knowledge: knowledge of mathematics, knowledge of students, and knowledge of instructional practices. Knowledge of mathematics includes not only understanding of concepts correctly and being able to perform procedures accurately, but also understanding mathematics "in ways that allow them to explain and unpack ideas in ways not needed in ordinary adult life" (p. 371). This definition can be viewed as the

"pedagogical content knowledge" idea used by Shulman and the teachers' standards. Knowledge of students and how they learn mathematics refers to how ideas develop in children including common difficulties with concepts and procedures. Knowledge of instructional practice refers to knowledge of curriculum, tasks and tool for teaching, design and management of classroom discourse, and knowledge of classroom norms. This kind of knowledge includes knowledge of both pedagogy and planning mentioned in a previous section.

These kinds of knowledge, according to this framework, must be connected so that their use has an impact in children learning. Furthermore, the authors suggest that the connection needs to be made to classroom practice. The suggestion is then to require several interrelated components of proficiency in the context of teaching: *conceptual understanding* of the core knowledge required in the teaching practice, *fluency* in carrying out basic instructional routines; *strategic competence* in planning effective instruction and solving problems that arise during instruction, *adaptive reasoning* in justifying and explaining one's instructional practices and in reflecting on those practices, and a *productive disposition* toward mathematics, teaching, learning, and the improvement of practice.

If we look carefully at this framework we can relate the first four components, that is all except for the *productive disposition*, to the levels of cognitive demand used for statistical content knowledge for students. For example, understanding of core knowledge and fluency would correspond to solving routine problems or in the case of statistics, statistical literacy. Adaptive reasoning would correspond to communicating understanding or statistical reasoning, and strategic competence would correspond to

93

solving non-routine problems or statistical thinking. The comparison is not perfect, but one can see similarities.

Taken from this point of view, the short-comings of the present national and state standards become clear. As opposed to standards for student learning, standards for teacher preparation generally describe what a teacher needs to know, but do not explicitly define the different levels of cognitive demands, nor do they relate these general domains of knowledge to specific tasks teachers must perform in the classroom.

Statistical Aspects of Teacher's Knowledge

Finally, these are the aspects of knowledge for teaching specifically statistics at the middle grade levels. The only document that explicitly refers to knowledge needed for teaching statistics is *The Mathematical Education of Teachers* (CBMS, 2001) report. Although this document does not provide a framework or separate categories as documents mentioned previously, one can isolate the following specific recommendations in terms of content in the discussion (CBMS, 2001, pp. 115-117).

- Prospective teachers must develop both skills for calculation and those for interpretation within the context of the problem.

- Teachers themselves need to learn to be critical consumers of data and statistical claims.

- A teacher's time is better spent on learning to interpret graphs and related summary statistics rather than undertaking tedious calculations.

- Prospective teachers should have practice with and develop understanding of the role of conjecturing using sample data, and they should understand that conjectures as to why certain patterns appear in data are part of the exploratory process. They should encourage their future students to think about data in this way.

- Examples of the misuse of statistical association to make cause-and-effect statements can be brought into class discussions.

- Teachers should understand the process of making inferences through simulated sampling distributions (which can be done effectively in middle grades) and its relationship with more mathematically based inference procedures taught at higher levels.

Other recommendations are also made but these listed above are the ones that relate to teaching the subject. This document focuses more on recommendations about understanding the statistical content than in the other domains. Hence, a textbook was also examined to identify more aspects for teaching statistics as an approximation of teaching statistics in practice. The textbook examined was the *Connected Mathematics* Teacher's Guide Grade 6 Statistics Unit: Data About Us (Lappan et al, 2002).

The content of data analysis and statistics is covered in two units, one in grade 6 called "Data About Us" and the other one in grade 8 called "Samples and Populations". The "Data About Us" unit focuses on formulating questions; gathering, organizing, representing, and analyzing data; interpreting results from data. The "Samples and Populations" unit focuses on using samples to reason about populations and make predictions; comparing samples and sample distributions. Each unit cover subtopics in small sections called "investigations" and each investigation consist of several problems which address a concept or procedure.

Two investigations of the unit of "Data About Us" were selected for examination. This unit and investigation were selected because they best represent aspects of content identified previously in the students' standards and *The Mathematical Knowledge of Teacher* (CBMS, 2001) report. One investigation entitled "Using graphs to group data" focuses on collecting and organizing numerical data in steam-and-leaf plots, locating measures of center and spread, describing the shape of the data, including the location of clusters and gaps, determining what is typical about the data, and comparing two data sets

using back-to-back stem-and-leaf plots and using statistics, such as median and range. The knowledge for teaching identified in this investigation in terms of content was knowing different graphical representation for numerical and categorical data, and knowing that a data set can be represented with different types of graphs. Teachers also need to judge when to use a specific type of graph and justify the choice. In terms of knowledge of students, teachers need to know how to respond to students who want to use an inappropriate type of graph, how to create questions that middle school students can ask in order to collect data and that might involve using a stem-and-leaf plot, and how to assess students' responses making judgments on their reasoning.

As for pedagogical knowledge, teachers need to engage students in exploration of the data given by having them suggest questions that might have originated from the data and methods for collecting the data, leading students in the process of constructing a stem-and-leaf plot and finally, pose questions about data organized in the stem-and-leaf plot that guide students to "read the data", "read between the data" by focusing on reading the stem and on identifying intervals; and "read beyond the data" by focusing on the mode, median, and shape. This investigation also includes reading back-to-back stem-and-leaf plots for comparing two data sets. Here the main aspect of knowledge for teaching identified was knowing how to conduct a discussion with students about the comparison of two data sets. In particular, a teacher needs to know how to respond to students that only focus on a particular statistic or students that have difficulties comparing data sets with different number of data points.

The second investigation of this unit, entitled "What do we mean by mean?", focuses on the concept of the mean as the "balancing" point of the distribution, finding and interpreting of the mean of a data set using physical models leading to the algorithm

and the proper use of the mean, median and mode. The investigation is split into five different problems. In the first problem, students explore different ways to describe the average number of people in six households and the mean is introduced through a visual model using cubes. The model used employs towers of the cubes to represent the 6 different observed data values. Students are then asked to move the blocks around to create 6 towers of equal height. This height represents the mean value. For this problem teachers need to know that there are different ways to determine the average of a data set and how to use them to solve problems. In terms of pedagogy, they need to know advantages and limitations of physical models to introduce the concept of the mean as the "evened out" number; how to make connections between the physical model and the line plot; how to pose questions to students to help them see that the physical model and the line plot display the same information. Teachers need to be aware that in the physical model it is easier for the students to find out the sum of the data values than in the line plot. Finally, teachers need to explain the mean as the balance point in the distribution and relate the "evening out" model to the line plot.

The second problem in this investigation is a continuation of the first and students are supposed to apply what they learn in the first problem. Here the teacher needs to understand how students are thinking about the data when they use different strategies and models to find the mean, and how to assess proper statistical reasoning for justifying students' strategies.

The third problem links the two previous problems and explores the idea that different sets of data may yield the same mean. As for content, teachers need to know how to create data sets with the same mean and different number of data points; create data sets with different mean and same number of data points; create data sets with the

97

mean as part of the data set and not part of the data set. Pedagogically, teachers must understand how to work with the physical models in the classroom. They need to know how to create data sets with the same mean but a different distribution using physical models, to relate the physical model with line plots, to make connections about the number of data points, the total value and the mean, to make the transition between having the sum of the data values as the unknown to the mean as the unknown using models, and to lead students to the discovery of the algorithm of the mean and why it works. Finally, in terms of knowledge of students, teachers need to know how to respond to students who think that it is impossible to have many data sets with the same mean.

In problem four, students use a larger data set than they have used before motivating them to develop an algorithm for computing the mean. For this problem teachers need to know how to assess student's explanations of their strategies for finding the mean of a large data set and why it works. They also need to understand the difference between the median as the physical middle of the values and the mean as the balancing point of the distribution.

The last problem was designed to broaden student's understanding of the mean by introducing some extreme values. The knowledge for teaching needed for this problem is to anticipate students' answers or interpretation to an investigation question such as "How many movies did you watch last month?" and be able to pose questions to students that lead them to see the effect that outliers and/or any new data values have on a stem plot and the mean.

The lack of a common framework and the different purposes of the documents reviewed in this section make it difficult to summarize the different aspects of knowledge addressed in them. However, as with the statistical content analysis, it is useful to list the

important tasks, in this case teaching tasks, teachers are called on to perform. The most

specific knowledge for teaching statistics was found in the examination of the *Connected*

Mathematics textbooks which approximate the actual practice of teaching statistics in

middle school. It is here where the identification of knowledge for teaching statistics was

found most useful in terms of developing items for assessment. Table 3.6 summarizes the

teaching tasks for knowledge for teaching. The rows represent the domains of knowledge

suggested by the state professional standards and the columns represent three of the

components of proficiency described in *Adding It Up* (NRC, 2001a).

Table 3.6

Summary of Teaching Tasks for Knowledge for Teaching Data Analysis and Statistics

	Understand core knowledge, fluency on teaching routines	Adaptive reasoning	Strategic competence
Students as learners	• Describe and anticipate students' misconceptions and limitations when analyzing data. • Know the development of statistical ideas according to age, abilities, and interest.	• Select language to use and context to describe a task. • Select and sequence statistical tasks that will yield the best results for student learning. • Determine and use students' prior knowledge in connection with new information. • Interpret students' oral and written responses in relation to statistical concepts. • Determine the statistical validity of a student strategy, solution, or conjecture.	• Create, adapt, and use tasks for diverse purposes. • Making decisions about when and how to probe for deeper understanding, to give feedback, to react to students mistakes. • Compose good questions. • Assist students with hints to move them along in their thinking. • Examine students' strategies and solutions to infer students' understanding and plan future instruction.
Pedagogy	• Describe different teaching strategies and give examples and counterexamples of statistical concepts. • List concrete material and technology available for instruction in statistics.	• Select powerful teaching methods according to the statistical concept to teach. • Revise textbooks or resource book lessons. • Select statistical models and manipulative to use.	• Use appropriate teaching strategies, materials and technology. • Develop and adapt instructional materials and plans.
Assessment	• Know formal and informal assessment strategies.	• Select assessment strategies according to educational purposes and students developmental level. • Align assessment strategies with what is taught and how it is taught. • Set criteria to make judgments about student work.	• Design and use a variety of formal and informal assessment. • Analyze and use information from assessment to guide student learning and inform students, their parents, and school.

Knowledge for Teaching

domains

Content

Applied to Teaching

dimensions

Topic ✕ Cognitive Demand

subdomains

Students as Learners

Assessment

Pedagogy
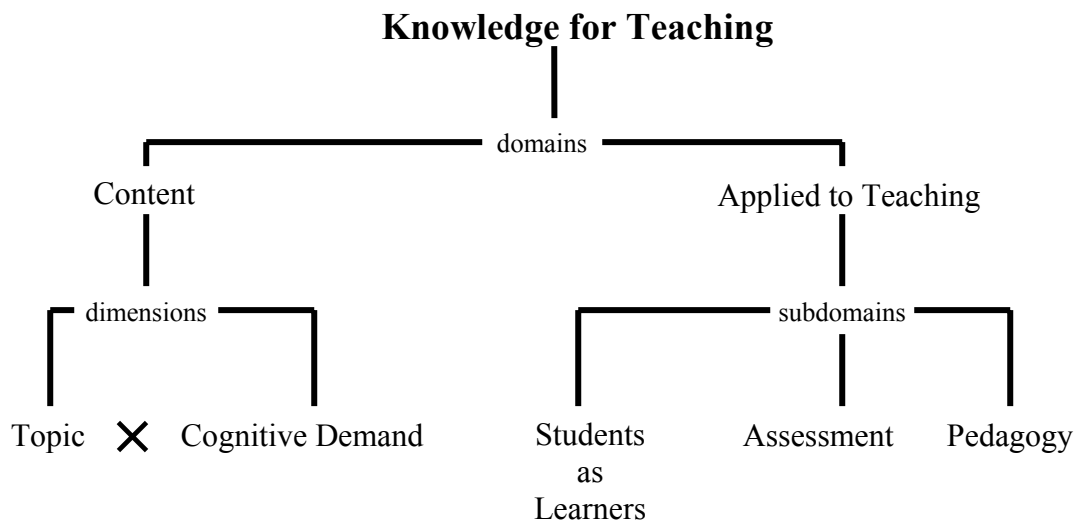
Figure 3.16. Structure for Analyzing Aspects for Knowledge for Teaching.

Summary and Discussion

Figure 3.16 displays the structure used to organize the analysis of documents in this chapter. Knowledge for teaching was separated into two major domains: Content Knowledge and Content Knowledge Applied to Teaching. As shown in Figure 3.16, each domain was then further subdivided. The aspects of content knowledge are tasks that represent a cross between topic and cognitive demand. The tasks needed for teaching in the middle school grades are suggested by state and national standards, and *The Mathematical Education of Teachers* (CBMS, 2001) and are summarized in Figures 3.13, 3.15 and Table 3.5. National standards suggest a broader list of topics and higher level of cognition than state standards. Furthermore, state standards vary considerably in their emphasis of statistical content as well as the level of cognitive demand.

The important aspects in data analysis and statistics identified in these documents, as a whole, are data representation, in particular numerical data representation such as line plots and histograms, and measures of center and spread. The emphasis is on

performing procedures (making the graphs and computing the measures of center and spread), and communicating understanding (appropriate selection and interpretation). Less emphasis was given to solving nonroutine problems (application of data analysis and statistics to real world problems) and conjecture, generalize, prove (make predictions and inferences from the data).

Knowledge applied to teaching is subdivided into three subdomains (Students as Learners, Pedagogy and Assessment) suggested mainly by national and state teachers' standards. Most of these documents are not subject matter specific and only provide general guidelines for what is needed to teach well. Like content standards at the student level, these documents vary in the domains of knowledge suggested for teachers and the cognitive demand. The documents place emphasis on the pedagogy domain, in particular teaching strategies. Second in importance is the domain of knowledge of students as learners, in particular classroom discourse; and informal assessment. The most specific aspects for knowledge for teaching statistics was found in the examination of the *Connected Mathematics* textbooks which approximate the actual practice of teaching statistics in middle school. The aspects, which represent teaching tasks, that were suggested by the NRC documents and the *Connected Mathematics* textbooks are summarized in Table 3.6.

CHAPTER 4

METHODOLOGY FOR DEVELOPING AN INSTRUMENT TO ASSESS
STATISTICAL KNOWLEDGE FOR TEACHING

The analysis reported in Chapter 3 identifies the important aspects of statistical

knowledge needed for teaching at the middle school level. A second question naturally

arises, what do prospective teachers know about the various aspects of statistical

knowledge for teaching? In particular, what do they know about the content and what is

their pedagogical content knowledge of data analysis and statistics? This chapter

describes procedures used to develop and administer instruments for assessing statistical

knowledge for teaching.

Instruments

Statistical Knowledge for Teaching Assessment

The Statistical Knowledge for Teaching Assessment instrument has been

designed to measure two major domains of knowledge. One is purely statistical

knowledge, and the other is statistical knowledge applied to teaching. The analysis in

Chapter 3 of documents, state and national standards, and several units from a set of a

curriculum materials identified many more aspects of knowledge than can be measured

by a single instrument. For example the content matrix in Table 3.2 contains 35 cells,

representing 35 potential different content aspects and Table 3.6 contains over 25

teaching tasks. To narrow the focus to a more manageable level, a selection of aspects

was necessary. To determine the topics and levels of cognitive demands to best assess purely statistical knowledge the content maps in Figures 3.13 and 3.15 were combined and modified to create Figure 4.1. The five levels of cognitive demands were reduced to three. Given the fact that none of the 14 documents analyzed contained any content at the level of memorization, this level was dropped. Furthermore, the two highest levels were collapsed. With this modification the levels of cognitive demands can be matched with the framework suggested by Garfield (2002) and delMas (2002). Perform Procedures corresponds to *Statistical Literacy*, Communicate Understanding to *Statistical Reasoning* and the combined levels of Solve Nonroutine Problems and Conjecture, Generalize, Prove correspond to *Statistical Thinking*. The different shaded cells on the figure correspond to the intersection of a topic with a particular cognitive demand. Examination of Figure 4.1 reveals that 15 of the 21 content cells have a percentage of 4 or higher. Of these content cells 9 were included in final instrument. The cell with the highest percentage, process of statistical investigation at the statistical thinking level, is associated with designing studies for inference and is beyond the scope of a timed written instrument. All cells with next highest percentages, representation of numerical data and measures of center, are included in the instrument. Items on the representation of categorical data were included in a piloted version of the instrument. However, they were found to be too easy to be of use and were dropped from the final instrument. The lowest level – statistical literacy - of the process of statistical investigation is associated with formulation of questions to produce data. Although an item related to this topic was included in the instrument its focus is on the domain of knowledge for teaching. The two levels – statistical literacy and reasoning - of the measures of spread with percentages

above 4% are included, but due to space and time constraints, bivariate data

representation was not. Shape of distributions, which has the lowest percentages of

coverage (below 2.9%), was not included in the instrument.   It should be noted however

that the distinctions between the different cells is not entirely clear cut. For this reason,

one item may measure more than one cell. For example, item 4a was created to asses the

interplay between measures of center and spread; so it was counted in two cells.
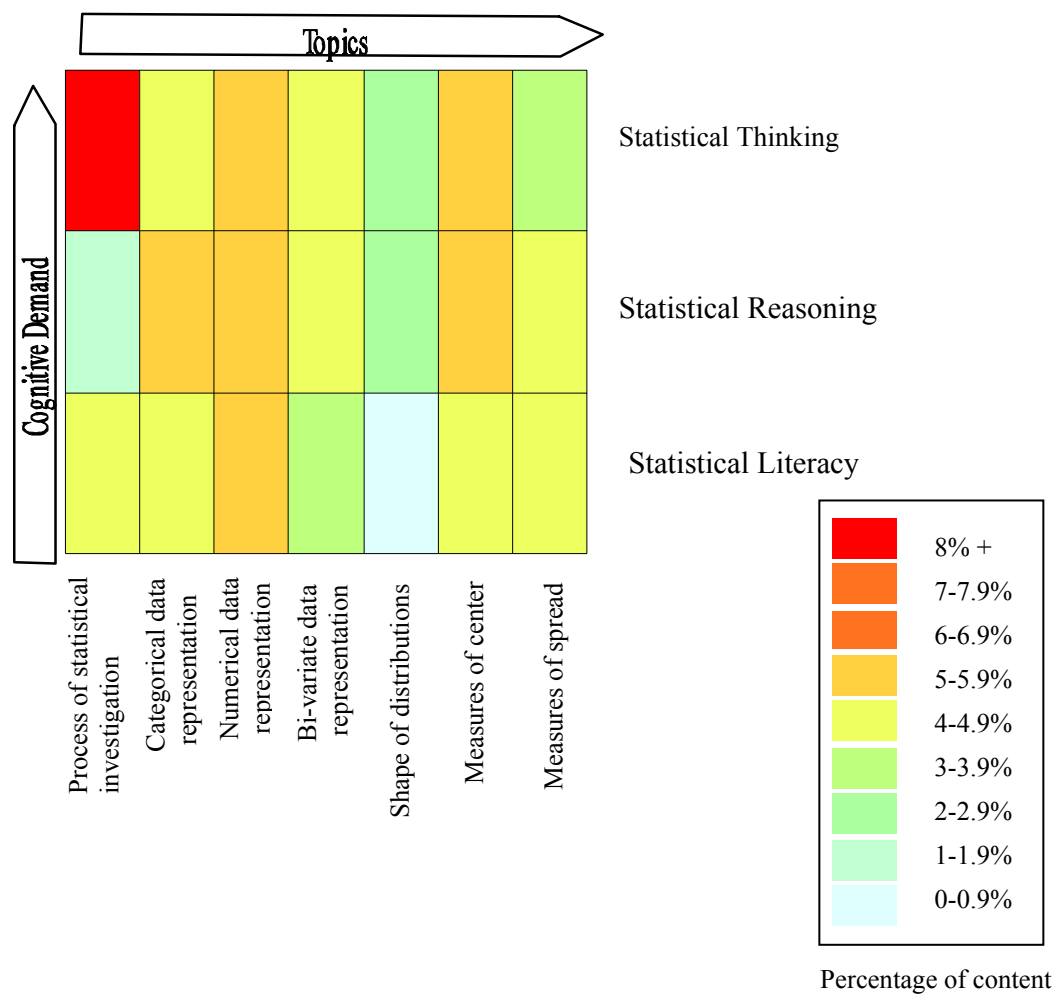
Figure 4.1. Content map for the ten states standards, the Mathematical and Problem-Solving Goals in *Connected Mathematics* Teacher's Guide Grade 6 and 8 Textbooks, *Principles and Standards for School Mathematics* (NCTM, 2000), *The Mathematical Education of Teachers* (CBMS, 2001) and the Topics Covered in PRAXIS II Middle School Mathematics Test.

As for statistical knowledge applied to teaching, the instrument focuses on the knowledge of students as learners. Pedagogical knowledge and assessment knowledge, the other two domains described in the previous chapter are not considered here and are left for future investigation. Twelve teaching tasks in this domain are listed in Table 3.6. The following two were considered most conducive to this type of assessment and are included in the instrument:

- Interpretation of students' oral and written responses in relation to the content.

- Examination of students' strategies and solutions to exercises to make inferences about their understanding.

Development Procedures

Two pilot studies were conducted. In the first study, several items were tested using only a written format. Items were analyzed and a second phase was conducted to refine the selected items and perform a couple of interviews. The participants in the first phase (Fall 2001) were prospective teachers planning to receive K-8 teaching certificates from three different institutions: Michigan State University, American University, and Montgomery College. A total of 42 prospective teachers participated in the first phase, the majority (30 out of 42) of which were females in their early twenties. All the participants had taken at least three credits of a mathematics for teachers' course, and an average of nine credits of college mathematics. About two thirds of the participants had at least three credits of college level statistics.

Seventeen items were tested for construct reliability, wording, and format. The main purpose of the first phase was to analyze these items, that is, to identify what items

were more successful measuring the construct desired, the effect of the format of the responses (close form or open ended), and the nature of responses. Furthermore, the analysis helped to develop the initial rubric or classification of the responses into patterns and dominant answers to help the researcher to create a rough picture of prospective teachers' statistical knowledge.

Items were examined in isolation first. For closed format items, responses were tabulated accordingly to the correct answer and distracters. For open ended and short format, direct responses were recorded creating naturally emerging categories, with the purpose of not imposing any preconceptions on the nature of responses. When available, these categories were compared against existent literature. Items were coded and a small database was created.

Items covered different levels of performances, from computational and procedural skills to higher-order thinking. The items were adapted and/or used from the literature to address the different aspects of knowledge for teaching described earlier. Item sources included the *Collecting, Representing, and Interpreting Data Module*, San Diego State University, TIMSS 1999 Assessment, Friel, Bright, Frierson, and Kader, (1997), PRAXIS II: Middle School Mathematics (ftp://ftp.ets.org/pub/tandl/0069.pdf) , Liu (1998), Watson (1997), Susan Jo Russell (personal communication July 16, 2001), and Senk et al (1998).

The seventeen items, each with multiple parts, were separated into two forms. Form 1 had eight items and Form 2 nine items. Items in the two forms were distributed so that the forms covered similar content and difficulty. Pilot responses were coded for correctness (utilizing a rubric for each item) and statistical tests of reliability were

conducted. Reliability of the pilot instrument was tested using a Cronbach-Alpha. The results showed the following values: Form 1 α = .4731, Form 2 α = .6627.

Items that did not provide extensive insight into either content knowledge or knowledge about students were discarded. Some were re-written to improve the wording and format, and others were added. As expected, prospective teachers were able to perform well on the items that required extracting information from a graph, and computing and defining measures of center and spread. Although it is necessary to keep these types of items because they help provide a fuller picture of teachers' knowledge, fewer of them appear in the final version of the instrument. For example, pilot results show that the majority of prospective teachers can extract information from and interpret bar and pie graphs. The mistakes identified were related to proportional reasoning or scale errors. The final instrument measures the use of graphs for numerical data instead.

In addition, the pilot instrument did not cover some of the important aspects of statistical investigation such as the prospective teachers' ability to formulate questions and their knowledge of measures of spread. These two aspects were added to the final version.

This new version was pilot tested in a second study (Spring 2002) with 16 prospective teachers from the state of Maryland. Two prospective teachers were invited to participate in a follow-up interview. The second instrument piloted consisted of 11 items, items 1 – 5 consisted in short questions about statistics and items 6 – 11 where short questions about statistics applied to teaching. Most of the items of the second pilot were kept for the final version. Some were used for the follow-up interviews.

Description of Items

The final version of the instrument is given in the Appendix C. It consists of 8 free-response items with multiple parts intended to be answered in 50 minutes. Some of the items cover statistical knowledge without reference to teaching while others cover statistical knowledge for teaching in relation to students' thinking. Table 4.1 shows the distribution of items for the domain of statistical knowledge. Note that a single item or part of an item may measure more than one aspect of statistical content. For example, item 4a measures numerical data representation *and* measures of center because this item consist of the presentation of a graphical display where the prospective teacher is asked to find the mean (see instrument assessment in Appendix C).

Table 4.1

Distribution of Items by Content and Cognitive Demand

Aspects of statistical content	Cognitive Demand		
	Statistical literacy	Statistical reasoning	Statistical thinking
Numerical data representation	1a, 4a, 6b	1b, 6a	1c, 8b
Measures of center	4a	2, 4b, 4c	1c
Measures of spread	8a	8b	

Items were also designed to assess statistical knowledge applied in teaching, in particular, to applications to students' thinking about data, data displays, and measures. In each item the prospective teacher is confronted with either a student response or solution to an exercise. In contrast to the items related to statistical knowledge, these items are not characterized by cognitive demand. In some items (3a, 3b, 5a, and 5b), the

prospective teacher is asked to judge whether a response is correct and then to explain what thought process the student might have used to arrive at that response. In others (items 7a and 7b) the prospective teacher is asked to describe the method or solution used by the student and then to make inferences about his/her understanding.

Item 1, which involves a stem-and-leaf plot, has been adapted from an example of an investigation presented in Friel et al. (1997). It was originally designed to measure the three levels of graph comprehension identified by Friel, Curcio, and Bright (2001): (a) extract information, (b) find relationships, and (c) move beyond the data. These levels closely correspond to the levels of cognitive demand of the framework used in this research.

Item 2 measures knowledge of the appropriate use of measures of center based on the shape of the data and the effect of outliers. This item is a variation of an item from the Statistical Reasoning Assessment (Lui, 1998) which originally was developed to measure the understanding of how to select an appropriate average, to identify the misconception that averages are the most common number, and the mistake of failing to take outliers into consideration. In addition, the context of this item suggests the view of the average as a signal in noise (Konold, 2002). According to this view, "each observation is an estimate of an unknown but specific value. Each observation is viewed as deviating from the actual weight by a measurement error, which is viewed as 'random'. The average of these scores is interpreted as a close approximation of the actual weight" (p. 269).

Item 3 measures graph comprehension (histogram) at the level of interpretation. Part (a) measures the ability of the teacher to recognize a common student misconception

110

and to interpret the source of misunderstanding. Part (b) measures interpretation and judgment of a student's oral response. While similar in nature, the two parts differ in the degree of misunderstanding demonstrated by the student. In part (a) the student answer is incorrect, while in part (b) the student statement is more incomplete than incorrect. The item hopes to measure the ability of the teacher to differentiate between the two. This item is a variation of a problem from the Collecting, Representing, and Interpreting Data Module developed at the Center for Research in Mathematics and Science Education at San Diego State University. The misconceptions and students' responses are authentic, taken from the pilot study.

The data for items 4, 5 and 6, the graphical displays and suggestions for student responses were taken from the *Connected Mathematics* units (Lappan et al., 2002). Item 4 measures knowledge of calculating the mean from a line plot, which implies assessing for knowledge of graph comprehension and the concept of the mean beyond the algorithm. It also measures properties of the mean and the ability to create distinct distributions with the same mean. Item 5 measures knowledge about formulating questions to generate data and measures of center and range for categorical data. This question measures the ability to identify errors in students' responses. Item 6 measures the proper selection of data representation, taking into account the shape of the data and the ability to create a graph.

Item 7 measures the level of interpretation of a graph and the ability to examine and make judgments about student work based on statistical reasoning. It also measures the ability to describe students' thinking and infer about their understanding. This item

was adapted from an item shared in a personal communication with Susan Jo Russell (July 16, 2001).

Finally, item 8 measures in part (a) computational knowledge of the mean, range and standard deviation from a line plot of two data sets. Part (b) measures the recognition of the two data sets' shape in relation to the measures calculated in part (a). This item is based on an example from *Functions, Statistics, and Trigonometry* (Senk et al., 1998).

Validity of the instrument was not conducted by the judgment of experts on the field but instead by using the analyses of documents described in chapter 3 and by using the theoretical perspective described in chapter 2.

As in the pilot study, reliability of the instrument was tested using a Cronbach-Alpha. The results showed greater reliability with $\alpha = .80$. Reliability was also tested for the two different domains of knowledge measured. The part of the instrument measuring statistical knowledge (items 1, 2, 4, 6, 8) has $\alpha = .74$ and the part that measures statistical knowledge for teaching (items 3, 5, 7) has $\alpha = .53$.

Procedures

The population under study is the set of prospective middle school teachers in the last stage of their professional education in the U.S. The states considered for the sample are those that both require a standardized mathematics content test[1] and offer Middle School Certification. Ten states meet these criteria, and they are Connecticut, Florida, Georgia, Kentucky, Missouri, North Carolina, Ohio, Oregon, Virginia, and West

Virginia. It is believed that colleges and universities in these states are more likely than others to have teacher preparation programs that focus on middle school content and pedagogy.

Three major universities in these states were selected to participate in the study. These are the University of North Carolina, Kenessaw State University and the University of South Florida. These universities were selected for two reasons. First, each of these institutions has a large program and strong reputation in teacher preparation. Second, personal contacts in each were willing to cooperate with the study. In addition students from the University of Maryland and Towson University participated in the study. Although the state of Maryland does not have middle school certification, the College of Education at the University of Maryland is among the top ranked programs in the nation[2]. Furthermore, their geographic proximity to the researcher made the process of interviewing convenient.

Within each university a faculty member, or some representative of the researcher, identified prospective teachers who qualify for the study. Although in most cases these students were clustered in a mathematics education class, some were identified by the faculty member within their education program. Subjects were asked to answer a few questions about the mathematics, mathematics education, and education courses they had taken. These questions were customized for each institution according to the official academic program found in the respective web site.

---

[1] PRAXIS II: Middle School Mathematics (0069) (Retrieved June 30, 2004 from ftp://ftp.ets.org/pub/tandl/0069.pdf)
[2] University of Maryland's Curriculum and Instruction Program raked 11th in the nation by 2004 U.S. News & World Report Guide.

Participants

A total of 42 prospective middle school teachers participated in the study. Most of subjects that participated in the study were female seniors in their twenties. Table 4.2 presents the distribution of the sample by gender, age, and class.

Table 4.2

Distribution of Subjects by Gender, by Age and by Class

Variable	Frequency
Gender	
Female	33
Male	9
Age	
19-23	22
24-29	7
30-35	1
Over 35	12
Class	
Junior	9
Senior	29
Graduate	4
Total	**42**

Subjects for this study have a strong mathematics background. All of them had taken at least one semester of Calculus, and 38 of the 42 prospective teachers surveyed had taken at least one course beyond calculus. The average number of mathematics classes starting at calculus is 5.2. Some of these classes include Calculus II and III, Differential Equations, Linear Algebra, Geometry, Number Theory, and History of Mathematics among others. All but four students had taken a basic or introductory statistics class at the college level. The performance of these four students was not different than the rest of the participants and therefore were included in the final analysis.

Mathematics education classes were harder to account for; for some institutions these belong to the education department and for others they belong in the mathematics department. The emphasis on content or pedagogy was also hard to tell, that is, it was unclear if the mathematics education classes were more focused on content knowledge for teaching or pedagogical knowledge. For the purpose of this study, all of these classes are labeled mathematics education classes. Subjects participating in this study had an average of 2 mathematics education classes.

Finally, participants were also strong in education classes, with an average of 4.5 classes. Some of the titles of the courses taken were Foundation of Education, Education Psychology, Human Development and Learning, Multicultural Prospective, Teaching & Schools, and Curriculum and Instruction.

Analysis of Statistical Knowledge for Teaching Assessment

The aim of the analysis of the written instrument is to create a global picture of prospective teachers' statistical knowledge for teaching. The responses were characterized not only by identifying levels of correctness but also by observing response patterns and solution strategies. Therefore, the analysis of the written instrument was done at two levels, item analysis and global analysis. Item analysis was conducted by coding each item's responses utilizing a rubric for levels of correctness. For many of the items in the final version of the instrument, rubrics were developed using the pilot data, but the rubrics were refined with the responses from the final version. To assure reliability of the scoring, two graders – the researcher and a trained grader - coded the

responses independently. The researcher adjudicated when there was disagreement between graders.

<div align="center">Rubric</div>

Since many questions on the instrument are free-response, a rubric was needed to score the items. For items that measure statistical knowledge, it was desirable to have a consistent scale in terms of level of performance and conceptual understanding that can be applied across all items. The scores can then be combined across the items that measure statistical knowledge to obtain an overall score and give a clearer picture of the level of performance of prospective teachers. A general rubric outlined below describes the important criteria that were taken into consideration for all items measuring statistical knowledge. This general rubric was then interpreted to develop a more specific rubric for each item.

A holistic-scoring method, adapted from two sources: an analytic-scoring method offered by Garfield (1993) for evaluating students' solutions to practical statistical projects and a holistic-scoring method used by Thompson & Senk (1993, 1998) to assess problem solving and conceptual mathematical understanding, was used to score the items in the instrument. Garfield's analytic scoring method considers multiple dimensions such as communication, visual representation and interpretation of results and uses criteria that are specifically adapted to statistical content. However, each item is given multiple scores, one for each dimension. A holistic-scoring method described by Thompson & Senk (1993) creates a single scale and makes the process of scoring and data entry faster and easier taking into account the assessment of procedural and conceptual understanding. By adapting Garfield's ideas to a single holistic scale, it is hoped to

obtain the best of both methods. The integration of both methods was done by associating the levels of correctness with the appropriate language used in statistics. The criteria listed below provide a generic rubric. Prior to actual grading, the criteria were further specified and customized to create item specific rubrics.

Successful responses

4 Solution is complete and correct. Language and notation used is correct; tables and graphs are correctly constructed; all decisions are made correctly; data sets are interpreted correctly using all appropriate information.

3 Solution is almost complete and correct, but some minor error is made, perhaps in use of language, missing labels and scale on graphs, or calculation error with valid reasoning.

Unsuccessful responses

2 Response is in the proper direction and contains some substance, such as a chain of reasoning. But either the prospective teacher stops about halfway through the solution or the complete solution contains some major conceptual errors. The use of language is partly appropriate *or* some decisions about the selection of graphs for representing data and summary measures seem inappropriate *or* the interpretation is too brief, the prospective teacher fails to interpret some important information *or* weak conclusions are made, but some attempt is made to look beyond the data.

1 Some work is correct, but the student reaches an impasse early. The work shows no evidence of a chain of reasoning. The prospective teacher uses statistical words in a context that does not make sense *or* errors in calculation lead to

unreasonable answer *or* the respondent is unable to interpret the plots and

measures *or* draws conclusions not substantiated by the data.

0   Work is all wrong or meaningless. No correct statistical knowledge is used for a

solution *or* there is not enough information to evaluate.

With numerical scores assigned to each item, percentages of prospective teachers

that reach a certain level of correctness on each item is reported. In the case of the items

that fall into the pure statistical knowledge category, a percent of prospective teachers

that reach each level of performance is reported as well. Since the goal is to characterize

teachers' knowledge as a whole, individual scores are not reported.

The global analysis was done by combining scores of the items that correspond to

the two major domains of knowledge: statistical knowledge and statistical knowledge

applied to teaching. To summarize the overall performance, average scores are reported

as well as averages of each domain of knowledge. Furthermore, the percentage of

prospective teachers who demonstrate at least a given level of correctness (0 to 4 scale)

on all the items within a content category is reported. This type of summary permits an

analysis and comparison of the performance in each category. For example, the

percentage of teachers who are successful on reading graphs (at least level 3 on all items

related to reading graphs) can be compared to the percentage who demonstrates success

on measures of center.

For statistical knowledge applied to teaching, the performance was assessed from

two perspectives. First, the ability of prospective teachers' to identify the correctness of

student responses was measured by combining the results from items testing knowledge

of student thinking. Secondly, for the items that prospective teachers are asked to explain what thought process the student likely used to arrive at that response. Responses were examined to characterize their understanding of the statistical reasoning of students. The key ingredients of this characterization was the level of agreement between the misconceptions cited by the prospective teachers and the literature on statistical education reviewed in chapters 2 and 3, and the depth of insight demonstrated by the prospective teacher's responses.

<div align="center">Follow-up Interviews</div>

<div align="center">*Purpose and Description*</div>

The second stage of data collection consisted of face-to face interviews with a subset of prospective teachers from the sample. The interview has two proposes: (1) clarify individual student responses of the written instrument and to probe more deeply into student understanding of knowledge for teaching, and (2) assure reliability on the written instrument. The information gathered from the written instrument is sufficient for a general description of some of the aspects of knowledge for teaching, but it is limited and sometimes hard to interpret. By asking prospective teachers to explain their thinking, reflect on their responses or give explanations for incomplete responses, a more accurate and detailed picture of the subjects' knowledge for teaching can be developed. Furthermore, the interview responses complement the information on questions that are too difficult to answer in another format.

Ten prospective teachers were selected to participate in the interview process. The criteria for selection consisted of physical proximity to the researcher and representation of low, middle, and high level of overall performance in the written instrument. Seven of

those indentified met the appointment for the interviews. Attempts to reach the other three were unsuccessful. After the selection and invitation to participate in the interview, subjects' individual responses to the written instrument were analyzed prior to the interview and a customized interview protocol was prepared for each subject (see Appendix E).

Customized protocol questions were guided by the nature of the written responses. For example, for Item 1 a new stem-and-leaf plot was presented to the interviewee according to what they have said was the typical value of the data set presented in the written instrument. The new plots varied according to what they had picked for the mean, median, or mode. This was done with the purpose of finding out whether prospective teachers took into consideration the shape of the distribution when choosing measures of center or they associated the word "typical" with any particular measure of center.

Of the seven subjects from the North Carolina and Maryland sites interviewed, two had reasonably high performance in the written instrument, two were about average, and three had low performance. Interviews were conducted following the protocols; besides the protocol questions, informal probing questions based on the specific answers each subject gave to the prepared questions were asked to clarify ambiguous responses and to unveil specific dimensions that seemed important.

Analysis

The interviews were audio taped. The analysis began by listening to the taped interviews and transcribing them. The complete transcripts are in Appendix F. Then each interview transcript was analyzed by item, and by domain of knowledge.

By item: For those items where several interviewees are questioned, a summary table for each item was created across subjects. Special attention was given to the method used for answering the item, ways of thinking about a concept, ideas, ways of assessing students' comments and responses, and other significant comments.

By domain of knowledge: As in the analysis of the written instrument, items that measure pure statistical knowledge and knowledge for teaching were be combined within each domain and responses were summarized. Special attention was given to the way prospective teachers think of concepts in and out of the teaching context.

The results of the interviews were used to enrich the item analysis of the written instrument and appear embedded when appropriate in Chapter 5.

CHAPTER 5

MEASURING STATISTICAL KNOWLEDGE FOR TEACHING

This chapter addresses the second research question - what do prospective teachers know about the various aspects of statistical knowledge for teaching? The first section describes the performance of prospective teachers at the item level followed by a section on overall performance. The last sections describe prospective teachers' knowledge of statistics and knowledge for teaching. The final section summarizes the findings.

Item Level Performance

The written instrument consisted of 8 free- response items with multiple parts. Each item was graded on a scale of $0 - 4$. For a general description of the rubric see Chapter 4. Each item is followed by a description of what the item was designed to measure and the corresponding distribution of scores. Next, the scoring rubric is described followed by the analysis of results. Last, a summary and discussion of results is presented.

Item 1

Item Specification

Item 1, which involves a stem-and-leaf plot, was adapted from an example of an investigation presented in Friel et al. (1997). It is designed to measure the three levels of

graph comprehension identified by Friel, Curcio, and Bright (2001): (a) extract

information, (b) find relationships, and (c) move beyond the data.

The stem-and-leaf plot below shows the number of minutes it takes

students in a class to travel to their school.

Minutes to Travel to School

0	3 3 5 7 8 9
1	0 2 3 5 6 6 8 9
2	0 1 3 3 3 5 5 8 8
3	0 5
4	5

a. How many students are in the class? _____

b. How many students took less than 15 minutes to travel to
   school? _____

c. What is the typical time it takes for students to travel to
   school? _____

Explain your answer.

Figure 5.1. Item 1

Table 5.1

Distribution of Scores for Item 1 (n = 42)

Item	Score				
	0	1	2	3	4
1a	4	0	0	0	38
1b	3	0	0	0	39
1c	3	8	6	9	16

Description of the Rubric for Item 1

For part 1a, a score of 4 was given to the correct answer, "26 students" and 0 otherwise. Similarly for part 1b, a score of 4 was given to the correct answer of "9 students" and 0 otherwise.

For part 1c, a score of 4 was given to responses that calculated measures of center correctly, that is, responses with the mean (18.46) or the median (18.5), or the mode (23). Score of 3 was given to responses that reported measures of center of clusters followed by appropriate statistical reasoning. For example, the following response got a score of 3, "the range of 20-28 is the largest range for minutes to travel to school, and there are 3 students who take 23 minutes. Therefore, I chose 23 minutes for typical time." Also, a score of 3 was given to responses with correct computation for the mean but rounded from 18.46 minutes to 18 minutes. Score of 2 was given to responses that chose some kind of measure of center but the explanation was vague. For example, "20 minutes. About ½ between 16 and 25 min. which is about 20". A score of 1 was given to responses that chose some kind of measure of center but made a mistake in the calculation or did not provide any work or explanation.

Analysis

In terms of correctness, 93% of the prospective teachers can successfully "extract information" and "find relationships" from a steam-and-leaf plot. More specifically, they can correctly count how many data points are represented in the plot and how many of those data points are below a certain data value (see Item 1a in Figure 5.1 and Table 5.1). In contrast, only 59% of the prospective teachers were able to "move

beyond the data" and provide a successful response (4 or 3 on the rubric scale) to the question "What is the typical time it takes for students to travel to school? Explain your answer."

The responses for part 1c, which reflect prospective teachers' conceptions of and reasoning about what is typical and their reasoning behind their choice, show that prospective teachers do not associate the word "typical" with measures of center. About a third of them used the mean to describe a typical value, about 21% chose the mode, and about 14% chose the median. The other third of prospective teachers chose an alternative measurement or description. Table 5.2 summarizes the type of responses for this part of the item.

Table 5.2

Distribution of Prospective Teachers' Conception of "Typical"

"Typical" as	Frequency
the mean	14
the mode	9
the median	6
a range of a cluster	5
a mode/median of a cluster or stem	5
Other	3
Total	42

Here are some examples of responses in the category of "typical as a range":

"Data clustered from 0 to 28."

"10 to 25"

"midteens to nearly twenties, basically it seems by look to be what the average would be"

Explanations are more sophisticated for the category of "typical as mode or median of a cluster or stem". Here are some examples:

A stem of 2 has most students in that stem. Most of data centered around 20 minutes.

The larger cluster occurs at 23 minutes, there are also clusters at 16 minutes, 25 minutes and 28 minutes. So it takes students, roughly, 20 minutes to get to school.

Most of the students were between 10-28 the majority between 20-28 with 23 occur 3 times while others 1 or 2.

Further inquiry was made about the interpretation of "typical" when six prospective teachers were given follow up interviews. Prospective teachers were presented with a second data set in which their previous interpretation of "typical" was no longer appropriate. Every prospective teacher interviewed used the same method of finding the "typical" value regardless of the distribution of the data set. The following vignette, from Prospective Teacher A, illustrates the choice of picking the median because the data are already organized. For this particular prospective teacher, another data set was presented where her first choice for typical was not appropriate.

Int.: Part c of Item 1 refers to what is the typical time for students to travel to school, and I am curious to know what each person thinks "typical" means.

A: It could mean any kind of a center, to me it could mean median, or it means…you know…not normally I don't look at mode, but it could mean mode.

Int.: So, you picked the median.

A: I thought it was easier, because it is already organized.

Int.: OK, so, let me ask you the same question for this other set of data.

Minutes to travel to School

```
0 | 3  3  5  5  6  7
1 | 0  0  1  4  8
2 | 1  1  2
3 | 5
4 | 3  3  5  6  7
```

A: I would find the median, it is organized, so…

Int.: Would you still stick with the median?

A: Ummm…[pause]…it is about 18, right? To me it is easier if it is organized or some kind numerical order.

Summary and Discussion

Prospective teachers show graph comprehension at the level of extracting information and finding relationships. However at the level of moving beyond the data, prospective teachers tend neither to use statistical measures nor statistical reasoning to infer about the data. It seems that the performance at the latter level is influenced by the undefined term of "typical", which has several meanings.

Item 1 is a variation of an item used by Friel et al. (1997) to assess graph comprehension for middle school students. Friel and others found similar results with middle school students after instruction. They found that students' alternative responses to measures of center fit into two categories, responses that identify a cluster of time and responses that provided a tally or range of numbers that occurred most frequently. An attempt to use the same framework was done with prospective teachers expecting similar reasoning. However, different patterns were found for prospective teachers. Prospective teachers not only looked at cluster of times, but also focus on the center of those clusters by finding the median or mode of the cluster.

Item 2

Item specification

Item 2 measures knowledge of the appropriate use of measures of center based on the shape of the data and the effect of outliers. This item is a variation of an item from the Statistical Reasoning Assessment (SRA) developed by Lui (1998) and Garfield (2003). In the SRA, the item has a multiple-choice format where incorrect and correct reasoning were given as choices. In this study, the item was changed to an open-ended format.

Nine students in a science class weighed a small object on the same scale separately. The weights (in grams) recorded by each student are shown below:

6.2   6.0   6.0   15.3   6.1   6.3   6.2   6.15   6.2

The students want to determine as accurately as they can the actual weight of this object. They may use the following methods:
    I. Use the most common number, which is 6.2.
    II. Use the 6.15 since it is the most accurate weighing.
    III. Use the result of adding up the 9 numbers and dividing by 9.

  As a teacher, what method would you prefer your students to use?
    a. Method I
    b. Method II
    c. Method III
    d. Other _____
Explain your choice.

Figure 5.2. Item 2

Table 5.3

Distribution of Scores for Item 2 (n = 42)

Item	Score				
	0	1	2	3	4
2	10	5	8	8	11

Description of the rubric for Item 2

The level of correctness for this item was measured judging both the method picked and its explanation. An ideal response (score 4) would involve choosing an alternative method of throwing out the outlier and taking the mean of the remaining data values with indication of the effect of the outlier in the explanation. For example, "I would not want students to use method III because of the outlying data value of 15.3. If this data value were used to find the mean weight it would raise the mean value (cause it to be high). It would be best for students to find the mean of the data while excluding the outlier value of 15.3".

A nearly ideal response (score 3) would involve responses that gave indication of the same correct reasoning as the ideal response but either did not communicate clearly the reasoning *or* choose the mode/median being aware of the outlier and viewing the measure of center as a balance point. For example, "I would use the mode because 15.3 would not give me an accurate average, it is considered to be an outlier. I would get rid of the outlier and do the mean if I had my way" *or* "I usually would have the students add up the numbers and divide by 9, but the 15.3 really throws the average off. So, 6.2 is the most common number and all the other numbers are slightly above 6.2 or slightly below 6.2."

A score of 2 was assigned to those responses that indicated awareness of the outlier but chose the mode with an incorrect reasoning. For example, "the most common is most likely the most accurate" or "measurement is not exact therefore having 6.2 as a measurement seems practical". A score of 1 was given to those responses that had no

evidence of awareness of the outlier and chose the mode *or* chose the mean but in their

explanation gave indication of the effect of outliers. For example "If the weight of the

object appeared the same a couple of times, then that would be the most accurate to me"

or "use the mean, this way each student's measurement is weighted the same. Although,

if there were any outliers, I would have them throw it out."

Finally, a score of 0 was assigned to those that chose any of the methods given

with none or limited explanation *or* gave incorrect reasoning. For example, "use the

mean because it is the method that can allow you to find the average" or "Taking the

average will be more accurate".

Analysis

A few less than half of the prospective teachers (19 out of 42) provided a

successful response to this item (see Table 5.3). Although there was evidence in all of

the responses of the effect of the outlier, their explanations vary in complexity. Some

were very simple as "It will skew the data if a straight average is taken." Others were

more sophisticated as:

> I would not want students to use method I because it is not usually the case that
> the most common answer is correct or close to correct. However, in this situation
> because all of the data values (with exception of 15.3) are so close in value, this
> method would not be a terrible method. I would not want students to use method
> II because the scale might not show zeros to the right of the decimal. If this were
> the case, every other reading (i.e. 6.2) would actually have a zero on the end (i.e.
> 6.20). Therefore, every reading would really contain numbers to the same place
> value. I would not want students to use method III because of the outlying data
> value of 15.3. If this data value were used to find the mean weight it would raise
> the mean value (cause it to be high). However, it would be best for students to
> find the mean of the data while excluding the outlier value of 15.3.

A few of the prospective teachers identified with the statement in the item that says "As a teacher, what method would you prefer your students to use?" and used some pedagogical language in their explanations; the following responses illustrate this point:

> I chose this method because I would explain to my student that by collecting all the data and figuring out average; that would be the most accurate way. I would go on to tell them that if we notice most of the numbers are in a range of 6.0 to 6.3, therefore our average is around those numbers. I would tell that the reason that we would not include 15.3 is because it is an outlier, that somewhere we made an error because it is not consistent with all the others.

> This is the method a real scientist would use. I would want my students to know that. Also, the student who answered "15.3" needs some help with measuring weights and balances. The students also need to know why methods I, II, and III are not the best choices.

Furthermore, Prospective Teacher B thought that choosing a method in this case is like choosing different strategies to solve a mathematics problem where the strategy is a matter of preference.

> *Int.:* Why not method I?

> B: You could, this is just another way to find the average. It depends of[sic] the average you want, it is not wrong. As a teacher you have to teach and accept other ways. But I would prefer the method I choose.

Unsuccessful responses also vary in their type, and many conceptions are unveiled with this open-ended format. Prospective teachers that choose the mode or the most accurate weighing as the most accurate method to determine the actual weight of the object have different levels of reasoning. Their responses fit mainly three categories. One category of responses is to choose the mode or most accurate weighing simply because of effect the outlier, for example:

> I choose method I because I felt the 15.3 would give a false average and there is no way to tell if 6.15 is the most accurate weight.

I would tell the students to pick the mode instead of mean because the last weight of 15.3 is too extreme to get an accurate measurement.

I choose method II because the most common doesn't make the weighing the most accurate and the average would be skewed by the 15.3.

Another category fit responses that chose these methods because of the conception that accurate measurements are reflective of the ability of a person to read a scale. These subjects do not view each observation as deviating from the actual weight and there is no evidence that takes the outlier into account.

Because when weighing something and in any case when the majority of students make an experiment and get the same answer, then I would go with the majority

Because if the weight of the object appeared the same a couple of times, then that would be the most accurate to me. In this case I might have the students take the average of 6.0 and 6.2 b/c they both appear more than once.

Small object is always the same weight; looked at the data to determine 6.2 since it is the mode.

Finally, a third category fit those responses that do not show evidence of statistical reasoning. Such as

The mode is the most reflective of the data set.

Might want to think about 6.15 because it is more 'accurate' maybe round it.

I would want my class to understand the importance of using the most accurate weight, so I would definitely have the class use method II.

Although choosing the mean (Method III) in this problem was a measure of failing to take outliers into consideration in the Statistical Reasoning Assessment (SRA) developed by Garfield (2003), five out ten prospective teachers that chose the mean said that they would pick it because they want to take into account the low and high values or because they want representation of every data value.

132

The best way to find the average weight in this example would be to find the mean. That would take into account the higher numbers, lower numbers, and the numbers in the middle. If the mode was used, the high and low end of the results wouldn't even be seen.

Every student has different weights on a small object…. It is unfair if I use other numbers instead of the average of their weights. Also, these numbers may not be the actual weight of this object.

This way, each student's measurement is weighted the same.

Method III, because I would want the students to find the average of mean of the weights they found. That way instead of picking one of their findings, the class would be able to see the average weight from all of their findings.

Again, only two students (out of the ten that picked the mean) viewed the average as a signal in noise interpretation. However they failed to take the outlier into consideration.

Sometimes or most of the times the weight is off because of human error so avg out all the weight <u>should</u> come out to be more precise.

I would use the third method, for several reasons. There is always room for error when measuring an object. How are students supposed to know that 6.15 is the most accurate weighing? In using the third method, I would make sure that the same object would be used for each group. Just because one # may be more common than others, does not mean that is the most accurate, so I would not use the I method. You could have a tenth group who got 6.0, what would the students then do?

A deeper view of prospective teacher's reasoning is illustrated with Prospective Teacher C's comments during the interview. Prospective Teacher C chose to use the mean even though she recognized the effect of the outlier. Her thinking revels that she knows for a fact that the mean is sensitive to high values but when she looks at the distribution of the data, she is confused.

Int.: You said that you chose method III (the mean) because it is more accurate; what do you mean by that?

133

C: because it's clustered all around a close sort of range. I thought OK with the mean…you know…I thought the mean and the median might be close in this situation, but some times when you have a really, really high, like say you have a score of 8.5, I might want to say I want to ignore the 8.5 and do median instead. I say that the extreme score is going to weight the mean more than it would with the median.

Int.: and in this case…

C: In this case the mean and the median are going to be really close, because the data are all clustered together.

Int.: around…cluster around…

C: Well, you have the one… that's what brought me to think about it. Like I said, well, even though I say maybe we want to take that one into account, because sometimes you do…so I said, well it might be weird but maybe we should take him into account . So I thought maybe in this situation you want to. That's why I said the mean, I just didn't know. They are all clustered except for that one guy, that's what made me think, "Do I really want to take it into account or not?"

Summary and discussion

Prospective teachers are able to identify outliers in a data set but not necessarily to use their effect to choose the appropriate measure of center. Some actually want to take the outlier into account for representative reasons; others for pedagogical reasons. It is important to note that prospective teachers may be using this reasoning because the item asks "to determine as accurately as they can" and one of the choices is to "use the 6.15 since it is the most accurate weighing" given the indication that the task is about finding the most accurate measurement and not the most appropriate measure of center.

Garfield (2003) describes the results of a version of this item given to college students. Her version was multiple choice with the three methods given here plus a fourth method: compute the mean with the outlier thrown out.   She concluded that college students who do not choose the fourth method are ignoring the outlier. Due to the

134

design of the item used in the current study, however, we can see that the attraction of the mean (method III) appears to be stronger and more complicated than first thought. Results here indicate that prospective teachers may recognize the outlier but still choose the mean (method III).

Konold and Pollatsek (2002) give this item as an example of a context in which the average is interpreted as signal in noise, where the average is a close approximation of the actual weight and each observation is viewed as deviating from the actual weight by a measurement error. There was no evidence that prospective teachers that chose to use the mean after throwing out the outlier take this view. Only 2 out of the 19 mention that the average "would help reduce error". Others mention that the average will be the "most accurate way".

Item specification

Item 3 measures graph comprehension (histogram) at the level of interpretation. Part (a) measures the ability of the prospective teacher to recognize a student's misconception and to interpret the source of misunderstanding. Part (b) measures interpretation and judgment of a student's oral response. This item is a variation of a problem from Collecting, Representing, and Interpreting Data Module developed at the Center for Research in Mathematics and Science Education at San Diego State University.

The following graph gives information about the adult female literacy rates in Central and South American countries.

Frequency (y-axis: 1, 2, 3)

45 50 55 60 65 70 75 80 85 90 95 100

Adult Female Literacy Rate (%)

Adult Female Literacy Rates in Central and South America

a. Suppose you ask your students to tell you how many countries are represented in the graph. One student says, "there are 7 countries represented".

Is this student right or wrong? _____

In your opinion, what is the student's thinking to arrive to that conclusion?

b. Suppose now you ask your students to explain what the third bar from the right indicates. One says, "It indicates 85% to 90% literacy rate". Comment on the response.

Figure 5.3. Item 3

Table 5.4

Distribution of Scores for Item 3 (n = 42)

Item	Score				
	0	1	2	3	4
3a	1	6	9	3	23
3b	15	4	3	5	15

Description of the rubric for Item 3

For part 3a, a score of 4 was assigned to responses that stated that the student is wrong *and* identified the proper source of the mistake, counting bars. A score of 3 was assigned to responses that answer the question with "yes" but identified correctly the source of the mistake. A score of 2 was assigned to responses that stated that the student is wrong, identified the mistake of counting bars, and suggested the wrong representation of the bars. This is an example of a response with a score of 2, "The student is wrong. The student thinks that each bar in the histogram represents a country, not the frequency of adult female". A score of 1 was assigned to responses that either stated that the student was wrong and did not identified correctly the source of mistake *or* stated that the student was neither right nor wrong because there was way to tell the number of countries represented. Finally, a score of 0 was assigned to responses that stated that the student was right. This is an example of a response with score of 0, "The student is right, different bars shows different rates. Every country has its own rate. So bars could be countries."

For part 3b, ideal responses, which were assigned a score of 4, were expected to say that the explanation is incomplete and the student needs to add that there are three countries with the rate mentioned. A score of 3 was assigned to those responses that only

stated that the student needs to focus on the frequency as well, without providing

evidence of understanding of the label "Frequency". Responses that only mentioned that

the comment of the student is incomplete or mentions in general terms, like "the student

is only look[sic] at the information on the x-axis and not how the x and y axis relate."

received a score of 2. Less clear responses such as, "you are on the right track, but look

closely at what percent rate goes with what bar" received score of 1. Scores of zero were

given to those participants who said that the student's comment was correct *or* provided

an incorrect interpretation of the bar in the histogram.

Analysis

For part 3a, the about three fourths correctly identified that the student's statement

in part (a) was incorrect. Nine said that the graph did not provided enough information to

determine if the student was right or wrong, and only one said that the student was right

(see Table 5.4). Out the 32 prospective teachers who correctly said that the student was

wrong, 20 also identified the source of the mistake by saying that the student focused

only on the number of bars (see Table 5.5).

Table 5.5

Distribution of Responses to Item 3a by Recognition of Mistake and Its Source.

	Student is right	Student is wrong	Can't tell/Don't know	Total
Correct source of mistake	0	20	0	20
Incorrect source of mistake	1	12	9	22
Total	1	32	9	42

Some prospective teachers went further in their explanation and explained what the student is not focusing on, providing information of their own knowledge of graph comprehension. Some of these extended explanations were correct and some were not. Examples of correct explanations are:

Each bar actually represents a range which had at least 1 country reporting that score.

Student did not take into account that the height of these bars showed the frequency (or number) of countries within a specific range of % females who are literate. Had the students correctly read the graph, he or she would have noticed that 15 countries are actually represented.

The student is thinking that each bar represents 1 country, not that the height of the bar tells the frequency of the countries.

Examples of incorrect explanations varied and fit different categories. This type of responses was assigned a score of 2 or 1 for level of correctness. Some prospective teachers thought that the student is not "reading" the *x*-axis correctly.

The student looked at the graph and counted the bars, but either did not read the x-axis or doesn't have a grasp on reading graphs.

The student is looking at the bar graph and assuming that the x-axis represents seven different countries, when in actuality it does not. I can understand why a student might interpret the bar graph that way, because it is misleading.

Other prospective teachers thought that the student made the mistake because he or she did not understand what the bar represents providing an incorrect explanation for the bar representation. These prospective students did not provide a successful response for the second part of the problem, either. Examples of misconceptions about the representation of the graph fit mainly two categories. Some prospective teachers thought

that the bar represents the *percentage* of literacy rate in Central and South American

countries.

> "The bars represent the % of literate adult females in the central and …"

> "The student thinks that the adult female literacy rate(%) represents the different
> countries"

> "They counted the bars on the graph, not realizing it is the literacy rate (%)"

Others thought that each bar represent the *frequency* of adult female.

> The student thinks that each bar in the histogram represents a country, not the
> frequency of adult female literacy rate.

> Due to frequency, I think that means the number of women at the Adult Female
> Literacy Rate."

The nine prospective teachers who said that they could not tell whether the

student was right or not gave the following explanations:

> The graph is not explicitly detailed enough to be able to tell, depending on the
> grade level some students will say 7 countries.

> We don't know if the student is right or wrong. He is assuming that if there is
> some frequency of female literacy in every country it would register.

> Yes and no, yes there are 7 bars, but there may be stats that fall on a frequency of
> 0.

Results for part 3b are consistent with the previous part. About half of the

prospective teachers successfully recognized that the student had a partially correct

answer and correctly identify what was missing in her/his explanation (see Table 5.4).

Four of those who answered correctly showed some pedagogical knowledge.

> The third column from the right does indicate a 85% - 90% female literacy rate,
> but it's that in one country, more than one country, or the entire region? I would

then ask the student to examine and explain what this graph is representing to get him or her to answer my previous question.

I would ask 85% to 90% literacy rate <u>of what</u>? I would ask them open-ended questions to get them to see that it explains that 3 countries have a 85% to 90% literacy rate.

I would ask the student what the graph is telling us about that rate. What does the height of the bar indicate?

Seven responses received scores of 2 ($n = 3$) or 1 ($n = 4$) and 15 responses fall into the category of zero score. Four of the 15 said that the student was correct or the response was "an excellent response". Three thought that the student's comment was incomplete by neglecting to say that "the frequency of 3 tell us how many women had 85% to 90% literacy rate". The rest provided responses that gave evidence of no knowledge for teaching. Such as

It is the interval of the literacy rate and not specifically 85% or 90% literacy rate.

This response is vague, does not define what 'it' is that indicates 85%-90% literacy rate.

I would explain that since it is a histogram, the bars are connected, but in reality that piece of data ends for 85% it is just connected to 90.

In particular, when Prospective Teacher A was asked to elaborate on her written response about not knowing if the student was right or wrong for the first part of the item, she indicated her struggle to understand student's mistakes.

Int.: When you say that "we don't know" if the student is right or wrong, do you mean that we don't have enough information in the graph to say how many countries are represented?

A: We assumed that we have data for everybody, but apparently we didn't . Sometimes you have something missing, like I thought, OK there is quite a bit missing here [pointing at the gap between 50% and 70%] why do we lay out our histogram this way, and I thought, maybe there is something where there was no literacy rate…some country where there is zero.

Int.: That's why those gaps?

A: You never know…I though it was a strange histogram. Normally, you would just said, well you have a bar for everything, but you don't always know that, right? When you have a gap, you don't always know about your gap. This bothered me a little, this gap.

Int.: Why do you think the student said 7?

A: I don't know, because he wasn't really looking …well, he was only looking at one bar or two and then saying…you know what I mean. I couldn't figure…error with kids always bothers me…in math…I look at the error and I go…it is hard to know what they mean.

Furthermore, the prospective teachers seemed to have a difficult time correcting the mistake or guiding the student in the right direction when they themselves have other limitations. In this item, some prospective teachers were able to say that the student was wrong in saying that there were 7 countries represented and that the student was in fact counting bars. However, when asked in an interview how they would correct the mistake or guide the student to a correct answer, prospective teachers did not seem to understand the information given by the histogram.

Int.: You said the student is wrong because…

D: The graph said is for Central and South American countries, it does not say which ones.

Int.: But you do understand why the student said 7 bars?

D: The student is counting bars.

Int.: What would you tell the student to lead him to the right answer?

D: I would tell him to look at the height of the graph and remember always to follow the height of the bar to the left to see how tall it is and that is the frequency.

Int.: If you ask the student what does that mean, frequency 3, what would you want him to tell you?

D: The bar indicates 85% to 90% literacy rate for 3 women.

Discussion and summary

About 75% of the prospective teachers can recognize a typical student's misconception when reading a histogram and about 50% can also identify the reason why the student is making the mistake. However, that does not necessarily mean that the prospective teachers have other limitations with the histogram and are unable to correct the mistake properly. In other words, prospective teachers correctly make the judgment that the student is wrong but give the incorrect way to correct the mistake.

Results here show that prospective teachers misinterpret the representation of the bar in a histogram mainly because there is only one variable labeled "Adult Female Literacy Rate (%)" and they do not understand how to interpret the meaning "Frequency" in the content of the problem.

Item 4

Item specification

Item 4 measures knowledge of calculating the mean from a line plot, which implies assessing for knowledge of graph comprehension and the concept of the mean beyond the algorithm. It also measures properties of the mean and the ability to create distinct distributions with the same mean. Graphical displays were taken from materials developed by the *Connected Mathematics Project* (Lappan et al., 2002).

The following line plot shows the number of people in households in a neighborhood.
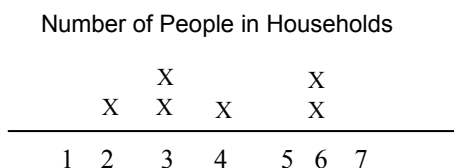
Number of People in Households

```
              X              X
        X   X   X          X
      _____
        1   2   3   4     5  6  7
```

a. Find the mean. Show how you find it.

b. Is it possible to have other sets of data with the same mean?
   Explain why or why not.

c. Is it possible to have a data set of six households with mean $3\frac{1}{2}$ people?
   If yes, give an example.
   If not, explain why.

Figure 5.4. Item 4

Table 5.6

Distribution of Scores for Item 4 (n = 42)

Item	Score				
	0	1	2	3	4
4a	3	0	4	7	28
4b	5	6	9	15	7
4c	10	0	4	4	24

Description of the rubric for Item 4

For part 4a, a score of 4 was assigned to responses that correctly stated that the mean was 4 people per household and provided a valid justification. A score of 3 was assigned to the responses that stated the correct mean but provided an incorrect equation or made a small computational mistake. A score of 2 was assigned to responses that either divided by 7 instead of 6 or did not give any justification. No score of 1 was assigned, and a score of 0 was assigned to responses that were blank or gave incorrect reasoning.

For part 4b, a score of 4 was assigned to responses that answered "Yes" to the question "Is it possible to have other sets of data with the same mean?" and provided a valid justification based on the algorithm connected to the concept of the mean as a balanced point of a distribution. A score of 3 was assigned to responses that stated "Yes" but only provided an example of another distribution with mean 4. A score of 2 was assigned to responses that said "Yes" but provided an unclear justification like, "there are many sets of data with the same mean". A score of 1 was given to even more unclear justifications, like, "Yes, it is possible" or, "Yes, as long as the average = 4". Finally, scores of 0 were assigned to responses of "No" or blanks.

For part 4c, a score of 4 was assigned to responses that stated "Yes" to the question "Is it possible to have a data set data set of six households with mean $3\frac{1}{2}$ people?" and provided a correct distribution as an example. A score of 3 was assigned to responses that stated "Yes" and in the attempt to create an example made a small mistake. A score of 2 was assigned to responses that stated "Yes" but did not provided an example or stated an unclear justification like, "because the mean is the average it does not have to

145

be whole number". A score of 0 was assigned to those responses that stated that is it not possible to have a mean of 3 ½ because of the fraction.

Analysis

For part 4a, which measures the ability of prospective teachers to "extract information" from the graph, the results are surprising. About two thirds of the prospective teachers correctly found the mean and wrote a valid mathematical equation to justify their answer (see Table 5.5). About 16% of the prospective teachers successfully identified the data values and the number of data points but made a notational error of equating the value of the sum to the ratio and therefore to the mean. An example of this mistake is shown Figure 5.5.

Show how you find it.
$$2+3+3+4+6+6 = 24 \div 6 = 4$$

Figure 5.5. Example of a Typical Notational Mistake from a Prospective Teachers' Work.

Only 2 students failed to actually divide the sum of the data value by the correct number of data points, 6. Instead these prospective teachers divided by 7. Note that there are 7 numerals on the axis of the plot. All of the prospective teachers found the mean using the mathematical algorithm. Only one person attempted to use the idea of center and unfortunately he used it incorrectly, saying, "using the idea of 'cluster' point associated with the mean, the mean is 3, for the data set above."

The most preferred method to compute the mean was to identify each data value, add them up and divide by the number of data points, reading the graph from left to right. For example, (2+3+3+4+6+6)/6 = 4. A quarter of the prospective teachers that successfully found the mean used a weighted mean approach by identifying the data value and its frequency. For example $\frac{2(1) + 3(2) + 4(1) + 6(2)}{6} = 4$. One person read the data values in horizontal "layers", (2+3+4+6+3+6)/6=4.

As expected, almost all of the prospective teachers could compute the mean of a given distribution. Interviews aimed to investigate if they could find or estimate the mean for a larger data set without reaching for the algorithm by providing either a larger data set where computation is not practical or asking them to find the mean by another method. The scenario used in the interviews was related to teaching. The interviewee was first presented with a new bigger data set (see Figure 5.6) and then asked how would a student find or estimate the mean without doing any computations.

```
       X                       X
       X                       X
       X         X             X
       X    X    X    X        X                   X
       X    X    X    X        X    X              X
       X    X    X    X    X   X    X    X         X         X
    _____
    26   27   28   29   30   31   32   33   34   35   36   37   38   39   40
```

Number of Raisins in a Box
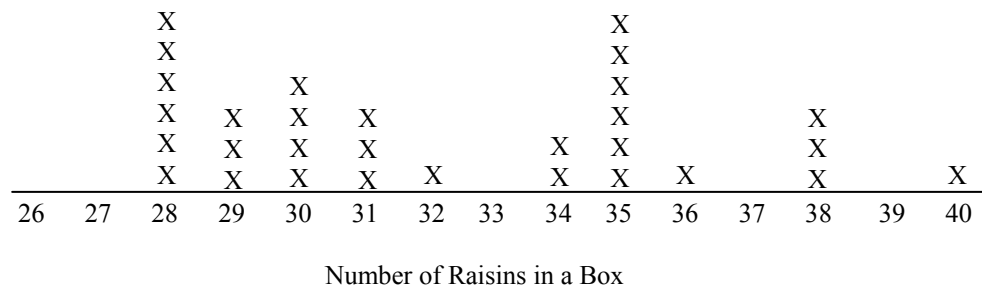
Figure 5.6. Data Set Used in Interview to Ask Prospective Teachers to Find the Mean without Computational Algorithm.

Prospective Teacher D had a method to estimate the median from a line plot but struggled to find a way to estimate the mean for a larger data set. In the intent to estimate

147

the mean, she "moves Xs around" and tried to relate the concept of balance but ended up

finding the average of the frequencies instead!

Int.: Suppose you are teaching the mean to your students and you want them to estimate the mean before they use the algorithm. How could they do it?

D: They could go…count one from each side [crossing out an X on each side]…they can take the average of the two left over and the answer will not be 4 but very close to it.

Int.: How about with a bigger data set?

D: Do the same…cross out one on each side, one, one, two, two,…[cross out one X on each side and ends up with two X's under the number 31], you have two 31's, then you would not have to divide by two. The answer is 31.
This is the median.

Int.: How about the mean? How can the students see the picture and estimate the mean?

D: It would not have to be exact.

Int.: No, but we want to find a good estimate.

[Long pause]

Int.: So the median would be easier to estimate for students than the mean?

D: That is what it seems to me…because I can't do it off top of my head…Maybe arrange the X's …like I would put this X down here and see how they average it out…that is the only way I can see how to do it. I would send them to the board and have them arrange the X's and make them line them up and that will be the average.
Move these over here, and erase these.
All the X's are even now, that would be three. Humm…[pause] I guess that is 3 raisins … hold on…[more thinking]…I am trying to figure out…
You could do adding up again and then divide. You add 28, 28, and 28 or 28 times 3.

Parts of 4a and 4c were designed to measure the ability of the prospective teacher

to explain why it is possible to have different distributions with the same mean and to

create distributions with a particular mean. About half of the prospective teachers gave a

successful response to Item 4b by providing a mathematical argument, or giving an

example of different distributions with equal means in the context of the problem (see

Table 5.6). However, only a third of those provided the ideal response of connecting the

algorithm with the concept of the mean as a balance point of a distribution. Following

are two examples of mathematical arguments given.

> Yes, because the average of numbers can be the same depending on what # you are dividing by and what #'s you are adding.

> All it takes is to have a numerator divisible by the denominator that comes out to be 4.

> Yes, as long as the total is 4x the number of responses.

About 12% of prospective teachers gave arguments in relation to the context of the

problem – people in households – to explain the possibility of having another data set

with the same mean. Here are some examples of these types of responses.

> Sure, the data could have been 20 people and 5 households, or 28 people and 7 households.

> Yes, the family that has 2 children and a family that has 3 children could move out of the neighborhood, then two new households could move in one with 1 person the other with 4 people.

> Yes, a house of 4 can be replaced by two houses of 2, and two houses of 3 can be replaced by one house of 6.

> Yes, you could. If you still only asked 6 households, you could still total 24 people. Data could be 2,3,3,4,5,7 = 24/6=4 or could ask less households have 12/3=4 or 32/8 if you asked more households. 4 has many multiples.

A third category of responses is one that talks about distribution, sample size, or

variation. However, these types of responses were too vague to be accepted as valid

arguments. Following are some examples.

> Yes, the distribution can be proportionately spread out.

Yes, you can have a smaller or larger sample size that may/may not add to 24 but still possible to have the same outcome.

Yes, other sets could be clumped together around the 4 or spread out.

Prospective Teacher E - who justified her answer for this part of the item by relying on the algorithm - was asked about this item during the interview.

Int.: How would you convince the children that is possible to have other data set with the same mean without reaching for the algorithm?

E: You can start with the same numbers and then switch them around , like if you have 4,4,4 take away one from the first 4 and make it into 3 and then put one more and make a 5. Just playing with numbers.

Int.: Can you do it with the bigger data set?

E: Let's see…[pause] I guess you should move… I don't how to do it with data, I just know how to do it with numbers. Because you can't move an X to 39 because this represents 3 of 38, isn't?

As for creating a data set with a specific mean (see Item 4c), about 70% of the prospective teachers were able to create a correct data set with the specified mean and number of data points. About half of those chose a symmetric distribution about the point 3.5, and the other half chose a non-symmetric distribution but with clear indication that they created it by finding the sum of the data values first (see Figure 5.7).
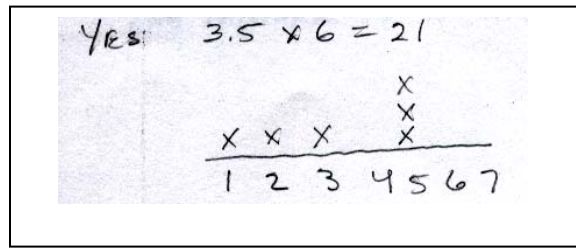
Figure 5.7. Example of a Procedure Used by a Prospective Teacher to Find a Distribution with Mean 3.5.

Finally, about 15% of prospective teachers argued that it was not possible to have six households with a mean of 3.5 people because when it comes to people, it does not make sense to have a non-integer value (see examples of responses below).

No because a household can not be represented by half point.

No! because the mean 3 ½ people is logically incorrect.

No, because there is no such thing as ½ people.

You can have a mean w/ 3.5, but not when it comes to people, because you can not have half of a person.

Prospective Teachers B and C were asked what it means to have a mean of 3.5 people per household and how they would explain it to a child. Prospective Teacher B seems to have a good way of explaining.

Int.: What does an average of 3.5 people per household mean?

B: That is the mean, you take all the numbers and add them all up and divide by the number you have.

Int.: Suppose you are explaining to a kid. Would the kid be bothered by the 3.5?

B: I don't think so, if you explain it in certain way.

Int.: Like what?

B: If you say, well nobody really has 4.5 kids, but over a big data set some people have 2 kids, some people have 3, some people have 6, some people have 8. So over a

big data set it could be 4.7, it could be 3.2, it just means over a big data set that is your average. I think kids understand that.

You have to say, well, you take an average and the average means that that matters, that every one matters, that's why you would get 4.7 kids. It just means that people are all over the range, some people have 1, some people have 2, 3, 4, 5, 10.

Int.: Now, I am going to play the kid. The kid would say "why doesn't the answer come out to be a whole number"?

B: I think you have to teach them about the median, maybe, that there is another number that could be a whole number, but not always. Even the median could be 4.5, but…

In contrast Prospective Teacher C was unable to explain the meaning of a mean of 3.5 people per household.

Int.: What does and average of 3.5 people per household mean? How do we interpret this number?

C: I guess when the average…many many houses average about 3.5. I don't know what that means; I don't know how to explain what that means though. Because the kids are gonna go "what is 3.5 of a person?!"

Summary and discussion

About half of prospective teachers are able to compute and estimate the mean of a small data set represented in a line plot. When given a bigger data set and asked to estimate the mean, they either estimate the median without observing the distribution or in their attempt to find the mean; they find the average of the frequencies instead. The latter seems to be a reflection of the methodology taught in elementary and middle school of leveling off stacks of cubes.

About half of prospective teachers know that it is possible to have many sets of data with the same mean. However, only a third of those can justify it with an argument

152

that relies on both the algorithm and the concept of the mean as a balance point. About two thirds of prospective teachers are able to create a distribution with a specific mean that is not a whole number.

More than half of prospective teachers interviewed and probed for knowledge for teaching did not show sufficient evidence of this type of knowledge. Most could not estimate the mean of a data set without reaching for the algorithm, or describe how to convince a child that there could be several data sets with the same mean, or explain what a average of 3.5 people means.

Item 5

Item specification

Item 5 measures knowledge about formulating questions to generate data and measures of center and range for categorical data. Furthermore, this question measures the ability of prospective teachers to identify errors in students' responses. The data and suggestions for student responses were taken from materials developed by the Connected Mathematics Project (Lappan et al., 2002).

One middle school class generated data about their pets shown below.

Pet	Frequency
bird	2
cat	4
cow	2
dog	7
duck	1
fish	2
goat	1
horse	3
rabbit	3

a. Give a possible question the teacher could have asked the students to generate the data.

b. Students were talking about the data and one said:
" The mode is dogs, the median is duck, and the range is 1 to 7."

If you think the student is right, explain why.
If you think the student is wrong, identify the mistake(s).

Figure 5.8. Item 5

Table 5.7

Distribution of Scores for Item 5 (n = 42)

Item	Score				
	0	1	2	3	4
5a	14	0	0	7	21
5b	18	21	2	1	0

154

Description of the rubric for Item 5

For part 5a, ideal responses that received a score of 4 were those that correctly stated a question that could have generated the data. A score of 3 was assigned to responses that did not state the answer in a question format but provided a procedure or way to generate the data. No scores of 2 or 1 were assigned. A score of 0 was assigned to responses that stated a question that can be answered about the pets, such as, "What pet does the majority of the students have?" or, "How many pets are there in all?"

For part 5b, the ideal response was to state that the student is only right when he or she says that the mode is "dog". The median and range do not apply to categorical data such as type of pets. A score of 3 was assigned to responses that came close to the ideal responses by giving some indication that the median and range did not apply in this case, but made some incorrect statement. A score of 2 was assigned to responses that stated that the mode and the range were correct, but the median was not because of the type of data. A score of 1 was assigned to responses that stated that the student was right saying that the mode was "dog". A score of 0 was assigned to responses that did not make any correct judgment about the student's statement.

Analysis

For part 5a, half of prospective teachers were able to create a correct question that teacher could have asked the students to generate the data. Seven out of 42 did not write their response in a question format, but indicated a way to create the given Table 5.. Of the remaining 14 responses, 12 stated a question that could be answered using pets and 2 did not provide any response.

The typical question created by prospective teachers that could be asked to generate the data was of the following kind.

"Everyone tell me what kind of pets you have?"

"What type of pets do you own and how many?"

As before, some prospective teachers identify with the teaching context and provide responses with some pedagogy related. Here are some examples.

List all the different types of pets and have the students come up to the board and put a tally next to the pets in their home.

Make the sound of your animal (that you own) and get a group by making those sounds. Now select a leader to count the number of kids in the group! How many of you own one of these animals. (Provide chart).

The second part of this item turned out to be the most difficult question of the assessment instrument. No one responded correctly that the student is right when he talks about the mode, but for categorical data the definition of measures of center and spread do not apply in the same way as they do for numerical data. Only one person got close to a correct response (score of 3) by saying:

The data is showing how many people have each type of pet. When we refer to measures of central tendency, we are using numerical data. … In this case dog is the most frequent choice, but it has no numerical value….We cannot order non-numerical data. …There are no data values to subtract, you do not use frequency to find the range, and if you did the range would be 7-1 = 6.

Two others showed some awareness of the type of data and mentioned something about not making sense of measures of center in this case. These two responses fall into the category of unsuccessful with the score of 2.

It is difficult to arrange non-numerical data into the same type of asending/desending[sic] order that numbers can be arranged into. 'Dog' is the most frequent response, so I might agree that it could be considered the 'mode'. However, 'duck', although it falls in the middle of the alphabetic listing, can

hardly be considered the median (since 9 answers are past it in the list and 15 are before it). Alphabetically, the range would be 'bird to rabbit'.

The student is right on the mode & the range but wrong on the median part. Median can't be duck. It has to be a number. Median is the average # of animals in each type of pet that a one middle school class should have.

The rest of the responses split into those that recognized that the student is correct in saying that the mode is "dog", but were incorrect about of the judgment of the rest of the student's statement (21 out of 42) and those that incorrectly judged the entire statement, receiving a score of zero (18 out of 42).

The typical response for the 21 that received a score of one is that the student is wrong about the median being "duck" because the student is just finding the middle animal in the list and he/she needs to took into account the frequency as well. Furthermore, prospective teachers take the frequency into account in many different ways, coming out with at least five different strategies to find the median.
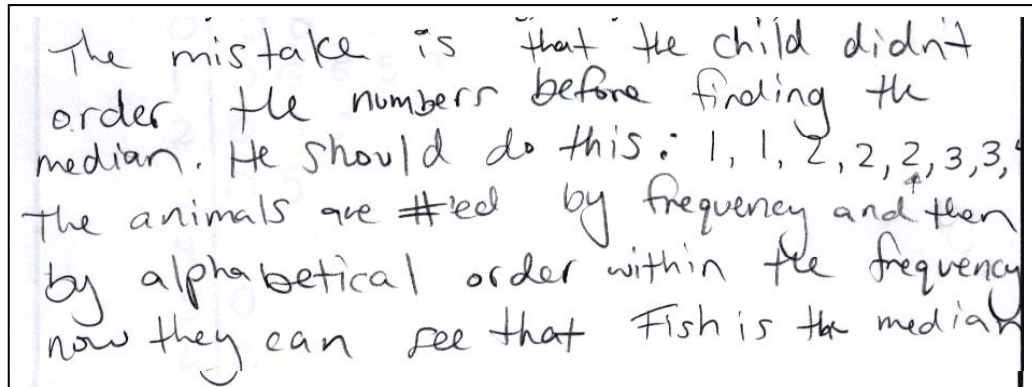
One strategy (4 responses) is to find the median by ordering the frequency values and picking the middle value. This is illustrated by the following response:

> *"The mode is dogs and the range is 1 to 7. However, the median is*
>
> *2....112223347 "*

The second strategy (2 responses), is not to pick the middle value of the frequencies but to pick the corresponding pet the goes with the middle frequency. As shown in Figure 5.5, the prospective teacher picks "cow" as the median.

A third strategy (1 response) is similar to the previous one. However, instead of picking a pet arbitrarily within the frequencies, the names of the pets are listed in alphabetical order. The median would then be the pet that corresponded to the median of

the frequency values (see Figure 5.6). When this prospective teacher was asked in the interview "you said that the median is 'fish', what does this tell you about the pets?" She responded, "that the average student in the class have fish [sic]".

Figure 5.9. Example of a Prospective Teacher's Strategy to Find the Median of Categorical Data.

A fourth strategy (3 responses) is to order the frequency values, not with the purpose of finding the median of these values, but with the purpose of making a list of the 25 pets being considered: {duck, goat, bird, bird, cow, cow,…, rabbit, cat, cat, cat, cat, dog, …, dog}. The median is the13th pet in the list, namely "rabbit". As one student put it "…the student incorrectly stated that the median is duck, because he or she failed to place the data in order of increasing frequency first. Once the data is arrange [sic] in the correct order, it is clear that the median is rabbit."

Finally, the fifth strategy (6 responses) is a modification of the fourth strategy. Instead of ordering the list of pets by its frequency, the prospective teachers decided to keep the list in alphabetical order: {bird, bird, cat, cat, cat, cat, cow, …, dog,…, rabbit, rabbit, rabbit}. Then, they picked the 13th pet, which in this case is "dog". In their own words "…median is incorrect. It should be the pet that is listed in the middle of the 25 pets, which would be dog."

One of the students who used the fifth strategy argued that this is the way to order the animals "…because when finding the median, you're suppose to put it in order from smallest to largest and when it comes to animals how to judge which is larger or smaller. There are dogs smaller than ducks, ducks smaller than cat and vice versa."

With respect to the statement about the range being 1 to 7, 27 out of 42 participants said that the statement was correct. The typical explanation was that the minimum number of pets is 1 and the maximum is 7, thinking more in terms of the range of the frequency values. Another group of participants (5 out of 42) said that the student was wrong because he or she should say that the range is the <u>difference</u> between the largest and the smallest number, in this case it would be $7 - 1 = 6$.

Only one prospective teacher gave evidence that the range should be consistent with the other measurers in terms of non-numerical value and said "…alphabetically, the range would be 'bird to rabbit'".

Summary and discussion

Almost all prospective teachers can create or formulate questions to generate data. However, most can not recognize students' mistakes about median or range for categorical data. Most prospective teachers could not describe what the measures of center of spread tell about the data when trying to find them. Instead they use a procedural method to get an answer. In this particular case, prospective teachers were very creative inventing their own methods to have a median for categorical data.

Item 6

Item specification

Item 6 measures the prospective teachers' ability to select and create a suitable representation for a data set, taking into account the shape of the data.. The data for this item was taken from materials developed by the Connected Mathematics Project (Lappan et al, 2002).

Student's initials	Time in minutes
DB	60
DD	50
SE	35
AE	30
FH	25
CL	25
DR	22
BN	20
VH	20
IW	17
AS	15
KS	15
VC	15
AS	10
MS	8
RS	5

A middle school class was wondering how much time it took each student to travel to school each morning and collected the following data.
Students are asked to make a graphical display of the data to show how much time the majority of students take to travel to school.

a. Which one of these plots seems the most appropriate for the data?

☐ Histogram   ☐ Stem-and-leaf Plot

☐ Bar Graph   ☐   Other _____

Explain your choice.
b. Make the plot picked in part a.

Figure 5.10. Item 6

Table 5.8

Distribution of scores for Item 6 (n = 42)

Item	Score				
	0	1	2	3	4
6a	13	6	7	12	4
6b	7	0	5	2	28

Description of rubric for Item 6

Correct responses for the first part of the item, 6a, were the choices of stem-and-leaf plot or histogram together with a correct explanation. Prospective teachers were expected to justify their choice by mentioning that these two representations would show the shape of the data and could answer the question in consideration. In addition, if they chose steam-and-leaf plot, they were expected to say that this plot shows all the data values. These responses were assigned a score of 4.

Still considered successful responses, but with a score of 3, were those responses that chose stem-and-leaf or histogram for practical reasons related to the type of data such as its range, frequency of 20s and teens, etc. These responses did not attend to the question needing to be answered – how much time the majority of students take to travel to school?

Unsuccessful responses with scores of 2 or 1 were those that chose the appropriate plots but provide a non-statistical reason such as – it is the quickest, it is less confusing, it would suit it better, it would give the best visual, etc. Scores of 0 were left for those that chose a bar graph with incorrect reasoning or provided no response.

For part 6b, a score of 4 was assigned to a correctly made graph that was consistent with the choice in part 6a. A score of 3 was assigned to graphs with minor mistakes, such as miscalculating the frequency of data values within an interval when constructing a histogram. A score of 2 was assigned to graphs that contained major mistakes, such as not showing the stem of 40s because there were no values in the 40s in the data given. No scores of 1 were assigned and a score of 0 was assigned to blank

responses or responses that showed a graph that was inconsistent with the choice in part 6a.

Analysis

Results show that almost 40% of prospective teachers provided successful responses. Approximately 30% scored 2 or 1 and another 30% scored 0.

A typical reasoning prospective teachers stated for choosing a bar graph is that they thought that it was important to pick a graph that would show the time with the corresponding student, claiming that this is more powerful than any other choice. One used a different reasoning.

> I would use a line graph, only because there are two pieces of info given to the student: initials and time. If there were only one piece available to the student then I would use a stat plot, or a histogram.

The second part of the item, 6b, had better results. Almost 75% of prospective teachers were able to create the graph they chose in part 6a, 28 correctly and 2 with minor mistakes. Five made a major mistake when constructing the graph. The typical mistake when constructing the stem-and-leaf was to omit the stem corresponding to the 40s since no values are represented in the table (see Figure 5.11).
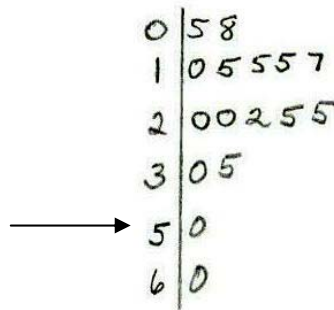
Figure 5.11. Example of Typical Mistake when Constructing a Stem-and-leaf Plot.

Seven responses received a score of zero, four of which were inconsistent with the choice in part 6a. Participants that chose a histogram made a bar graph instead. Two were blank responses.

In the interview, prospective teachers were also asked why they did not pick the alternative choices. Most of them based their choices on non-statistical reasoning as the following vignette shows.

Int.: Tell me why did you pick the steam-and-leaf plot?

A: I thought the histogram is easier, but I always do the steam-and-leaf plot before I do the histogram because I want to see how does [sic] the data actually clusters. Cause there are so many in two [points to the numbers in the 20s] and then I might… I would have to split my groupings slightly so I wouldn't have so many in just one part of the histogram.

Int.: How about in terms of the data? Would you do the same for another set of data?

A: It depends on the students' ability level, but you can do a histogram.

Int.: What can you tell me about the bar graph?

A: Well, time is continuous in general, it's a measure I suppose as discrete data. I don't know, you could do it I suppose. I wouldn't tell a child no you can't ever try this. They can try it and see how it works, right?

Int.: Can you answer the question with a histogram?

C: No, you can do it with the histogram…it would look good. You could actually visualize the frequency a little bit better. But I thought this would be easier for a student to draw on its own, because with a histogram you have to find the frequency that you are going use between each number and there is a lot more stuff you have to do.

Discussion and summary

Prospective teachers were able to create graphical representation of data, however they often based their choice on practical reasons for construction. There was very little

163

evidence that prospective teacher took the question needing to be answered into account for the choice of the graph.

The most common choice was a bar graph showing the initial of the students on the *x*-axis. It seems important to them to show the correspondence between the time and each student. This is an interesting point since statisticians understand data more as distributions rather than as representing separate pieces of information.

Item specification

Item 7 was designed to measure the level of interpretation of a non-standard data representation and the ability to examine and make judgments about student work based on statistical reasoning. It also measures the ability to describe students' thinking and infer about their understanding. This item was adapted from an item shared in a personal communication from Susan Jo Russell (July 16, 2001).

Imagine that two second-grade students in the same class have created the following representations to show the number of teeth lost by their classmates.

Student 1:   Student 2:

a. Compare and contrast the two representations. That is, how are they alike? How are they different?
b. What do you think each student understands about the data?

Figure 5.12. Item 7

Table 5.9

Distribution of scores for Item 7 (n = 42)

Item	Score				
	0	1	2	3	4
7a	7	7	23	4	1
7b	18	9	11	4	0

Description of the rubric for Item 7

For the first part of the item, 7a, where participants were asked to compare and contrast the two representation, a response with a score of 4 or 3 was expected to mention, or give some indication, that the representations are alike and different in the sense that a bar graph and a line plot or histogram are different when used to represent the same data.

An unsuccessful response, with a score of 2, was identified as a response that compared and contrasted the representation in terms of physical construction and gave indication of graph comprehension. An example of this type of response is

> Student 1 has represent[sic] each student with a picture placed beside the number of teeth he/she has lost, starting with the least and moving to the most. Student 2 represents each student with a vertical stack of blocks, each block representing one tooth lost. The representations are similar in that they are very visual and did not require a lot of explanation.

Responses that compared the representation in terms of superficial characteristics yet with some indication of graph comprehension received a score of 1. An example of such a response is the following:

> "Both students show the same information. 19 students were polled and both charts show that the same amount of teeth were lost by the 19 students. Student 1 did a better job of labeling and using creativity."

166

A score of zero was left for blank responses and responses of that focused solely on superficial characteristics.

> "alike: represent same form of data (one is horizontal and the other vertical), different: no labels and not quite clear"

For part 7b, the ideal response was expected to state, or give some indication, that Student 1 has grouped or reduced the raw data represented by Student 2 and that student 1's representation enables him/her to answer different questions about the number of teeth than Student 2's representation.

A score of 3 was assigned to responses close to the ideal response by either mentioning that Student 1 has a picture of the distribution of number of teeth or by mentioning that Student 1 is grouping or reducing the data. A score of 2 or 1 was assigned to responses that stated some evidence of understanding at least what each student is focusing on. Following are two examples of this type of responses.

> Student 1 understands that for each amount of teeth lost, a picture could represent the number of children who lost those teeth. Student 2 understands that every child who lost teeth should get their own stack to represent them.

> They both understand how to model the situation. Yet Student 1 represented # of people that lost teeth and how many vs. where[sic] the Student 2 represented # of teeth lost by the whole class, if you count the squares individually.

Finally, a score of 0 was assigned to blank responses or incorrect reasoning.

Analysis

For part 7a, only one prospective teacher provided the correct ideal response and 4 participants got close to it. A little more than half scored 2, which means that they were able to at least describe the differences of the graphs in a physical way.

167

The second part of the item, 7b, was even more difficult than the first. Prospective teachers had a hard time talking about students' understanding about the data. No one provided an ideal response. About half of the participants scored 2 or 1 which means that they mentioned or gave some indication of understanding of student's thinking.

More than 40% of the sample scored 0. Seven were blank responses. From the eleven that answered the question, some categories of responses can be identified. About half of them think that both students had the same understanding about the data but represented it differently.

> "Frequency is represented in both graphs, I think they both understand and show that."

> "Same but different perspectives."

> "Both understand what to do, however, they choose to express it in different way."

A couple of other prospective teachers provided answers to questions that the graphical representation may answer, such as

> "more students have lost 8 teeth than any other group."

> "Well, there was a total of 115 teeth lost by their classmates."

Finally, there were a couple of prospective teachers that expressed that they could not comment on students' understanding because of a lack of explanation:

> "Depends on how it is interpreted. There is no explanation so I have no idea about what each student is thinking."

> "how to count, student 1 realizes to separate but student 2 is going somewhere and I don't know where."

When interviewed, prospective teachers were asked to think statistically, that is, to think about how the students' work differed in a statistically sense. Prospective teacher A thought that Student 1's representation was like a pictograph and Student 2's was like a bar graph. Although Student 1 has pictures of children, the representation does not resemble a pictograph but a line plot. Towards the end of the conversation, she noted that Student 1 reduced or grouped the data. This was what was expected for an ideal response.

Int.: You described how these representations are different and understand what each one tells you about the teeth lost. Now, I want you to think statistically. This one resembles what type of graph? [Pointing to Student 1's]

A: A pictograph

Int.: And this one? [Pointing to Student 2's]

A: like a bar graph

Int.: Now think how are they different

A: They are very similar, if you put this one sideways [pointing at Student 1's] it is also a bar graph, the different[sic] the pictograph is that there is only one number, but then you can have as many pictures as that many number kids are represented and then… but in the graph bar every student is represented…where… will be a lot more space …this is… I personally will do a pictograph. You can see that 3 people had 7 then for the bar graph type you have to count each one up, where the numbers are right there for you.

Int.: How about in terms of questions each one answers? For example, what is the total number of teeth lost?

A: I think the pictograph because you can see 2 people in the 2s, so that is 2 times 2, and 2 people in the 3s, so 2 times 3, and so on. While this one you have to add up because it is not laid out for you.

Int.: What is the most common number of teeth lost?

A: The pictograph shows it a little bit better, you can see it in the bar graph, it is very close to the seventh category it is not as visual as this one.

Int.: If you were assessing these two students understanding about the data, what would you say? Remember to think statistically.

A: The pictograph because that involves more knowledge to group them and make it more readable whereas this one they have the knowledge of how to make the graph and all the information they need to put in it, but they do not understand how to put it together yet and so they just laid it all up.

Differently than Prospective Teacher A, Prospective Teacher C judges students' understanding about the data in terms of their communication and presentation.

Int.: If you were assessing these two students understanding about the data, what would you say? Remember to think statistically.

C: I would actually say Student 1 understands better. Student 2 does not have anything that tells me what it is. If they have told me a key, or if they have told me what this axes represents here and what this represents I would have an easier time to…but they organize it beautifully. I like this second grade work, it shows that they did it themselves or in a little group and they really thought about it.

Summary and Discussion

Analyzing and judging students' own data representation in terms of statistical reasoning was extremely hard for prospective teachers. When asked to compare and contrast their work, most focused only on the procedural and physical differences of the graph representations. Since both students' work are representations of the data, prospective teachers do not perceive any difference in students' understanding of the data claiming that "they both show the same information in different way". Part of the difficulty of this item is the fact that prospective teachers are asked to infer about students' statistical thinking for data representation that are non-standard.

Item specification

Item 8 measures in part 8a computational knowledge of the range and standard deviation from a line plot of two data sets. Part 8b measures the recognition of the two data sets' shape in relation to the measures calculated in part 8a. This item is based on an example from *Functions, Statistics, and Trigonometry* (Senk et al., 1998).

Consider the dot frequency distributions below with the heights in inches of the 10 players on two basketball teams.

a. The mean of each distribution is 75. Calculate the following for each distribution.

	Team A	Team B
Mean	75	75
i)Range		
ii)Standard deviation		

b. Explain how you could have determined which distribution has the largest standard deviation, without calculating.
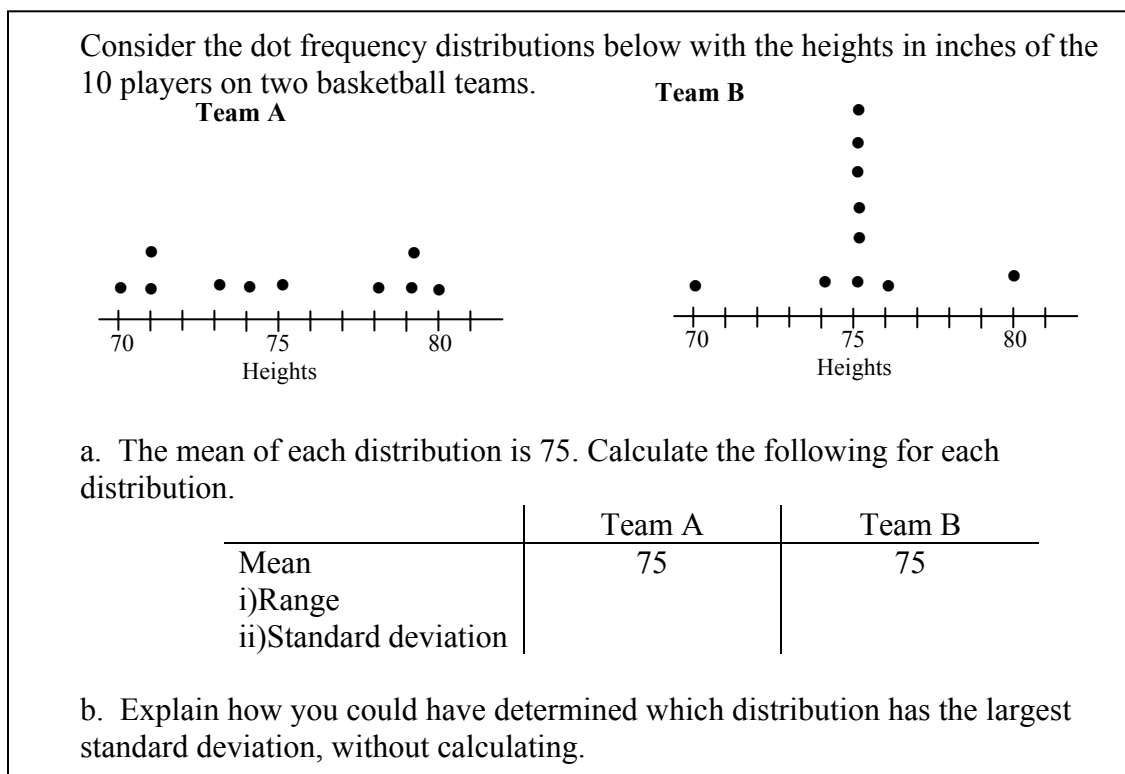
Figure 5.13. Item 8

Table 5.10

Distribution of scores for Item 8 (n = 42)

	Score				
Item	0	1	2	3	4
8ai	2	0	0	25	15
8aii	19	3	7	3	10
8b	21	3	3	7	8

Description of the rubric for Item 8

 In Item 8a, computations for the range and standard deviations were scored

separately. For the range, a score of 4 was assigned to responses that computed the range

as a single measurement value of 10. A score of 3 was assigned to those that reported the

range as "70-80" or "70 to 80". No scores of 2 or 1 were assigned for this item and a

zero score was assigned for incorrect computations.

 Scoring the computational knowledge of the standard deviation, took the whole

range of the scoring rubric. Again, a score of 4 was assigned to correct calculations using

either formula, $s = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n-1}}$ or $\sigma = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n}}$. A score of 3 was assigned to

those responses that made a small computational mistake when dividing by n or $n - 1$, *or*

the value was incorrect in the tenths or hundredths digit. A score of 2 was assigned to

those computations with a major mistake such as not dividing by n or $n - 1$ *or* computing

the mean average deviation (MAD) *or* not giving a numerical answer by indicating that

Team B has a smaller standard deviation than Team A. A score of 1 was assigned to

responses that made a mistake in the early stages of the procedure to compute the

standard deviation. Scores of 1 were also given to responses that indicated that all they

remember about the standard deviation was the fact that for normal distributions 68% of

the data values are within one standard deviation. Finally, responses that did not

provided any or incorrect values were given scores of 0.

 Item 8b was designed to measure the understanding of the measures of center and

spread in relation to the shape or distribution of the data, that is, the application of the

measures of center and spread as a tool to compare two distributions. Scores of 4 were

those that correctly communicated the meaning of the standard deviation applied to the

context of the problem. An example of this type of response was "Team A has a more spread out height distribution, whereas Team B had more players closer to the mean." Responses that did not mention that the spread was about the mean *or* did not refer to the context of the problem received a score of 3. For example, "The standard deviation told me that graph A is more spread out than graph B". Score of 2 and 1 were assigned to responses that indicated partial understanding of what the measures calculated in the first part say about the distribution but either did not communicate clearly *or* part of the response was incorrect. Responses with scores of 0 were those that did not provide any answer *or* all of the statements were incorrect.

Analysis

Results for Item 8ai show that more than 90% of prospective teachers have computational knowledge of range. Out of the 40 who were successful on this item, 15 gave the answer as a single measurement and 25 reported the range as an interval.

About a quarter of the prospective teachers correctly found the standard deviation for both data sets (five used calculators and five did it by hand). Three of them made a small mistake. Thus, 13 out of 42 (31%) successfully responded to this part of the item. No patterns of major computational mistakes were identified here, and the alternative algorithm used for the standard deviation was the mean absolute deviation (MAD). However, the subjects made no indication of the fact that this was a different measurement.

About 45% of the participants scored 0. Six participants left the question blank and the typical response for those that provided an answer was an equal value of 5

173

standard deviation units for each distribution. No computational procedure was shown

by any of the participants that arrived at this answer. However, one of them indicated in

the second part of the problem that "The students' height range in total of 10 in. from

shortest to tallest with 75 being in the middle so 5 inches difference up or down."

During the interviews prospective teachers that successfully computed the range

and the standard deviation were asked whether their answers made sense and why. They

were also asked to say what the number they computed means in terms of the data. Their

answers reveal different levels of understanding about this measure of spread.

Prospective Teacher A

> *Int.*: Do your answers make sense?

> A: It does, this stuff is all scattered so it is a higher standard deviation. This stuff, you only have two aberrations and the rest is all clustered right at the mean, so it makes sense that one would be smaller than the other.

> *Int.*: What does 2.14 means?

> A: 2 units, it means that from your mean one standard deviation takes into account 68% of your data and two standard deviations take a 95%, that's all it means.

Prospective Teacher C

> *Int.:* Do the answers make sense?

> C: Yes, I think my answers are right.

> *Int.:* Because…what does the standard deviation tell you?

> C: It is the distance away from the average.

> *Int.:* Does it make sense that this one [Team B's] is bigger?

> C: Huumm, yeah because these are more the same…and these come down and out, I guess, so that it become…I am not…I want to say that make sense. But I would actually do it in paper to see if it is correct.

Int.: So, you are saying that number represent how far each data is away from the mean. So 3.57 would mean that…

C: That the average will go ???….that far away.

Prospective Teacher B

Int.: Do the numbers make sense?

B: The standard deviation is from the lowest number to the highest number how much they are changing.

Int.: So, does it make sense that Team A has a higher standard deviation?

B: Probably not to a child, Team A is more spread out but Team B is way up there towards the middle.

Prospective Teacher D

Int.: Do your calculations make sense?

D: For team B we see that it is 2.28, that means that the standard deviation is smaller than team A which is 3.58. That makes sense because none of the answers are closer together for team B given a smaller deviation and further apart make a larger deviation.

Int.: What are the units for the standard deviation? What does 2.28 means?

D: Is it inches? no…ah..I don't know what that means.
How far apart, how frequently the grades occur?
….

More than a third of prospective teachers scored 3 or 4 on this item, giving indication of understanding of the meaning of what the standard deviation measures. Eight of the fifteen successful subjects mentioned that the variation is about the mean and the rest only mentioned the variation. Nine of the fifteen successful responses for this item also provided a successful response for the calculation of the standard calculation. That is, nine out of 42 prospective teacher were able to compute *and* make sense of the

measurement. Six either made a major mistake on the computation or computed the MAD but still understood the meaning of the standard deviation.

For this item, responses were more dichotomous. Very few participants (6) were assigned scores of 2 or 1. They either knew about what the standard deviation tells us about the data or they did not. Consequently, this is the item among all the items in the instrument that has the highest percent of 0 scores (21 out of 42). The majority of these (17/21) could not compute the standard deviation, either. The remaining four subjects showed computational knowledge but incorrect statistical reasoning. Some of the misconceptions are illustrated in the following responses:

> Because the standard deviation of Team B is less than that of Team A, I can conclude that team B represents a more normal distribution than does team A. The less the standard deviation the more normal shaped distribution of data because values are located closer to the mean of the data.

> "Standard deviation indicates the + or – away from the mean and the percentage of subjects in such deviations eg: ± standard deviation (approximately 17%)"

> "that number represent how far each individual is away form the average height"

The results are summarized on Table 5.11. Successful responses for this item correspond to conceptual understanding about the standard deviation; unsuccessful responses with score of 2 or 1 corresponds to partial understanding; and scores of 0 corresponds to no conceptual understanding.

Table 5.11

Distribution of Responses of Computational Skills and Type of Responses for the Standard Deviation.

	Correct computation	Incorrect computation	Total
Successful responses (score 3 or 4)	9	6	15
Unsuccessful responses (score 2 or 1)	1	5	6
Unsuccessful responses (score 0)	4	17	21
Total	14	28	42

Discussion and summary

Prospective teachers who have computational knowledge of the range and the standard deviation do not necessarily have understanding of the concept. However, results here show that the number of prospective teachers who have computational knowledge and poor or none conceptual knowledge is lower than the number of those who actually understand or partially understand the concept but do not have the computational knowledge. That is, there are more prospective teachers that show understanding without knowing how to compute than prospective teacher that can compute without understanding.

Overall Performance

Prospective teachers' overall performance on the written instrument was obtained by scoring each question (some items had more than one question) on a scale of 0 to 4. The total number of questions was 18, making a total of 72 possible points. Totals for each participant were converted to a percentage. Descriptive statistics across all 42 percentage scores were computed. There were four prospective teachers that did no report any background in statistics at the college level, however their performance in the assessment was about or above average. Therefore, they were not excluded from the analysis. Figure 5.14 shows the distribution of scores; the distribution is slightly skewed left, with mean 58.86%, median 60.42%, and standard deviation of 16.21%. None of the participants showed mastery or near mastery level (score 4 or 3 from rubric) on all the questions of the instrument.
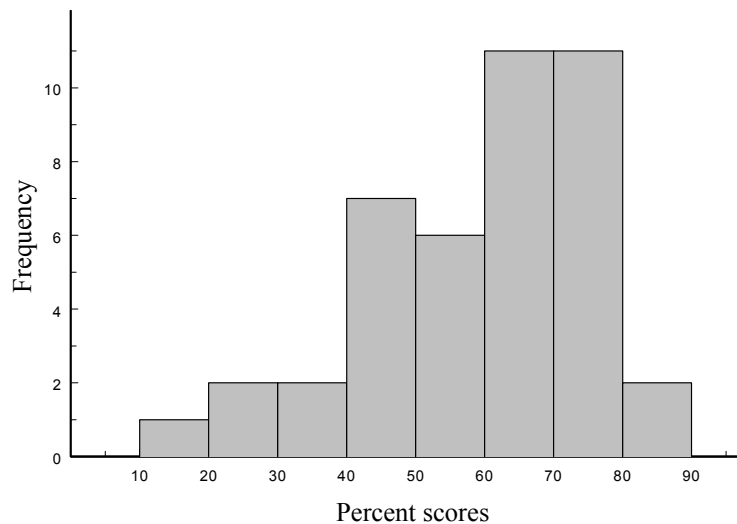
Figure 5.14. Distribution of Percentage Scores of 42 Prospective Teachers.

178

The lowest score was 19% and the highest 83%. The distribution does not show large spread about the mean or gaps between scores. Half of the scores (22) fall between 60% and 80% and the majority fall between 50% and 80%.

Performance by Domain of Knowledge

The assessment instrument was designed to measure two main domains, statistical knowledge and statistical knowledge for teaching. The domain of statistical knowledge was measured with 12 (out of 18) questions from the assessment, making a total of 48 possible points for that part. The 12 questions are Items 1a, 1b, 1c, 2, 4a, 4b, 4c, 6a, 6b, 8ai, 8aii, 8b. The domain of knowledge for teaching was measured with 6 (out of 18) questions, making a total of 24 possible points. The 6 questions are Items 3a, 3b, 5a, 5b, 7a, and 7b. Scores were added for each part and converted to percentages.

Prospective teachers performed better in the domain of statistical knowledge than the domain where they had to apply this knowledge to teaching with means 65.7% and 45.1%, respectively (see Figure 5.15). Furthermore, scores for the domain of knowledge for teaching are more spread out about the mean than the scores in the statistical knowledge domain (see Table 5.12). Again, none of the prospective teachers showed mastery or near mastery level of correctness in either domain of knowledge considered here (see Appendix D for complete database).

Table 5.12

Descriptive Statistics of Percentage Scores by Domain of Knowledge

	Mean(%)	Median(%)	SD(%)
Statistical knowledge	65.7	67.7	17.6
Statistical knowledge for teaching	45.1	45.8	18.1
Overall	58.9	60.4	16.2
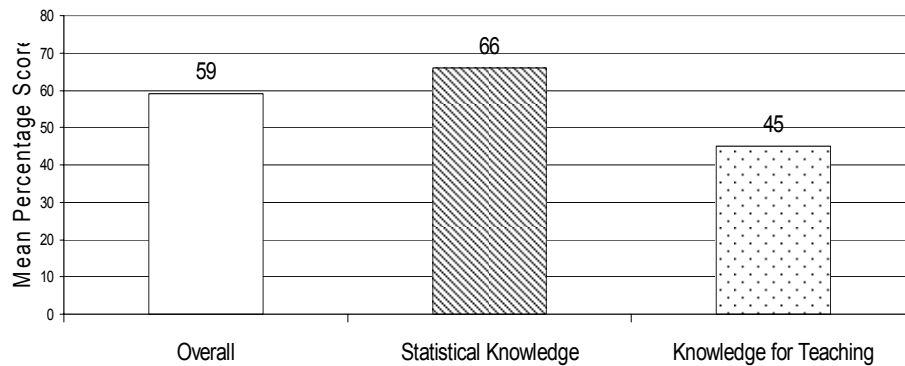
Figure 5.15. Mean Percentage Scores by Domain of Knowledge and Overall.

Distributions of scores for separate domains of knowledge show a better picture
of how these compare. Figure 5.16 and Figure 5.17 shows that the distribution of
statistical knowledge has more of a bell shape than the distribution of knowledge for
teaching. The distribution for the knowledge for teaching looks pretty uniform for scores
between 40 and 70%, while the distribution for statistical knowledge is skewed left.
About 76% of the prospective teachers scored 60% or *better* on the statistical knowledge
part of the assessment, while about 76% of the participants scored 60% or *less* on the
knowledge applied to teaching.

Another major difference between the two distributions is the percentage of
prospective teachers that score below 30%. For statistical knowledge about 7% of the

prospective teachers score below 30%, while about 29% of the participants scored below

30% for knowledge for teaching.
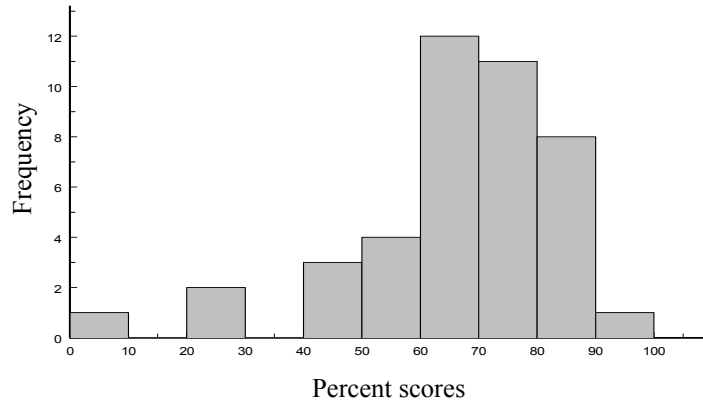
Figure 5.16. Distribution of Scores for Statistical Knowledge
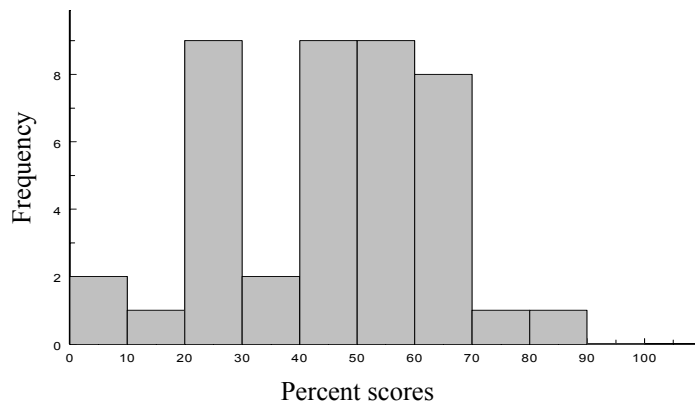
Figure 5.17. Distribution of Scores for Knowledge for Teaching

Performance by Cognitive Demand

The assessment instrument also measures three different levels of cognitive

demands for the domain that only involves statistical knowledge. The level of Statistical

Literacy was measured with 5 questions (Items 1a, 4a, 8ai, 8aii, and 6b); the level of

Statistical Reasoning was also measured with 5 questions (Items 1b, 2, 4b, 4c, and 6a);

and the level of Statistical Thinking was measured with 2 questions (Items 1c, 8b).   As

expected, prospective teachers perform better at the lowest level of performance −

181

Statistical Literacy. This involves mainly extracting information from a graph and recognition, identification, computation or basic understanding of concepts. At the higher levels, statistical reasoning and thinking, prospective teachers do progressively worse (see Table 5.13). Statistical reasoning refers to the way students reason with statistical ideas when asked why or how results are produced. Statistical thinking refers to the application of students' understanding of real world problems; for measures and distribution this might mean using them to make predictions and inferences about the group to which the data pertain.

Table 5.13

Descriptive Statistics by Cognitive Demand

	Mean(%)	Median(%)	SD(%)
Statistical Literacy	74.0	75.0	18.5
Statistical Reasoning	63.1	65.0	20.9
Statistical Thinking	51.5	50.0	30.4
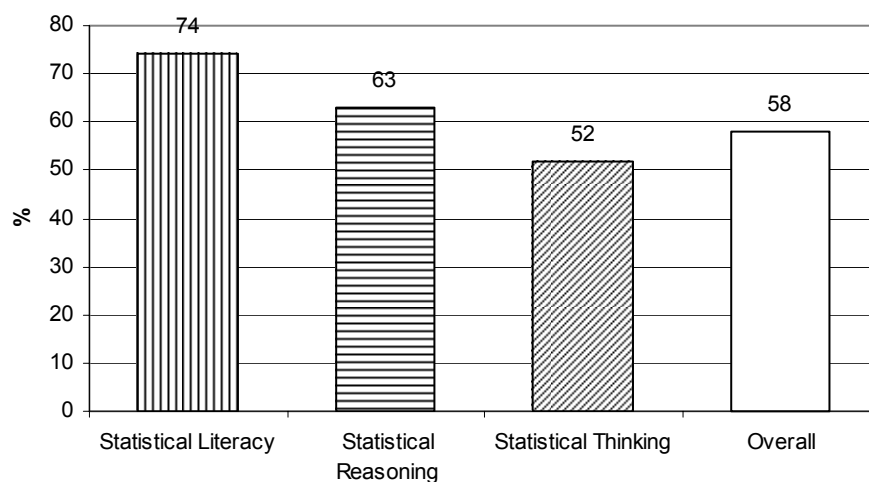
Figure 5.18. Mean Percentage Scores by Cognitive Demand

Percentage Success on Assessment

The two different domains of knowledge assessed can be analyzed by observing the percent of prospective teachers that successfully scored 3 or 4 on the items that fall into each category. Furthermore, since each item was designed to measure a specific aspect of content or content applied to teaching; the percent of success on these aspects can be analyzed. Table 5.14 shows the percent of success for statistical knowledge.

Table 5.14

Percent Success on Items Measuring Statistical Knowledge

Item	Percent Successful	Topic	Cognitive Demand
8ai	95	Range	Literacy (compute)
1b	93	Steam-and-leaf plot	Reasoning(interpret)
1a	90	Steam-and-leaf plot	Literacy (read)
4a	83	Mean	Literacy (compute)
6b	71	Graph representation	Literacy (construct)
4c	67	Mean	Reasoning(property)
1c	60	Steam-and-leaf plot	Thinking(infer)
4b	52	Mean	Reasoning(property)
2	45	Median	Reasoning(proper use)
6a	38	Graph representation	Reasoning(proper use)
8b	36	Standard deviation	Reasoning/Thinking
8aii	31	Standard deviation	Literacy (compute)
Mean % successful	63.42		

Note that most of the items with the highest percent of success are the items that fall into the level of statistical literacy, except for item 1b. Statistical literacy includes mainly computational skills, constructions of graphs; and the lowest level of graph comprehension (reading the data). These items with high percent of success do not fall into a single category of topics, they range from graphical representation, range, and mean. In contrast, the items with the lowest percent of success all, except for item 8aii,

fall into the statistical reasoning. Interestingly this is the level of cognitive demand that showed the highest emphasis suggested by state and national standards. Topics for the most difficult items also vary, however the topic of standard deviation show low percent of success both at the level of statistical literacy and statistical reasoning.

Similar analysis can be made for items that measured the domain of knowledge applied to teaching. Table 5.15 shows the percentage success for this domain and the corresponding teaching tasks.

Table 5.15

Percent Success on Items Measuring Knowledge Applied to Teaching

Item	Percent Successful	Topic	Teaching Task
5a	67	Formulating question	Select language appropriate for students' level
3a	62	Histogram	Identify students' mistake
3b	48	Histogram	Judge students' comment
7a	12	Data representation	Analyze students' work
7b	9.5	Data representation	Assess students' thinking
5b	2.4	Mean, median, mode	Judge students' comment
Mean % successful	33.48		

For this domain the highest percentages of success is shown for teaching tasks that involved the prospective teachers formulating questions that may have generated a data set using appropriate language for middle grade students. Also with high percent success among these items is identifying students' mistakes; however, judging the validity of students' comments shows greater degree of difficulty. With a very low percentage of success (2.4%) is item 5b which involved judging students' comments about measures of center. Note that computing, recognizing properties or using measures

184

of center had a higher percent of success for items that did not involve the application of teaching (see Table 5.14).

Summary and Discussion

Prospective teachers of middle school level that participated in this study showed better performance for items related to statistical knowledge than for items that had some application to teaching. These results are consistent with other studies related to other areas (Ball, 2001; Even, 1993; Even &Tirosh, 1995; Ma, 1999) which show that prospective teachers may know the subject matter well but not well enough to teach it based on children's conceptions. In this particular study, it is shown that most prospective teachers have the ability to formulate a question that may have generated a data set; construct correctly a graphical display and be able to extract information from it; compute and know how to find measures of center and spread for a set of values or small data displayed in a line plot.

In contrast, the majority of prospective teachers were neither able to identify correctly the source of erroneous or incomplete conceptions, nor could they infer or judge a child's level of understanding based on their work.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

Two central questions were investigated in this study. First, what are the important aspects of *content knowledge for teaching* data analysis and statistics at the middle school level? More specifically, what are the important statistical *content topics* taught in middle school that teachers need to be prepared for?; What are the *cognitive demands* (such as memorization, performing procedures and solving non-routine problems) that are related to the content?; What are the important aspects of *knowledge for teaching* that relate to the content aspect above?. Secondly, what are the conceptions and misconceptions prospective middle school teachers have with respect to these important aspects of knowledge for teaching data analysis and statistics?

Chapter 3 describes the important aspects for teaching data analysis and statistics and Chapter 5 describes prospective teachers' knowledge in relation to these aspects. Although these descriptions and findings may not tell the complete story, they do motivate discussion and ask better questions for future research on teacher preparation. As Ball, Lubienski, and Mewborn (2001) point out, the problem of what teachers need to know remains an unsolved one.   Stylianides and Ball (2004) suggest that one of the reasons for this is "that people have been looking at it from different angles, having available only limited research that concentrated on connecting the knowledge produced

by different perspectives" (p. 3). Stylianides and Ball (2004) suggest a framework that

brings together, in an *iterative cycle*, different perspectives such as analyzing experts'

perspectives, teachers' mathematics curricula, teachers' mathematical knowledge,

students' mathematical curricula, students' mathematical knowledge, and school

mathematics practice with the hope that the integration of all these resources may help

the study of the different domains of teachers' knowledge for teaching.

This study partially investigates some of the approaches suggested by the

Stylianides and Ball framework. The choice of the approaches studied here are practical

and based on the availability, at the time, of resources in the field of statistics education.

The approaches taken here are the analysis of experts' perspectives by examining

national and state standards at the student and teacher level, teachers' mathematical

knowledge by administrating an assessment, and the examination of a students'

mathematical curriculum. Teachers' mathematical curricula and school mathematics

practices were not examined for this study.

The chapter is divided into three parts: summary and discussion of main findings,

significance and implications, and recommendations for future research. A reminder note

for the reader, findings and conclusions of this experience are based on limited sources

and its generalizations should be done with caution. Aspects of content knowledge for

teaching were identified by looking at written documents and recent research, not by

observations of teaching practice. Prospective teachers' knowledge was measured by a

written instrument only on a convenience sample of 42 subjects taken from five eastern

universities.

Summary and Discussion of Main Findings

The identification of the most important aspects of knowledge for teaching, which addresses the first research question, was motivated by the need to create a framework upon which to develop an instrument to measure this construct. The identification of these aspects was done by analyzing experts' perspectives expressed in state and national standards, reports and a curriculum. A systematic analysis of these documents required searching for tools that have been applied to other similar purposes, such as content map analysis (Porter, 2003). This methodology was more successful with standards than with the rest of the documents given the format and language in which the standards are written. Complementing this methodology there were many other sources that help in understanding this complex idea. One is the literature on content knowledge for teaching in areas like numbers and operations or functions. Research here is well documented (see Ball, Lubienski, and Mewborn (2001) for a review of literature in these areas) and provides clear examples of the important aspects and its organization. Other resources include the most recent documents focused on teacher preparation published by the National Research Council (NRC, 2001) and the Conference Board of the Mathematical Sciences (CBMS, 2001) which give clear suggestions of the mathematical and statistical education of future teachers. Other documents are the numerous set of national and state standards at the student and teacher level on which many of the curriculum and programs for teacher preparation are based. Finally, in an attempt to get as close as possible to teaching practice, units developed by the Connected Mathematics Project (Lappan et al., 2002) on statistics and data analysis at the middle school level were examined.

Recently, current research on knowledge of teachers has shifted from looking at teachers' background to looking at teachers' knowledge to finally looking at knowledge *in* and *for* teaching (Ball et al. 2001). Because there is very little evidence of any association between teachers' credentials, such as course work in mathematics, and teacher effectiveness or student performance; researchers have taken the approach of studying the nature of teachers' mathematical knowledge focusing on the different kinds of knowledge. This last approach, although valuable, generates results that are based on hypothetical settings and does not provide evidence that teachers have applied it in practice. Hence, a new approach is suggested by Ball and others, mathematics knowledge *in* and *for* teaching that focuses on the core activities of teaching. This new approach requires watching teachers using their knowledge while they teach. "What emerges from these observations is that being able to talk about mathematics is different from doing it" (Ball et al., 2001, p. 450). This new approach has given new directions and hope for better answers to the problem of teachers' mathematical knowledge.

For this study, the approach taken was a mixture of the two latest approaches. Observation of actual teachers was beyond the scope of this study; so, in order to identify the aspects of knowledge for teaching statistics and data analysis a systematic analysis of documents and a textbook was conducted. As for measuring these aspects, a reliable instrument was constructed basing the teaching settings as close as possible to real practice, with the understanding that this instrument can only partially tell the story about what prospective teachers know and how they know it when confronted with teaching the subject.

Aspects of Knowledge for Teaching Data Analysis and Statistics

First a summary of the content that is expected of students at the middle grades level is given, not because this study is about students, but because the introduction of data analysis and statistics in the middle school curriculum is relatively new, in relation to other areas such as geometry and algebra. Minimally teachers have the responsibility to meet these expectations and given the newness of the statistics curriculum may have difficulty fulfilling this responsibility. This new material creates an extra challenge to teachers and prospective teachers, they themselves need to become familiar with an area that may never have been part of their pre-college education and must rely only on one or two statistics courses – not necessarily geared towards teachers – at the college level.

A systematic analysis of national and state standards shows that prospective teachers, as future teachers, should know and be able to teach the following content:

- Process of statistical investigation (includes formulating questions, collecting data, designing studies)

- Categorical data representation (includes bar/pie graphs, pictographs, tables)

- Numerical data representation (includes stem-and-leaf plots, histograms, box plots)

- Bivariate data representation (includes scatter plots, line graphs, regression line)

- Shapes of Distributions (includes shapes of distributions, skweness, gaps, outliers, clusters)

- Measures of center (includes mean, median, and mode)

- Measures of spread (includes range, interquartile range)

Prospective teachers are also responsible for engaging students in cognitive activities related to these content topics, the cognitive demands identified on the national and state standards are:

- *Perform routine procedures/Statistical literacy* (includes collecting data, reading/ creating graphs and data displays, finding and computing measures of center and spread, identifying clusters, gaps, outliers)

- *Communicating understanding/Statistical reasoning* (includes making decisions on what and how to measure, justifying the use and selection of graphical representation and measures of center and spread, interpreting graphs, and explaining findings and results)

- *Solve nonroutine problems/make connections/Statistical thinking* (includes applying statistics to real work contexts, analyzing data recognizing patterns, judging or critiquing statistical methods)

- *Conjecture/generalize/prove/Statistical thinking* (includes making inferences, predictions, and generalizations from data displays and measures)

The list above is a comprehensive one; results show that state and national standards at the student level vary a great deal in terms of the inclusion of and emphasis given to topics and activities. In particular, state standards differ on the level of cognitive demand for each topic. When both topic and cognitive demand are accounted for, only measure of center at the level of communicating understanding is found in the standards of all 10 states analyzed. The national and state standards differ significantly as well. On the whole, state standards emphasize categorical data representation and performing procedures more than the national standards do.

Fewer documents exist which provide standards for middle school teachers as opposed to middle school students. Two documents were reviewed for this study, the PRAXIS II and *The Mathematical Education of Teachers* (CBMS, 2001). The latter places emphasis on the process of statistical investigation and shapes of distributions, while the former emphasizes categorical data representation. Both documents place higher emphasis on bivariate data representation at the level of solve non routine problems than do the standards for students.

Using a methodology to identify content adapted from Porter (2003) it was found that the average emphasis of all the documents reviewed is given to: 1) graphical representation, more numerical than categorical, at the level of creating and using them properly; and 2) measures of center and spread at the level of computation and proper selection. Hence, these aspects of content are included in the assessment instrument developed for this study. A limitation on using this methodology is that it depends on the language used in the documents and not on the intent experts have given in the recommendations, therefore documents that are not written with clear distinction between topics and cognitive demands were very hard to analyze (e.g. *The Mathematical Education of Teachers* (CBMS, 2001) report).

Aspects of knowledge for teaching data analysis and statistics were identified from teachers' standards, National Research Council (NRC) documents, and the Connected Mathematics Project's two units on statistics from the middle grade curriculum. Each document suggest a different organization for looking at teachers' knowledge. State and national teachers' standards give *general* recommendations and organize the knowledge into four kinds:

- Knowledge of the subject matter

- Knowledge of pedagogy (some include pedagogical content knowledge)

- Knowledge of students' learning

- Knowledge of assessment

Adding It Up (NRC, 2001) besides focusing on these kinds of knowledge, also adds cognitive activities that teachers engage in relation to these kinds of knowledge more *specifically* tied to mathematics. The authors argue that teachers' learning, like students' learning, can be conceived in terms of interwoven strands. The strands or components that are required in teaching for mathematical proficiency are the following.

- *conceptual understanding* of the core knowledge required in the practice of teaching;

- *fluency* in carrying out basic instructional routines;

- *strategic competence* in planning effective instruction and solving problems that arise during instruction;

- *adaptive reasoning* in justifying and explaining one's instructional practices and in reflecting on those practices so as to improve them; and

- *productive disposition* toward mathematics, teaching, learning, and the improvement of practice.

Another NRC document, *Knowing and Learning Mathematics for Teaching* (NRC, 2001), focuses instead on practical activities of teaching practice that are directly related to mathematics. Six categories of recurrent teaching tasks are identified that require the use of mathematics: 1) managing class discussions, 2) establishing a classroom culture for mathematical reasoning, 3) designing and selecting tasks, 4)

analyzing student thinking and work, 5) planning instruction, and 6) assessing student learning.

All these different frameworks and organizations intersect in one or more points. Each one of the categories, strands, components or kinds of knowledge for a particular organization could be mapped into another framework. For example, we can map the teaching task of "analyzing student thinking and work" onto "knowledge of students" in the first framework and also to "strategic competence" in the second.

The different organizations of knowledge for teaching help create a general vision of the different aspects that one needs to consider for assessment, which is the ultimate goal of this study. However, these organizations are too general. It is not clear if all of these aspects apply in all areas within mathematics, in particular in data analysis and statistics. If they do, do they apply to all topics? The necessary intersection of the actual content and all of these aspects of knowledge of teaching is still absent. As mentioned before, observing teachers teach the subject would have been the ideal path. Failing that, relying on research done in this area would suffice. However, the first was beyond the scope of this study and to date there is no research in this area. The best approximation of this intersection was the teachers' guide to a unit in statistics from a middle grade curriculum.

The book examined was the Teachers' Guide of the *Connected Mathematics* Statistics unit: Data About Us (Grade 6) (Lappan et al., 2002). Although this unit covers several topics identified previously, only the two topics which received greater emphasis in the standards and documents were examined, namely organizing and representing data, and measures of center.

For teaching data representations the knowledge of students in relation to the content was identified to be:

- Know how to create questions that middle school students can ask in order to collect data and might involve using a specific type of graph;

- Know how to respond to a student who want to use an inappropriate type of graph

- Assess students' responses making judgments about their reasoning;

The Teachers Guide also identified aspects related to pedagogy for data representation:

- Engage students in the exploration of data by having them suggest questions that might have originated the data and methods of collecting the data;

- Lead students in the process of constructing a stem-and-leaf plot, and;

- Pose questions that lead to "read the data", "read between the data" and "read beyond the data".

As for measures of center, in particular, the mean; the following pedagogical aspects were identified, most of them dealing with the proper use of physical models:

- Know the advantages and limitations of physical models to introduce the concept of the mean as the "evened out" number;

- Know how to make connections between the physical model and the line plot;

- Know how to create data sets with the same mean but different distribution using physical models; and

- Lead students to the discovery of the algorithm of the mean and why it works.

Finally, aspects of knowledge of students in relation to the measures of center were also identified:

- Understand how students are thinking about the data when they use different strategies and models to find the mean;

- Assess proper statistical reasoning for justifying students' strategies;

- Respond to students who think that is impossible to have many data sets with the same mean, and

- Anticipate students' answers or interpretation to an investigation question and be able to pose questions to students that lead them to see the effect of outliers or/and new data values have on the distribution and the mean.

Some of these aspects were considered in the measurement of prospective teachers' knowledge for teaching, especially the ones related to knowledge of students. Some were addressed as items in the written instrument and some in the interviews.

Prospective Teachers' Statistical Knowledge

To address the second research question about measuring teachers' knowledge an instrument was developed, piloted twice, and administered at five eastern universities. The instrument was divided into two major domains, statistical knowledge and statistical knowledge applied to teaching. Number and allocation of items measuring the statistical knowledge were selected to represent the percentage of content identified in the content maps. The applications to teaching were restricted to students' work and responses in relation to statistical concepts given the constraints of time and the written format. The reliability of the instrument is $\alpha = .80$.

Although the items were divided into content and content applied to teaching, the items measuring the latest domain are rooted in content and therefore related to the first

domain. However, the division provides a way to compare how prospective teacher know the content with and without teaching contexts.

The sample consisted of 42 senior prospective teachers mainly females in their early 20s with strong background in mathematics and education. All of them had taken at least one course in mathematics education and all but four had taken a course in statistics. Prospective teachers were selected conveniently from five different universities which have a program in middle school mathematics and are located in a state that requires a middle grade certification. A subset of the sample (7) participated in follow-up interviews.

Written responses were scored according to a 0 – 4 scale rubric. Analyses at the item and overall level were conducted. Interview responses helped enrich the analysis of the instrument. Following are summaries of the results organized by topics or big ideas within each domain of knowledge.

Knowledge of Statistics

The aspects of statistical content measured in the instrument were: 1) Formulation of questions to generate data, 2) Numerical data representation; and 3) Measures of center and spread. An attempt to measure each of these aspects at all levels cognitive demand was made, but it was not possible given the limitations of a written instrument. However, the instrument has the following representation of aspects of content and cognitive demand.

Formulation of questions

Participants in this study were given a table showing different kinds of pets and its respectively frequency (see Item 5 of instrument). They were asked to create a possible question that could have generated the data. Two thirds of the prospective teachers were able to create a correct possible question, although some of them did not use a proper question format.

Numerical data representation

About 70% of the prospective teachers were able to create graphical representation of data; however they based their choice on practical reasons for construction. There was very little evidence that prospective teacher took the question needing to be answered into account for the choice of the graph.

The data given was a list of students (initials) and their time to get to school (in minutes). The most common choice was a bar graph showing the initial of the students on the *x*-axis. It seems important to prospective teachers to show the correspondence between the time and each student. This is an interesting point because statisticians understand data more as distributions rather than as representing separate pieces of information. Some of the prospective teachers argued that because of the time variable, they needed to use a line graph.

Nearly all (93%) of the prospective teachers were able to read and interpret data in a stem-and-leaf plot. However, about 60% of prospective teachers used neither statistical measures nor statistical reasoning to make inferences about the data. When interpreting histograms, the main difficulty was to interpret what each bar represented.

Many prospective teachers focused on the *x*-axis variable and did not know how to relate that to the "Frequency".

Measures of center and spread

About 80% of prospective teachers were able to compute and estimate the mean of a small data set represented in a line plot. When given a bigger data set and asked to estimate the mean, they either chose the median as an estimate without observing the distribution or in their attempt to find the mean, they found the average of the frequencies instead. The latter seems to be a reflection of the methodology taught in elementary and middle school of leveling off stacks of cubes.

About half of prospective teachers knew that it is possible to have many sets of data with the same mean. However, only a third of those could justify it with an argument that relies on both the algorithm and the concept of the mean as a balance point. About two thirds of prospective teachers were able to create a distribution with a specific mean that is not a whole number. Prospective teachers interviewed and probed for knowledge for teaching did not show sufficient evidence of this type of knowledge. They could neither estimate the mean of a data set without reaching for the algorithm, nor convince a child that there could be several data sets with the same mean, nor explain what an average of 3.5 people means.

About half of prospective teachers were able to identify outliers in a data set but not necessarily to use this knowledge to choose the appropriate measure of center. Some actually wanted to take the outlier into account for representative reasons, others for pedagogical reasons. When asked to say what is the typical value for a data set given, prospective teachers choose a measure of center or a range of values independently of the

shape of the distribution. Furthermore, about half of them could not make sense of the computed measures of center and the corresponding shape of the distribution. One likely explanation for these results is the use of the word "typical" which is not a well defined statistical term. It could be thought as the mode, median or mean. As prospective teachers relate this term to one of the measures of center, they tend to attach a single measure of center to the word "typical" (e.g. the mean) and use it to describe the center of the distribution regardless of the shape.

Prospective teachers who had computational knowledge of the range and the standard deviation did not necessarily have understanding of the concept. This result was somehow expected. However, results here show that the number of prospective teachers who have computational knowledge and poor or no conceptual knowledge is lower than the number of those who actually understand or partially understand the concept but do not have the computational knowledge.

Knowledge for Teaching Statistics

As for statistical knowledge applied to teaching, the instrument focuses on the knowledge of students as learners. The aspects considered for the instrument within this domain are: 1) interpretation of students' oral and written responses in relation to the content, and 2) examination of students' strategies and solutions to exercises to make inferences about their understanding. Results in this part of the assessment were lower than the part on statistical content and unveiled conceptions and misconception about statistical content that were not captured by items with no teaching contexts. One of the reasons for lower performance might be due to the new format of items. Most

prospective teachers' responses were based on their pedagogical rather than statistical content knowledge. The expectation that the responses needed to be focus on the statistical content was not explicitly given in wording of the item. This attempt was done only with the interviewed subjects (see Interview Protocols in Appendix E). For example, when analyzing student work, prospective teacher were asked how students' own graphical representations resemble the standard statistical graphs. Some prospective teachers were able to correctly identify the features in students' graphs that would resemble a bar graph or a stem-and-leaf graph. In contrast, others would focus on the physical drawing and said that the students' graphs would resemble a pictograph because the student drew classmate faces.

Interpretation of students' written responses

About three fourths of prospective teachers could recognize a typical student's misconception when reading a histogram and half could also identify the reason why the student is making the mistake. However, the results show evidence that the prospective teachers have other limitations with the interpretation of the histogram and are unable to correct the mistake properly. In other words, prospective teachers correctly make the judgment that the student is wrong but give the incorrect way to correct the mistake.

In contrast, most could not recognize students' mistakes about the inappropriateness of using the median or range for categorical data. Most prospective teachers could not describe what the measures of center or spread say about the data when trying to find them. Instead they use a procedural method to get an answer. In this

particular case, prospective teachers were very creative inventing their own methods to compute a median for categorical data.

Examination of students' strategies and solutions

Analyzing and judging students' data representation was extremely hard for prospective teachers. When asked to compare and contrast their work, they focused only on the procedural and physical differences of the graph representations. Since the work of both students presented represented the same data, prospective teachers generally could not describe any difference in students' understanding of the data claiming that "they both show the same information in different way".

It is not clear from this investigation the reason why prospective teachers do not have the knowledge required to teach statistics. One potential cause, and that needs further investigation, may be the lack of opportunity to learn applications to teaching in the way they were assessed in their own teacher preparation programs. Since this variable was not measured, these particular prospective teachers should not be held accountable for their lack of knowledge. Another reason could be the integration of pedagogy issues and content in each item; prospective teachers may not have had a clear notion what was expected from them. A successful response was expected to focus on statistical content, but most of prospective teacher focused on pedagogical issues. Better efforts are needed to create items to measure this construct so that prospective teachers know what is expected of them.

Although the assessment instrument was clearly divided into two domains, statistical content and statistical content applied to teaching, inferences on what

prospective teachers know about statistics for teaching can not be separated in this fashion. As statistical content is present in both domains, claims can only be made about the way the context of teaching changes how prospective teachers think and understand the content.

Significance and Implications

The measure and description of prospective teachers' knowledge, its application in teaching, and understanding about basic statistical concepts should be of interest to several communities: teacher educators, statistics educators, mathematics teacher educators, prospective teachers' curriculum developers, statistics professors, and assessment developers. The implications of the results are theoretical, practical, and methodological. The first two implications are related to the goals of the study – the identification of important aspects of statistical knowledge for teaching, and the descriptions of the knowledge prospective teachers have with respect to the above aspects. The third implication has to do with the way the aspects were identified and the development of a reliable instrument to measure them.

Several aspects of statistical content knowledge were identified as very important for middle grades teachers. The choice of aspects was based on a systematic integrated analysis of several documents such as students' state and national standards, which in part, are based on theoretical and empirical work on student learning. This approach of identification of content is of particular interest to those that are trying to make hard choices on what to include – or not to include – in curriculum guides and assessment for preservice teachers and professional development developers for inservice teachers.

The aspects of statistical knowledge for teaching were not identified in the same fashion as the content. This kind of knowledge is much more complex and has too many dimensions to be analyzed the same way that students' content is. However, several pieces of work were examined and integrated: teachers' state and national standards, research and theoretical work on teachers' knowledge and its role in teaching, and teachers' guide to a statistics textbook. These documents provided a general framework to view teachers' knowledge as well as specific aspects that come from the actual practice of implementing curriculum. They are all important as we need the "big picture" to create vision in teacher preparation programs and the "little picture" to make it happen in the classroom and to create authentic assessment instruments. The identification of these aspects for teaching is a starting point for a discussion of what do middle grade teachers need to know about statistics in order to teach it well and a continuation of how to measure this knowledge in the other areas of mathematics.

Statistics and mathematics educators are concerned about the way statistics is taught at all levels, in particular at the elementary and middle grades levels where the inclusion of this content is relatively new. They are also concerned about the way teachers understand statistical concepts, as they are different from mathematical concepts (Bright & Friel, 1998; Burrill, 1998; Cobb & Moore, 1997; Coob, 1992; Friel & Bright, 1998; Garfield, 1995; Garfield & Ahlgren, 1998; Mokros & Russell, 1995; Moore, 1997; Shaughnessy, 1992). The measurement and description of the statistical knowledge that prospective teachers have about central statistical topics taught in middle grades in this study raises concerns about the opportunities future teachers are having in their preparation to learn statistics in a way that they can teach it to young adolescents. Lack

of statistical reasoning to justify and judge proper use of graphical representations and measures of center and spread, knowledge of procedures and facts without understanding where the procedures come from or what do the facts tell you about the data, are just some of the findings that may keep teachers from teaching statistics in a meaningful way.

Current reform efforts in teacher preparation (e. g., Wilson, Floden, & Ferrini-Mundy, 2001; NRC, 2001a) point out that the subject matter preparation is important and the current research results are disappointing. They also argue for the need for studies that give insights into the nature of content and quality across areas and levels. This study provides needed information about the limitations of middle grades teacher preparation programs in the area of statistics and data analysis. Changes should be made so that middle grades mathematics teachers will understand more profoundly the basic concepts of statistics and data analysis taught in middle grades, be exposed to children's work and strategies to represent and reason about data, and understand the study of statistics as a process of data investigation, not as a collection of procedures and formulas.

Furthermore, this study contributes to the argument that the knowledge of the subject matter is a necessary but not sufficient condition for teachers to teach it well. Most of the subjects that participated in this study were prospective middle school teachers, seniors with strong mathematical and education backgrounds, including two mathematics education courses, on average; and one basic course in statistics. That is, making a course or two a requirement in their program because they are likely to teach it, does not guarantee that they would be able to teach it well. It seems that a required statistics course needs to be backed up with a course that integrates the content and its

application in teaching because results show that although the prospective teachers have taken a course in statistics their performance is low in the area of application. Statistics teachers at the university and college level may use this information to make arguments for developing new courses geared toward future teachers that will better serve them in practice.

The methodological implications of this study are concerned with the systematic way to analyze important content and the development of a reliable instrument to measure a construct such as knowledge for teaching. Although the written instrument is limited to some aspects of data analysis emphasized in middle school, it has the advantage of being able to assess a fairly large scale number of subjects and what is most important, it provides a quantity associated to those aspects. Conscious of the limitations, some written responses were followed-up with interviews complementing some of the gaps and providing more detail.

Recommendations for Future Research

This study investigated statistical knowledge for teaching by studying only some of its components. Mainly by analyzing experts' perspectives and measuring teachers' statistical knowledge. The answer to the problem is not given in this single study conducted primarily by a single person and it is far from complete. For example, the study of teachers' statistical curricula, students' statistical knowledge, and school statistical practices is needed. What it is hoped to be accomplished is a contribution to the solution to the problem by understanding some of its pieces. There is much more to be learned about this process of teaching and learning statistics for teaching.

One of the difficulties in developing the instrument was to identify the important aspects of statistics applied to teaching situations. There is no research that focuses on the statistical knowledge in the context of teaching, that is, studying how the knowledge of statistics interacts with the actual work of teaching it. Research is needed in this area to strengthen the validity of the instrument and create more authentic teaching-related items.

The instrument developed for this study only covers a limited number of statistical topics and aspects of knowledge for teaching. Topics like bivariate data representation and distributions need to be added as well as other aspects of knowledge for teaching such as how to help students learn statistical concepts, evaluation of curriculum materials, and development of student assessment to name a few.

It would be useful, as well, to generate more items and pilot them to create a closed version of the instrument. At the moment the instrument is short and free response which is hard to grade, time-consuming, and limits the number of participants.

One of the beauties of measurement is that we can assign a quantity to a construct, not necessarily for accountability, but to make association with other variables. One of the most needed associations is between teachers' knowledge and students' achievement. The ultimate goal of this project is to be able to measure that association. As researchers, we would like to answer the question of whether there is a relationship between what teachers know for teaching and their students' understanding of statistics.

APPENDIX A

State and National Student Standards

Connecticut State Department of Education
Mathematics Curriculum Framework
Division of Teaching and Learning
Retrieved June 22, 2004 from http://www.state.ct.us/sde/dtl/curriculum/Frmath.pdf

Content Standard 7: Statistics and Probability

Students will use basic concepts of probability and statistics to collect, organize, display and analyze data, simulate events and test hypotheses.

K-12 Performance Standards

Educational experiences in **Grades 5-8** will assure that students:

- make conjectures; design simulations and samplings; generate, collect, organize and analyze data; and represent the data in tables, charts, graphs and creative data displays;

- make inferences and formulate and evaluate hypotheses and conclusions based on data from tables, charts and graphs;

- describe the shape of the data using range, outliers, and measures of central tendency, including mean, median and mode;

- select and construct appropriate graphical representations and measures of central tendency for sets of data.

Florida Department of Education
Sunshine State Standards
Mathematics Grade 6-8
Retrieved June 22, 2004 from http://www.firn.edu/doe/curric/prek12/frame2.htm
Data Analysis

Standard 1:

The student understands and uses the tools of data analysis for managing information.

- collects, organizes, and displays data in a variety of forms, including tables, line graphs, charts, bar graphs, to determine how different ways of presenting data can lead to different interpretations.

- understands and applies the concepts of range and central tendency (mean, median, and mode).

- analyzes real-world data by applying appropriate formulas for measures of central tendency and organizing data in a quality display, using appropriate technology, including calculators and computers.

Standard 3:

The student uses statistical methods to make inferences and valid arguments about real-world situations.

- formulates hypotheses, designs experiments, collects and interprets data, and evaluates hypotheses by making inferences and drawing conclusions based on statistics (range, mean, median, and mode) and tables, graphs, and charts.

Georgia Department of Education
Quality Core Curriculum Standards
Retrieved June 22, 2004 from http://www.glc.k12.ga.us/qcc/homepg.asp

Grade 6 Mathematics Statistics

- **Topic:** Data Collection, Data Organization, Data Display, Scale
  Standard: Collects and organizes data, and determines appropriate method and scale to display data.

- **Topic:** Data Collection, Data Organization
  Standard: Constructs tables, charts, pictographs and bar, circle, and simple line graphs to display data.

- **Topic:** Mean, Median, Mode, Range
  Standard: Finds median, mean, mode, and range of a given set of data.

Grade 7 Mathematics Statistics

- **Topic:** Charts, Tables, Graphs, Distributions
  Standard: Collects, organizes data, determines appropriate method and scale to display data, and constructs frequency distributions, bar graphs, line graphs, circle graphs, tables, and charts.

- **Topic:** Measures of Central Tendency and Spread
  Standard: Uses mean, median, and mode to describe central tendencies of a data set, and uses range to describe spread of the data.

- **Topic:** Charts, Tables, Graphs, Distributions
  Standard: Reads and interprets data in frequency distributions, diagrams, charts, tables, and graphs; and makes predictions or conclusions based on this data.

Grade 8 Mathematics Statistics

- **Topic:** Data Collection, Data Organization, Data Display, Scale
  Standard: Collects and organizes data, determines appropriate method and scale to display data, and constructs frequency distributions; bar, line, and circle graphs; tables and charts; line plots, stem-and-leaf plots, box-and-whisker plots, and scatter plots.

- **Topic:** Mean, Median, Mode, Range
  Standard: Uses mean, median, mode, and range to describe tendencies of a data set and make predictions.

Grade 6 through 8 with Assessment in Grade 8th

Statistics

Concepts - Students will describe properties of, define, give examples of, and/or apply to both real-world and mathematical situations:
- Meaning of central tendency (mean, median, mode)

- Meaning of dispersion (range, cluster, gaps, outliers)

- Characteristics and appropriateness of graphs (e.g., bar, line, circle), and plots (e.g., line, stem-and-leaf, box-and-whiskers, scatter)

Skills - Students will perform the following mathematical operations and/or procedures accurately and efficiently, and explain how they work in real-world and mathematical situations:
- Organize, represent, analyze, and interpret sets of data

- Construct and interpret displays of data (e.g., table, circle graph, line plot, stem-and-leaf plot, box-and-whiskers plot)

- Find mean, median, mode, and range; recognize outliers, gaps, and clusters of data

Relationships - Students will show connections and how connections are made between concepts and skills, explain why procedures work, and make generalizations about mathematics in meaningful ways for the following relationships:
- How different representations of data (e.g. tables, graphs, diagrams, plots) are related

- How data gathering, bias issues, faulty data analysis, and misleading representations affect interpretations and conclusions about data (e.g., changing the scale on a graph, polling only a specific group of people, using limited or extremely small sample size)

- How probability and statistics are used to make predictions and/or draw conclusions

North Carolina Department of Education
Current Standard Course of Study and Grade Level Competencies (1998)
Mathematics Curriculum – Middle Grades 6-8

Retrieved June 22, 2004 from
http://www.dpi.state.nc.us/curriculum/mathematics/middle.html

Data, Probability, and Statistics
Competency Goal 4
Grade 6

- Create and evaluate graphic representations of data.

- Use measures of central tendency to compare two sets of data.

- Construct convincing arguments based on analysis of data and interpretation of graphs.

Grade 7

- Interpret and construct histograms.
- Compare and relate bar graphs and histograms.
- Construct circle graphs using ratios, proportions, and percents.
- Create, compare, contrast, and evaluate both orally and in writing, different graphic representations of the same data.
- Identify appropriate uses of different measures of central tendency.

Grade 8

- Interpret and construct box plots.
- Collect data involving two variables and display on a scatter plot; interpret results; identify positive and negative relationships.
- Interpret the mean, explain its sensitivity to extremes, and explain its use in comparison with the median and the mode.

Missouri Department of Elementary and Secundary Education
Missouri's Frameworks for Curriculum Development
Mathematics

Retrieved June 22, 2004 from
http://dese.mo.gov/divimprove/curriculum/frameworks/math.html

By the end of grade 8, all students should be able to

- develop, analyze, and explain methods utilized to collect, organize, and describe data

- make, read, and interpret multiple representations including tables, charts and graphs of data

- formulate, predict, and defend positions taken that are based on data collected

- analyze information and arguments that are based on data collected

- investigate the power of making decisions based on statistical methods and the applications of probability in the real world

- use computers, graphing calculators, and/or other forms of technology to enhance understanding of numbers, data, and the resulting analysis

Ohio Department of Education
Joint Council of the
State Board of Education and the Ohio Board of Regents
Mathematics Academic Content Standards

Retrieved June 24, 2002 from
http://www.ode.state.oh.us/academic_content_standards/acsmath.asp

Data Analysis Standard

Students pose questions and collect, organize, represent, interpret and analyze data to answer those questions. Students develop and evaluate inferences, predictions and arguments that are based on data.

By the end of the 6 – 8 program…

- Select, create and use appropriate graphical representations of data including histograms, circle graphs, box plots and scatter plots and justify the selection of the graph types.

- Evaluate different graphical representations of the same data to determine which is the most appropriate representation for an identified purpose.

- Find, use and interpret measures of center and spread, including mean and interquartile range, and use these measures to compare two sets of data.

- Interpret the mean, explain its sensitivity to extremes, and explain its use in comparison with the median and mode.

- Construct convincing arguments based on analysis of data and interpretation of graphs.

Mathematics – Statistics Grade 6-8

- Find, use, and interpret measures of center and spread, including mean and interquartile range for given or derived data.

- Formulate questions and design experiments or surveys to collect relevant data.

- Represent and interpret data using frequency distribution tables, box-and whisker-plots, stem-and-leaf plots, and single- and multiple- line graphs.

- Determine the graphical representation of a set of data that best shows key characteristics of the data.

- Recognize distortions of graphic displays of sets of data and evaluate appropriateness of alternative displays.

- Analyze data from frequency distribution tables, box-and whisker-plots, stem-and-leaf plots using measures of center and spread and draw conclusions.

- Predict and evaluate how adding data to a set of data affect measures of center.

- Use observations about differences between two or more samples to make conjectures about the populations from which the samples were taken.

Virginia Department of Education
Standards of Learning
Mathematics
Retrieved June 22, 2004 from http://www.pen.k12.va.us/go/Sols/math.html

Statistics

Grade 6

- The student, given a problem situation, will collect, analyze, display, and interpret data in a variety of graphical methods, including line, bar, and circle graphs and stem-and-leaf and box-and-whisker plots. Circle graphs will be limited to halves, fourths, and eighths.

- The student will describe the mean, median, and mode as measures of central tendency and determine their meaning for a set of data.

Grade 7

- The student will create and solve problems involving the mean, median, mode, and range of a set of data.

- The student will display data, using frequency distributions, line plots, stem-and-leaf plots, box-and-whisker plots, and scattergrams.

- The student will make inferences and predictions based on the analysis of a set of data that the student(s) collect.

Grade 8

- The student will use information displayed in line, bar, circle, and picture graphs and histograms to make comparisons, predictions, and inferences.

West Virginia Department of Education
Mathematics Content Standards and Objectives for West Virginia Schools
Retrieved June 22, 2002 from http://wvde.state.wv.us/policies/p2520.2_ne.pdf

Standard 5: Data Analysis and Probability

Students will:
- formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them;

- select and use appropriate statistical methods to analyze data; develop and evaluate inferences and predictions that are based on models.

Data Analysis and Probability Objectives

Grade 6
- collect, organize, display, and interpret data using line graphs, circle graphs, bar graphs, histograms, stem-and-leaf plots, tables, and charts.

- Create and solve problems involving the mean, median, mode, and range of a data set of data.

Grade 7

- Read and interpret multiple line graphs;

- extrapolate information from multiple-line graphs, circle graphs, bar graphs, histograms, tables, and frequency distributions (tally charts);

- collect, organize, graphically represent, and interpret data using frequency distributions, line-plots, stem-and-leaf plots, box-and-whisker plots, and scatter plots;

- determine measures of central tendency (mean, median, mode, range) and dispersion from data, table, and experiment.

Grade 8

- draw inferences and construct convincing arguments based on data analysis.

Principles and Standards for School Mathematics
National Council of Teachers of Mathematics (2000)

Data Analysis Standard for Grades 6 – 8

In grades 6 – 8 all students should –

- formulate questions, design studies, and collect data about a characteristic shared by two populations or different characteristics within one population;

- select, create, and use appropriate graphical representation of data, including histograms, box plots, and scatterplots.

- Find, use, and interpret measures of center and spread, including mean and interquartile range;

- discuss and understand the correspondence between data sets and their graphical representations, especially histograms, stem-and-leaf plots, box plots, and scatterplots.

- Use observations about differences between two or more samples to make conjectures about the populations from which the samples were taken;

- make conjectures about possible relationships between two characteristics of a sample on the basis of scatterplots of the data and approximate lines of fit;

- use conjectures to formulate new questions and plan new studies to answer them.

APPENDIX B

Codes Assigned to Student Standards

Table B.1

Codes for Connecticut Students Standards

State:	Connecticut
Grade:	Grades 5-8

	Rating Standard: CT1		Rating Standard: CT2		Rating Standard: CT3		Rating Standard: CT4		Rating Standard:	
	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand
1	1	E	2	E	5	B	2	C		
2	1	D			6	B	6	C		
3	1	B			7	B				
4	2	B								
5										
6										
7										
8										
9										
10										

Table B.2

Codes for Florida Students Standards

State:	Florida
Grade:	Grades 6-8

	Rating Standard: FL1		Rating Standard: FL2		Rating Standard: FL3		Rating Standard: FL4		Rating Standard:	
	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand
1	1	B	6	C	2	D	1	E		
2	2	B	7	C	3	D	2	E		
3	4	B			6	D	3	E		
4	2	C					6	E		
5	4	C					7	E		
6										
7										
8										
9										
10										

Table B.3

Codes for Georgia Students Standards

State:	Georgia
Grade:	Grades 6-8

	Rating Standard: GA1		Rating Standard: GA2		Rating Standard: GA3		Rating Standard: GA4		Rating Standard: GA5		Rating Standard: GA6		Rating Standard: GA7	
	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand
1	1	B	2	B	6	B	3	B	2	D	3	B	3	B
2	2	C	4	B	7	B	3	C	2	E	3	C	3	C
3	4	C							4	D				
4									4	E				
5														
6														
7														
8														
9														
10														

Table B.4

Codes for Kentucky Students Standards

State:	Kentucky
Grade:	Grades 6-8

	Rating Standard: KY1		Rating Standard: KY2		Rating Standard: KY3		Rating Standard: KY4		Rating Standard: KY5		Rating Standard: KY6		Rating Standard: KY7	
	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	
1	6	C	5	C	2	C	2	B	5	B	1	E	2	C
2			7	C	3	C	3	B	6	B			3	C
3					4	C	4	B	7	B			4	C
4														
5														
6														
7														
8														
9														
10														

Table B.4 (continued)

Codes for Kentucky Students Standards

State:	Kentucky
Grade:	Grades 6-8

	Rating Standard: KY8		Rating Standard: KY9	
	Topic	Cogn. Demand	Topic	Cogn. Demand
1	2	D	6	E
2	3	D	7	E
3	4	D		
4				
5				
6				
7				
8				
9				
10				

Table B.5

Codes for North Carolina Students Standards

State:	North Carolina
Grade:	Grades 6-8

	Rating Standard: NC1		Rating Standard: NC2		Rating Standard: NC3		Rating Standard: NC4		Rating Standard: NC5		Rating Standard: NC6		Rating Standard: NC7	
	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	
1	2	B	6	D	2	E	3	B	2	C	2	C	2	C
2	2	B					3	C	3	C			3	C
3														
4														
5														
6														
7														
8														
9														
10														

Table B.5 (continued)

Codes for North Carolina Students Standards

State:	North Carolina
Grade:	Grades 6-8

	Rating Standard: NC8		Rating Standard: NC9		Rating Standard: NC10		Rating Standard: NC11	
	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand
1	6	C	3	B	4	B	6	C
2			3	C	4	C	6	D
3								
4								
5								
6								
7								
8								
9								
10								

Table B.6

Codes for Missouri Students Standards

State:	Missouri
Grade:	Grades 5-8

	Rating Standard: MO1		Rating Standard: MO2		Rating Standard: MO3		Rating Standard: MO4		Rating Standard:	
	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand
1	1	B	2	B	1	C	1	D		
2	1	d	2	C	1	E				
3			3	B						
4			3	C						
5			4	B						
6			4	C						
7										
8										
9										
10										

Table B.7

Codes for Ohio Students Standards

State:	Ohio
Grade:	Grades 6-8

	Rating Standard: OH1		Rating Standard: OH2		Rating Standard: OH3		Rating Standard: OH4		Rating Standard:	
	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand
1	2	B	6	C	6	C	2	E		
2	2	C	6	B			3	E		
3	3	B	7	C						
4	3	C	7	B						
5										
6										
7										
8										
9										
10										

Table B.8

Codes for Oregon Students Standards

State:	Oregon
Grade:	Grades 6-8

	Rating Standard: OR1		Rating Standard: OR2		Rating Standard: OR3		Rating Standard: OR4		Rating Standard: OR5	
	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand
1	6	B	1	B	3	B	3	C	5	B
2	6	C	1	D	4	B	4	C		
3	7	B	1	E						
4	7	C								
5										
6										
7										
8										
9										
10										

Table B.8 (continued)

Codes for Oregon Students Standards

State:	Oregon
Grade:	Grades 6-8

	Rating Standard: OR6		Rating Standard: OR7		Rating Standard: OR8	
	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand
1	3	D	6	C	1	E
2	6	D				
3						
4						
5						
6						
7						
8						
9						
10						

Table B.9

Codes for Virginia Students Standards

State:	Virginia
Grade:	Grades 6-8

	Rating Standard: VA1		Rating Standard: VA2		Rating Standard: VA3		Rating Standard: VA4		Rating Standard: VA5		Rating Standard: VA6	
	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand
1	1	B	6	B	6	D	3	B	1	D	2	E
2	2	B	6	C	7	D	4	B	1	E	3	E
3	2	C										
4	2	D										
5	3	B										
6	3	C										
7	3	D										
8	4	B										
9	4	C										
10	4	D										

Table B.10

Codes for West Virginia Students Standards

State:	West Virginia
Grade:	Grades 6-8

	Rating Standard: WV1		Rating Standard: WV2		Rating Standard: WV3		Rating Standard: WV4		Rating Standard: WV5		Rating Standard: WV6		Rating Standard: WV7	
	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand
1	1	B	6	D	4	B	2	E	4	B	6	B	1	D
2	2	B	7	D	4	C	3	E	4	C	7	B	1	E
3	2	C					4	E						
4	3	B												
5	3	C												
6	4	B												
7	4	C												
8														
9														
10														

Table B.11

Codes for NCTM Students Standards

	State:	NCTM
	Grade:	Grades 6-8

	Rating Standard: NCTM1		Rating Standard: NCTM 2		Rating Standard: NCTM 3		Rating Standard: NCTM 4		Rating Standard: NCTM 5		Rating Standard: NCTM 6		Rating Standard: NCTM 7	
	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand	Topic	Cogn. Demand
1	1	C	3	B	6	B	3	C	1	E	4	E	1	E
2	1	B	3	C	6	C	4	C						
3	1	D	4	B	7	B								
4			4	C	7	C								
5														
6														
7														
8														
9														
10														

APPENDIX C

Statistical Knowledge for Teaching Assessment

MICHIGAN STATE UNIVERSITY
DEPARTMENT OF MATHEMATICS

January, 2002

Dear Prospective Teacher:

I am a doctoral student in Mathematics Education at Michigan State University. I am conducting a study as part of my dissertation about the conceptions of prospective teachers about data analysis and statistics as they relate to the teaching of these subjects.

Please answer the following questionnaire to the best of your knowledge without consulting another person or any other resources. The approximate time needed to complete the questionnaire is about 50 minutes. Responses will be confidential; that is, your privacy will be protected to the maximum extent allowable by law. Responses will be used only with the purpose of understanding how prospective teachers know and apply statistical knowledge in teaching. In some cases, actual written work may be used in the discussion of results in potential subsequent publications. If this is done, your name will not be used.

You indicate your voluntary agreement to participate by signing your name on the next page and by completing and returning this questionnaire. Please provide your e-mail as well, in case we need to contact you for clarification of your responses. If you have any questions about this study, please contact me by email: sortomar@msu.edu. If you have questions or concerns regarding your rights as a study participant, or are dissatisfied at any time with any aspect of this study, you may contact – anonymously, if you wish – Ashir Kumar, M.D., Chair of the University Committee on Research Involving Human Subjects (UCRIHS) by phone: (517) 355-2180, fax: (517) 432-4503, e-mail: ucrihs@msu.edu, or regular mail: 202 Olds Hall, East Lansing, MI 48824.

Sincerely,

Maria Alejandra Sorto
Mathematics Department
Michigan State University
East Lansing, MI 48824

MICHIGAN STATE UNIVERSITY
DEPARTMENT OF MATHEMATICS

Statistical Knowledge for Teaching Assessment

Instructions

Please write all your responses in this booklet. Some questions ask about "your students."
When answering these questions, imagine you are actually teaching students at a grade
level you hope to teach.
You may use a calculator

Background Questions

Name:_____**e-mail:**_____

1. Male ☐   Female ☐

2. Age range: 19 – 23 ☐   24 – 29 ☐   30 – 35 ☐   over 35 ☐

3. Class: Freshman ☐   Sophomore ☐   Junior ☐

        Senior ☐   Post-grad ☐

4. Discipline major : _____
   Discipline minor : _____

5. Please check the <u>mathematics</u> courses you have taken at the college level.

    ____ MATH 10 : Algebra
    ____ MATH 30: Trigonometry and Analytic Geometry
    ____ MATH 31: Calculus of One Variable I
    ____ MATH 32: Calculus of One Variable II
    ____ MATH 67: Number and Algebra
    ____ MATH 81: Discrete Mathematics
    ____ MATH 131: Eucledian and Non-Eucledian Geometry
    ____ MATH 111: Developing Math Concepts
    ____ MATH 115: History of Mathematics
    ____ STAT 11: Basic Concepts of Statistics and Data Analysis
    ____ STAT 31: Introductory Statistics
    ____ Other: _____

6. Please check the <u>math education</u> courses you have taken at the college level.

    ____ EDUC 86 Teaching Mathematics in the Middle Grades

    ____ Other: _____

7. Please check the <u>education</u> courses you have taken at the college level
    ____ EDUC 65: Introduction to Teaching
    ____ EDUC 66: Planning for Teaching in the Middle Grades
    ____ EDUC 69: Teaching Skills Laboratory
    ____ EDUC 96: Teaching Internship
    ____ EDUC 97: Seminar on Teaching
    ____ Other: _____

8. Are you using a calculator? Yes ☐   No ☐

    If Yes, indicate the brand and model (e.g. TI83) _____

1. The stem-and-leaf plot below shows the number of minutes it takes students in a class to travel to their school.

Minutes to Travel to School

0	3 3 5 7 8 9
1	0 2 3 5 6 6 8 9
2	0 1 3 3 3 5 5 8 8
3	0 5
4	5
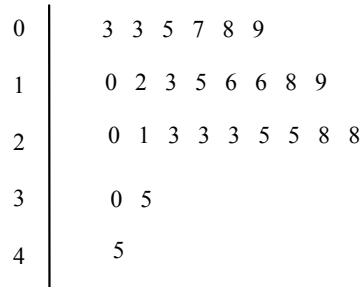
a. How many students are in the class? _____

b. How many students took less than 15 minutes to travel to school? _____

c. What is the typical time it takes for students to travel to school? _____

Explain your answer.

2. Nine students in a science class weighed a small object on the same scale separately. The weights (in grams) recorded by each student are shown below:

6.2   6.0   6.0   15.3   6.1   6.3   6.2   6.15   6.2

The students want to determine as accurately as they can the actual weight of this object.

They may use the following methods:
   I.   Use the most common number, which is 6.2.
   II. Use the 6.15 since it is the most accurate weighing.
   III. Use the result of adding up the 9 numbers and dividing by 9.

As a teacher, what method would you prefer your students to use?

   a.   Method I

   b.   Method II

   c.   Method III

   d.   Other _____

Explain your choice.

3. The following graph gives information about the adult female literacy rates in Central and South American countries.

Adult Female Literacy Rates in Central and South America

a. Suppose you ask your students to tell you how many countries are represented in the graph. One student says, "there are 7 countries represented".

   Is this student right or wrong? _____

   In your opinion, what is the student's thinking to arrive to that conclusion?

b. Suppose now you ask your students to explain what the third bar from the right indicates. One says, "It indicates 85% to 90% literacy rate". Comment on the response.

4. The following line plot shows the number of people in households in a neighborhood.

Number of People in Households

```
            X           X
      X   X   X         X
 _____
  1   2   3   4   5   6   7
```
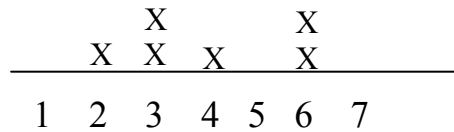
a. Find the mean. Show how you find it.

b. Is it possible to have other sets of data with the same mean?
    Explain why or why not.

c. Is it possible to have a data set of six households with mean $3\frac{1}{2}$
    people?
    If yes, give an example.

    If not, explain why.

5. One middle school class generated data about their pets shown below.

Pet	Frequency
bird	2
cat	4
cow	2
dog	7
duck	1
fish	2
goat	1
horse	3
rabbit	3

a. Give a possible question the teacher could have asked
the students to generate the data.

b. Students were talking about the data and one said:
" The mode is dogs, the median is duck, and the range is 1 to 7."

If you think the student is right, explain why.

If you think the student is wrong, identify the mistake(s).

6. A middle school class was wondering how much time it took each student to travel to school each morning and collected the following data.
Students are asked to make a graphical display of the data to show how much time the majority of students take to travel to school.

Student's initials	Time in minutes
DB	60
DD	50
SE	35
AE	30
FH	25
CL	25
DR	22
BN	20
VH	20
IW	17
AS	15
KS	15
VC	15
AS	10
MS	8
RS	5

a. Which one of these plots seems the most appropriate for the data?

☐ Histogram   ☐ Stem-and-leaf Plot

☐ Bar Graph   ☐ Other _____

Explain your choice.

b. Make the plot picked in part a.

7. Imagine that two second-grade students in the same class have created the following representations to show the number of teeth lost by their classmates.
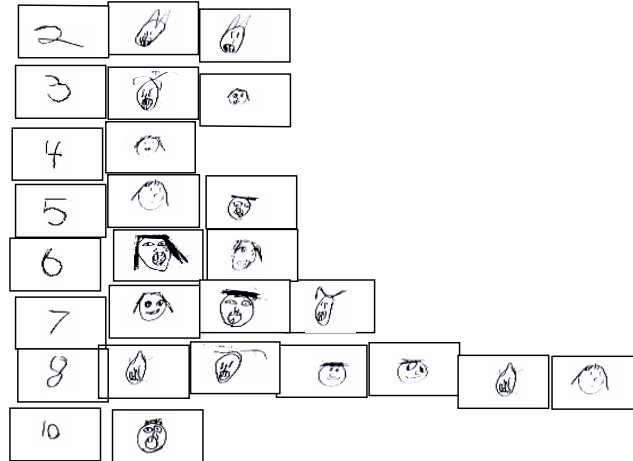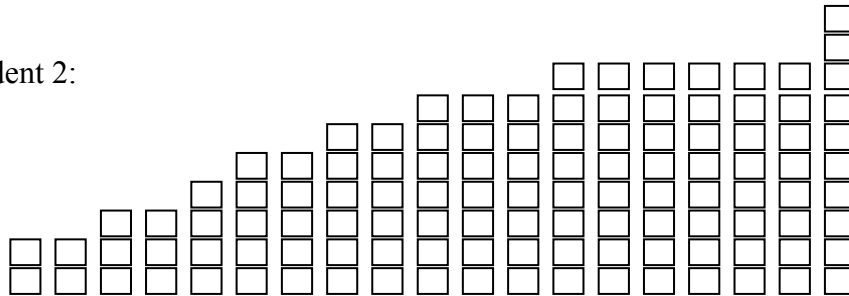
Student 1:

Student 2:

a. Compare and contrast the two representations. That is, how are they alike? How are they different?

b. What do you think each student understands about the data?

8. Consider the dot frequency distributions below with the heights in inches of the 10 players on two basketball teams.
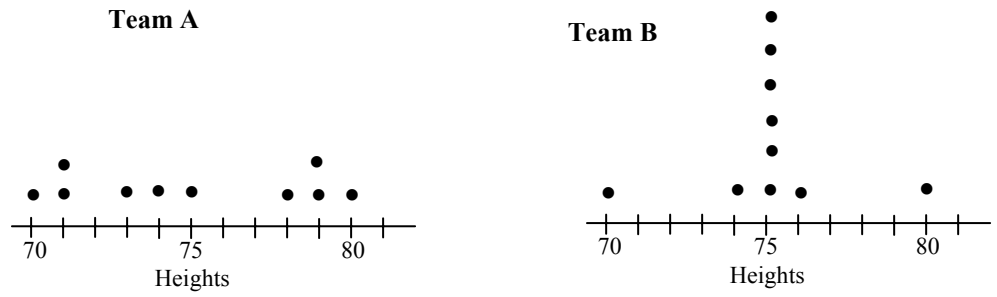
Team A

Team B

a. The mean of each distribution is 75. Calculate the following for each distribution.

	Team A	Team B
Mean	75	75
Range		
Standard deviation		

b. Explain how you could have determine which distribution has the largest standard deviation, without calculating.

APPENDIX D

Item Scores and Statistics for Assessment

Table D.1
Item Scores and Statistics

ID	1a	1b	1c	2	3a	3b	4a	4b	4c	5a	5b	6a	6b	7a	7b	8ai	8aii	8b	Sum	%	Pure Stat	For Teach	Literacy	Reasoning	Thinking
1	4	4	3	4	4	4	4	4	4	4	2	4	4	2	0	3	0	3	57	79	85	67	75	100	75
2	4	4	4	3	2	0	4	2	4	4	1	3	4	1	0	4	0	0	44	61	75	33	80	80	50
3	4	4	4	3	4	4	4	3	4	4	0	1	4	2	1	3	3	0	52	72	77	63	90	75	50
4	4	4	3	2	2	1	0	0	0	0	1	3	4	0	0	4	0	0	28	39	50	17	60	45	37.5
5	4	4	2	1	4	4	4	3	4	3	0	3	4	2	1	4	4	4	55	76	85	58	100	75	75
6	4	4	2	4	4	4	3	0	4	4	1	3	2	2	1	4	0	0	46	64	63	67	65	75	25
7	4	4	4	4	4	4	4	2	4	4	0	1	4	2	2	3	4	3	57	79	85	67	95	75	87.5
8	4	4	3	4	4	4	4	4	3	4	3	3	2	2	3	4	0	4	59	82	81	83	70	90	87.5
9	4	4	4	4	4	4	4	4	4	3	1	4	4	2	2	4	4	0	60	83	92	67	100	100	50
10	4	4	4	3	1	0	4	3	4	3	1	0	2	3	3	4	1	0	44	61	69	46	75	70	50
11	4	4	3	4	4	3	4	0	4	0	0	2	4	2	2	3	0	0	43	60	67	46	75	70	37.5
12	0	0	0	4	2	3	4	1	4	0	0	0	4	1	2	4	0	2	31	43	48	33	60	45	25
13	4	4	4	4	3	3	4	3	4	4	0	0	4	2	2	4	4	3	56	78	88	58	100	75	87.5
14	4	4	4	2	4	4	4	2	4	3	1	0	4	2	1	3	3	4	53	74	79	63	90	60	100
15	0	0	0	0	0	1	4	1	0	0	1	0	4	2	1	4	0	0	18	25	27	21	60	5	0
16	4	4	1	2	1	0	4	3	2	0	0	2	2	3	2	3	2	0	35	49	60	25	75	65	12.5
17	4	4	4	2	4	0	4	3	4	4	0	3	4	2	0	4	0	0	46	64	75	42	80	80	50
18	4	4	2	3	4	4	4	1	3	0	0	3	4	3	0	3	2	1	45	63	71	46	85	70	37.5
19	4	4	3	0	4	0	3	4	2	4	0	2	4	2	1	3	0	0	40	56	60	46	70	60	37.5
20	4	4	4	3	4	1	2	1	4	4	0	0	4	1	2	3	0	0	41	57	60	50	65	60	50
21	4	4	3	3	1	2	2	2	2	3	0	1	4	1	0	3	0	2	37	51	63	29	65	60	62.5
22	4	4	2	2	4	4	3	3	4	4	1	3	3	1	0	4	2	1	49	68	73	58	80	80	37.5
23	0	0	0	4	3	2	0	0	0	3	0	0	0	1	1	0	0	0	14	19	8	42	0	20	0
24	4	4	1	0	4	3	3	1	0	4	1	2	4	0	0	3	0	0	34	47	46	50	70	35	12.5
25	4	4	3	3	1	0	2	3	0	0	0	0	0	0	0	4	0	0	24	33	48	4	50	50	37.5
26	4	4	4	2	4	4	4	2	4	3	1	1	3	2	3	3	2	4	54	75	77	71	80	65	100
27	4	4	1	1	1	2	0	0	0	0	0	0	0	2	0	3	0	0	18	25	27	21	35	25	12.5
28	4	4	3	4	2	4	4	3	4	4	1	3	4	1	0	3	1	0	49	68	77	50	80	90	37.5
29	4	4	3	1	2	0	4	1	0	4	1	1	4	0	0	3	0	0	32	44	52	29	75	35	37.5

248

30	4	4	2	1	4	4	3	3	4	4	1	1	4	2	1	3	2	4	51	71	73	67	80	65	75
31	4	4	1	2	3	0	2	2	2	4	0	3	2	0	0	4	0	2	35	49	58	29	60	65	37.5
32	4	4	1	2	4	4	3	3	4	0	0	0	4	2	0	3	1	0	39	54	60	42	75	65	12.5
33	4	4	1	3	4	0	4	3	4	0	2	2	4	3	2	3	0	0	43	60	67	46	75	80	12.5
34	4	4	1	1	1	3	4	2	3	0	1	2	4	2	0	3	4	4	43	60	75	29	95	60	62.5
35	0	4	2	0	4	0	4	3	0	0	1	4	4	2	0	0	3	0	31	43	50	29	55	55	25
36	4	4	4	0	2	0	4	2	0	0	0	2	0	0	0	3	4	3	32	44	63	8	75	40	87.5
37	4	4	1	4	4	1	4	4	4	4	1	3	4	2	3	3	4	1	55	76	83	63	95	95	25
38	4	4	4	0	4	4	4	4	4	0	1	0	0	2	2	3	4	3	47	65	71	54	75	60	87.5
39	4	4	4	0	4	0	4	2	4	4	1	3	4	2	2	4	2	4	52	72	81	54	90	65	100
40	4	4	4	0	2	0	3	3	4	4	1	0	0	0	0	3	2	3	37	51	63	29	60	55	87.5
41	4	4	4	0	2	0	4	4	3	4	1	4	4	4	2	3	4	3	54	75	85	54	95	75	87.5
42	4	4	4	0	2	0	4	3	0	4	1	0	0	2	1	3	4	4	40	56	63	42	75	35	100
Mean	4	4	3	2	3	2	3	2	3	3	1	2	3	2	1	3	1.6	1	42	59	**66**	**45**	**74**	**63**	**51**
Median	4	4	3	2	4	2	4	3	4	4	1	2	4	2	1	3	1	1	44	60	**68**	**46**	**75**	**65**	**50**
SD	1	1	1	2	1	2	1	1	2	2	1	1	2	1	1	1	1.7	2	12	16	**18**	**18**	**19**	**21**	**30**

APPENDIX E

Interview Protocols

Interview Protocol for subject A

1. In question 1c you the median for typical, what would you say is the typical time for this data?

Minutes to Travel to School

0	3 3 5 5 6 7 7 8
1	0 0 1 4 8 9
2	1 1 2 3
3	5 5
4	3 3 5 6 7 8

2. Suppose that there are two modes, how would you reconcile that?

3. There is no way to find out how many countries are represented? What does "frequency" means in this case?

4. Suppose you did not know the algorithm of the mean, how would you estimate the mean from the picture? How would you explain 4b?
   How would you interpret 3.5 people as the mean of the data?

   Take a look at this other set of data, can you estimate the mean?

```
        X                           X
        X                           X
        X           X               X
        X   X   X   X               X               X
        X   X   X   X           X   X               X
        X   X   X   X   X       X   X   X           X           X
   ─────────────────────────────────────────────────────────────
   26  27  28  29  30  31  32  33  34  35  36  37  38  39  40
```

Number of Raisins in a Box

5. Interpret what "bird 2" means for your question.

6. You would need to list students initials to do the bar graph or the histogram?
     So, how much time the majority of students take to travel to school?

7. You did not answer this one, do you want to think out loud how would you answer this questions?
Statistically speaking, how would you rank these students?
You said that the second student does not understand order, what do mean by that?

8. Do you calculations make sense? What do these numbers mean? How did you get them?

 Which team would you bet on? and why?

Interview Protocol for subject B

1. In question 1c you the median for typical, what would you say is the typical time for this data?

Minutes to Travel to School

0	3 3 5 5 6 7 7 8
1	0 0 1 4 8 9
2	1 1 2 3
3	5 5
4	3 3 5 6 7 8

2. Why would the mean be more accurate?

3. There is no way to find out how many countries are represented? What does "frequency" means in this case?
 What would be the ideal response here?

5. Suppose you did not know the algorithm of the mean, how would you estimate the mean from the picture? How would you explain 4b?
 How would you interpret 3.5 people as the mean of the data?

Take a look at this other set of data, can you estimate the mean?

```
    X                   X
    X                   X
    X       X           X
    X   X   X   X       X               X
    X   X   X   X       X   X           X
    X   X   X   X   X   X   X   X       X           X
   26  27  28  29  30  31  32  33  34  35  36  37  38  39  40
```
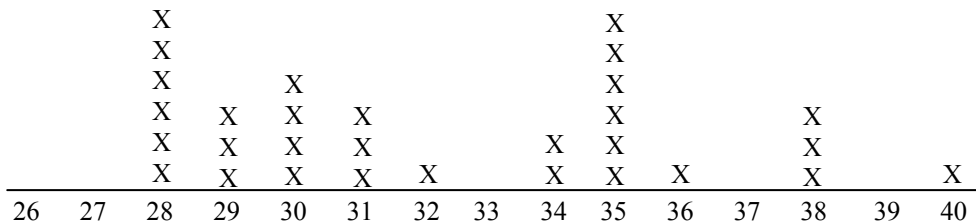
Number of Raisins in a Box

5. What would be the actual question? Interpret what "bird 2" means for your question.
 So, what would be the right response for the mode, median, and range?

6. How about in terms of the data? Does the data tell you what display to use?
 So, how much time the majority of students take to travel to school?

7. You did not answer this one, do you want to think out loud how would you answer this questions?
Statistically speaking, how would you rank these students?
You said that the second student does not understand order, what do mean by that?

8. Do you calculations make sense? What do these numbers mean? How did you get them?

 Which team would you bet on? and why?

Interview Protocol for subject C

1. In question 1c you picked a range of values for typical, what would you say is the typical time for this data?

Minutes to Travel to School

0	3 3 5 5 6 7 7 8
1	0 0 1 4 8 9
2	1 1 2 3
3	5 5
4	3 3 5 6 7 8

2. Were you going to say more about method III?, what about method I and II?

3. How would you correct this student?
   It seems that you did not have enough time to finish part b, what you tell the student?

4. Suppose you did not know the algorithm of the mean, how would you estimate the mean from the picture? How would you explain 4b?
   How would you interpret 3.5 people as the mean of the data?

   Take a look at this other set of data, can you estimate the mean?

```
    X                                  X
    X                                  X
    X           X                      X
    X     X     X     X                X                 X
    X     X     X     X          X     X                 X
    X     X     X     X     X     X     X     X           X                 X
 _____
   26    27    28    29    30    31    32    33    34    35    36    37    38    39    40
```

Number of Raisins in a Box

5. Interpret what "bird 2" means for your question.
   You wrote up here that the median is 12.5 but you say down here that the median is horse, which one is it?

6. So, how much time the majority of students take to travel to school? What do you mean by rightly represented?

7. You did not answer this one, do you want to think out loud how would you answer this questions?

Statistically speaking, how would you rank these students?
You said that the second student does not understand order, what do mean by that?

8. Do you calculations make sense?
   You mention that are positive, does this tell you something? You say that "s.d. is 7 away from the mean", 7 what?

 Which team would you bet on? and why?

1. In question 1c you picked the mean for typical, what would you say is the typical time for this data?

Minutes to Travel to School

0	3 3 5 7 8 9
1	0 2 3 5 6 6 8 8 9
2	0 1 3 3 3
3	1
4	
5	
6	2

2. What about method I or method II?

3. The bars are not countries, so what do they represent?
 How do you think we can correct this mistake?
What would be the ideal response you are expecting the student to answer?
You said the graph is misleading, in what way?
So, is the student's response wrong?

4.   For 4b, why is it possible?
     For 4c, you say the only way possible is if some of the numbers are 21/2 and so on?
Suppose you did not know the algorithm of the mean, how would you estimate the mean from the picture? How would you explain 4b?
How would you interpret 3.5 people as the mean of the data?

   Take a look at this other set of data, can you estimate the mean?

```
     X                           X
     X                           X
     X        X                  X
     X   X    X    X             X                X
     X   X    X    X        X    X                X
     X   X    X    X    X    X   X    X            X            X
  _____
  26   27   28   29   30   31   32   33   34   35   36   37   38   39   40
```

Number of Raisins in a Box

5. Interpret what "bird 2" means for your question.

You can see how students come to this conclusions, but is the student correct?

6. Is this a histogram?

7. Tell me what student 1 understand? Do you think questions about the number of teeth lost can be answer by any of the two representations?
Statistically speaking, how would you rank these students?

8. Do you calculations make sense?

 Which team would you bet on? and why?

Interview Protocol for subject E

1. Can you show me how you got 87 minutes for the typical time?
   In question 1c you picked the average for typical, what would you say is the typical time for this data?

Minutes to Travel to School

0	3 3 5 7 8 9
1	0 2 3 5 6 6 8 8 9
2	0 1 3 3 3
3	1
4	
5	
6	2

2. Why not the most common or the most accurate?

3. The bars are not countries, so what do they represent?
   How do you think we can correct this mistake?

4. Suppose you did not know the algorithm of the mean, how would you estimate the mean from the picture? How would you explain 4b?
   How would you interpret 3.5 people as the mean of the data?

   Take a look at this other set of data, can you estimate the mean?

```
        X                           X
        X                           X
        X           X               X
        X     X     X     X         X                 X
        X     X     X     X         X     X           X
        X     X     X     X     X   X     X     X     X           X
_____
  26    27    28    29    30    31    32    33    34    35    36    37    38    39    40
```

Number of Raisins in a Box

5. Interpret what "bird 2" means for your question.
   What would be a correct way to order the data? What would the median be?

6. What do you mean by "group into range"? Why not the others?

7. Explain to me again how are they alike and different?

Do you think questions about the number of teeth lost can be answer by any of the two representations?
Statistically speaking, how would you rank these students?

8. You said "that number" which one? The SD? How did you find these numbers? Do they make sense?

 Which team would you bet on? and why?

Interview Protocol for subject F

1. In question 1c you picked the mode for typical, what would you say is the typical time for this data?

Minutes to Travel to School

0	3 3 5 5 5 9
1	0 2 3 5 6 6 8 8 9
2	0 1 3 3 4
3	1
4	2

2. Why is it 'unfair' to use the other methods?
    What do you mean by "these numbers may not be the actual weight of this object"?

3. The bars are not countries, so what do they represent?
 How do you think we can correct this mistake?

4. Suppose you did not know the algorithm of the mean, how would you estimate the mean from the picture? How would you explain 4b?
 How would you interpret 3.5 people as the mean of the data?

   Take a look at this other set of data, can you estimate the mean?

```
     X                              X
     X                              X
     X        X                     X
     X    X   X   X                 X                   X
     X    X   X   X             X    X                   X
     X    X   X   X    X        X    X    X              X              X
  _____
  26   27   28   29   30   31   32   33   34   35   36   37   38   39   40
```

Number of Raisins in a Box

5. Interpret what "bird 2" means for your question.

6. What do you mean by "continuing"? Explain how is this a histogram. Why not the others? Can you tell from here how much time the majority of students take to travel?

7. Do you think questions about the number of teeth lost can be answer by any of the two representations?

Statistically speaking, how would you rank these students?

8. How did you find these numbers? Do they make sense?

 Which team would you bet on? and why?

Interview Protocol for subject G

1. In question 1c you picked the average for typical, what would you say is the typical time for this data?

Minutes to Travel to School

0	3 3 5 7 8 9
1	0 2 3 5 6 6 8 8 9
2	0 1 3 3 3
3	1
4	
5	
6	2

2. What do you mean by "many of the data are varied"?
   When you say that it is a mathematical way, you mean…

3. You say that the bars represent a range of countries between certain literacy rates, so how many countries are represented here?
   How do you think we can correct this mistake?

4. Why a 'clump' data about 4 would have the same mean, what about spread out?
   How would you interpret 3.5 people as the mean of the data?

   Take a look at this other set of data, can you estimate the mean?

```
     X                               X
     X                               X
     X           X                   X
     X    X    X    X                X                    X
     X    X    X    X           X    X                    X
     X    X    X    X    X       X    X    X              X              X
   ─────────────────────────────────────────────────────────────────────
    26   27   28   29   30   31   32   33   34   35   36   37   38   39   40
```

Number of Raisins in a Box

5. Interpret what "bird 2" means for your question.

   What would be a correct way to order the data? What would the median be?

7. Do you think questions about the number of teeth lost can be answer by any of the two representations?

 Statistically speaking, how would you rank these students?

263

8. Which team would you bet on? and why?

Interview Protocol for subject H

1. In question 1c you the mode for typical, what would you say is the typical time for this data?
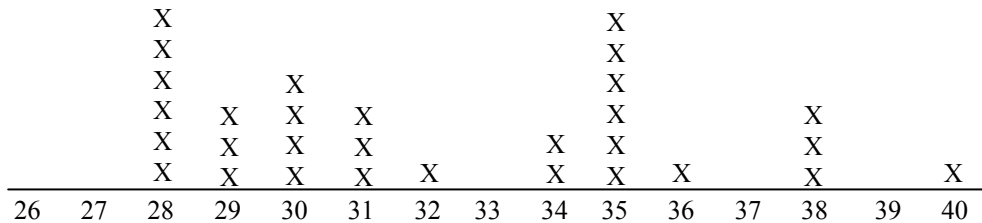
Minutes to Travel to School

```
0  |  3  3  5  5  5  9
1  |  0  2  3  5  6  6  8  8  9
2  |  0  1  3  3  4
3  |  1
4  |  2
```

2. So, you want all the data to be represented? Take a closer look at the data.

3. There is no way to find out how many countries are represented? What does "frequency" means in this case?

4. Suppose you did not know the algorithm of the mean, how would you estimate the mean from the picture? How would you explain 4b?
   How would you interpret 3.5 people as the mean of the data?

   Take a look at this other set of data, can you estimate the mean?

```
X                           X
X                           X
X              X            X
X     X   X    X            X                   X
X     X   X    X       X    X                   X
X     X   X    X   X   X    X    X         X          X
26  27  28  29  30  31  32  33  34  35  36  37  38  39  40
```

Number of Raisins in a Box

5. Interpret what "bird 2" means for your question.
   Explain more how to find the median here.

6. So, how much time the majority of students take to travel to school?

7. You did not answer this one, do you want to think out loud how would you answer this questions?
Statistically speaking, how would you rank these students?
You said that the second student does not understand order, what do mean by that?

8. Do you calculations make sense? What do these numbers mean?
   You mention that are positive, does this tell you something? You say that "s.d. is 7 away from the mean", 7 what?

 Which team would you bet on? and why?

Interview Protocol for subject I

1. In question 1c you picked the median for typical, what would you say is the typical time for this data?

Minutes to Travel to School

0	3 3 5 5 6 7 7 8
1	0 0 1 4 8 9
2	1 1 2 3
3	5 5
4	3 3 5 6 7 8

2. In this case, is it important to take all the number into account? What about method II?

3. The bars are not countries, so what do they represent?
 How do you think we can correct this mistake?
What would be the ideal response you are expecting the student to answer?

4. Suppose you did not know the algorithm of the mean, how would you estimate the mean from the picture? How would you explain 4b?
How would you interpret 3.5 people as the mean of the data?

   Take a look at this other set of data, can you estimate the mean?

```
    X                           X
    X                           X
    X         X                 X
    X    X    X    X            X                   X
    X    X    X    X       X    X              X
    X    X    X    X    X  X    X    X         X         X
 _____
   26   27   28   29   30   31   32   33   34   35   36   37   38   39   40
```

Number of Raisins in a Box
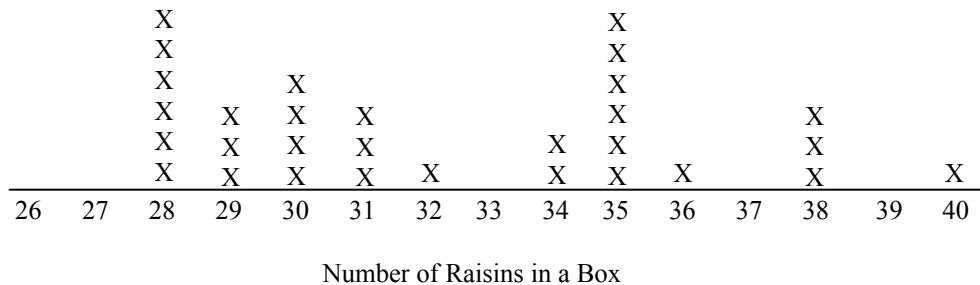
5. Interpret what "bird 2" means for your question.
   Is the mode and range correct?
   What is the meaning of the fish as median?

6. What about a histogram?

7. Do you think questions about the number of teeth lost can be answer by any of the two representations?
Statistically speaking, how would you rank these students?

8. Do you calculations make sense?
    You said "these facts also are true because the mean and range are the same", does that mean that if the mean and range weren't the same…

 Which team would you bet on? and why?

1. In question 1c you picked the mean for typical, what would you say is the typical time for this data?

Minutes to Travel to School

0	3 3 5 7 8 9
1	0 2 3 5 6 6 8 8 9
2	0 1 3 3 3
3	1
4	
5	
6	2

2.

3. When you say "yes and no" you mean that the student is right and wrong? Explain.
   So, he/she is not taking the correct range?

5. For 4b, you gave an example, how would you show that there are many data sets?
   For 4c, for this context it is not possible, right?
Suppose you did not know the algorithm of the mean, how would you estimate the mean from the picture? How would you explain 4b?
How would you interpret 3.5 people as the mean of the data?

   Take a look at this other set of data, can you estimate the mean?

```
    X                       X
    X                       X
    X           X           X
    X   X   X   X           X               X
    X   X   X   X       X   X               X
    X   X   X   X   X   X   X   X           X           X
   ─────────────────────────────────────────────────────
   26  27  28  29  30  31  32  33  34  35  36  37  38  39  40
```

Number of Raisins in a Box

5. Interpret what "bird 2" means for your question.
   What about the median?

6. So, how much time the majority of students take to travel to school with your line graph?

7. Do you think questions about the number of teeth lost can be answer by any of the two representations?
Statistically speaking, how would you rank these students?
You said that the second student does not understand order, what do mean by that?

8. Do you calculations make sense?
   What about Team B?

 Which team would you bet on? and why?

APPENDIX F

Interview Transcripts

Subject B

Question 1

Researcher: Part c of question 1 refers to what is the typical time for students to travel to school, and I am curious to know what you think "typical" means.

Student: It could mean any kind of a center, to me it could mean median, or it means…you know…not normally I don't look at mode, but it could mean mode.

Researcher: So, you picked the median.

Student: I thought it was easier, because it is already organized.

Researcher: OK, so, let me ask you the same question for this other set of data.

Student: I would find the median, it is organized, so…

Researcher: Would you still stick with the median?

Student: Ummm…[thinking]…it is about 18, right. To me it is easier if it is organized or some kind numerical order.

Question 2

Researcher: You said that you chose method III (the mean) because it is more accurate, what do you mean by that?

Student: because it's clustered all around a close sort of range. I though OK with the mean…you know…I though the mean and the median might be close in this situation, but some times when you have a really, really high, like say you have a score of 8.5, I might want to say I want to ignore the 8.5 and do median instead. I say that the extreme score is going to weight the mean more than it would with the median.

Researcher: and in this case…

Student: In this case the mean and the median are going to be really close, because the data are all cluster together.

Researcher: around…cluster around…

Student: Well, you have the one… that's what brought me to think about it. Like I said, well, even though I say maybe we want to take that one into account, because sometimes you do…so I said, well it might be wired but maybe we should take him into account . So I thought maybe in this situation you

want to. That's why I said the mean, I just didn't know. They are all cluster except for that one guy, that's what made me think, do I really want to take it into account or not.

Question 3

Researcher: When you say that "we don't know" if the student is right or wrong, do you mean that we don't have enough information in the graph to say how many countries are represented?

Student: We assume that we have data for everybody, but apparently we didn't . Sometimes you have something missing, like I though, OK there is quite a bit missing here [pointing at the gap between 50% and 70%] why do we lay out our histogram this way, and I though, maybe there is something where there was no literacy rate…some country where there is zero.

Researcher: That's why those gaps

Student: You never know…I though it was a strange histogram. Normally, you would just said, well you have a bar for every thing, but you don't always know that, right? When you have a gap, you don't always know about your gap. This bother me a little, this gap.

Researcher: Why do you think the student say 7?

Student: I don't know, because he wasn't really looking …well, he was only looking at one bar or two and then saying…you know what I mean. I couldn't figure…error with kids always bother me…in math…I look at the error and I go…it is hard to know what they mean.

Question 4

Researcher: How do you know there are more data sets with the same mean?

Student: If you want to have another data set with the same mean 4, all you do, you can move one up and one down or two up and two down. You have to keep it equal, the same number.

Researcher: Estimate the average with this data set? What would the kid would say?

Student: Their best guess would be 32 or 33 because half are here and half are over here.

Researcher: What is the meaning of 3.5?

Student: That is the mean, you take all the numbers and add them all up and divide by the number you have.

Researcher:	Suppose you are explaining to a kid. Would the kid be bother by the 3.5?
Student:	I don't think so, if you explain it in certain way.
Researcher:	Like what?
Student:	If you say, well nobody really have 4.5 kids, but over a big data set some people have 2 kids, some people have 3, some people have 6, some people have 8. So over a big data set it could be 4.7, it could be 3.2, it just means over a big data set that is your average. I think kids understand that. You have to say, well, you take an average and the average means that that matters, that every one matters, that's why you would get 4.7 kids. It just means that people are all over the range, some people have 1, some people have 2, 3, 4, 5, 10.
Researcher:	Now, I am going to play the kid, the kid would say, why does the answer come out to be a whole number?
Student:	I think you have to teach them about the median, maybe, that there is another number that could be a whole number, but not always. Even the median could be 4,5, but…

Question 5

Researcher:	What is the question the teacher asked?
Student:	How many people have…?
Researcher:	What does "bird 2" means"
Student:	That would mean that two people have one bird OR one people have two birds.
Researcher:	For question b, is the student completely wrong?
Student:	He is analyzing the data as continuous and it is a discrete data. For pets, you can't really have a median or mode, it is discrete data…I don't see that, because five people could have count for all the pets and does not mean anything.
Researcher:	What do you think about the range?
Student:	We don't even know that, we don't even know what the range is of the number of pets. That doesn't tell us, we don't know how many pets does the average student has, so we don't know what the range is. Is the range 0 to 10 pets? We don't know that, it does not tell us that.

Question 6

Researcher: Tell me why did you pick the steam-and-leaf plot?

Student: I thought the histogram is easier, but I always do the steam-and-leaf plot before I do the histogram because I want to see how does the data actually clusters. Cause there are so many in two and them I might I would have to split my groupings slightly so I wouldn't have so many in just one part of the histogram.

Researcher: How about in terms of the data? Would you do the same for another set of data?

Student: It depends on the students' ability level, but you can do a histogram.

Researcher: What can you tell me about the bar graph?

Student: Well, time is continuous in general, it's a measure as suppose as discrete data. I don't know, you could do it I suppose. I wouldn't tell a child no you can't ever try this. They can try it and see how it works, right?

Researcher: How much time the majority of students take to travel to school?

Student: There in their 20s

Question 7

Researcher: What type of graphs?

Student: There are bar graphs.

Researcher: What is the total number…?

Student: Student 1 has an easier time counting them up, because is a pictograph and in second grade they are more into pictograph, they are more into seeing it visually. So he or she can just count.

Researcher: Assessing?

Student: I would actually say student 1. Student 2 does not have anything that tells me what it is. If they have told me a key, or if they have told me what this axes represents here and what this represents I would have an easier time to…but they organize it beautifully. I like this second grade work, it shows that they did it themselves or in a little group and they really though about it.

Researcher: You want students to understand each other?

Student: You would have to talk to each other and give advice to each other.

Question 8

Researcher: Do your answers make sense?

Student: It does, this stuff is all scatter so it is a higher standard deviation. This stuff, you only have two aberration and the rest is all cluster right at the mean, so it makes sense that one would be smaller than the other.

Researcher: What does 2.14 means?

Student: 2 untis, it means that from your mean one standard deviation takes into account 68% of your data and two standard deviations take a 95%, that's all it means.

Researcher: If you were a coach…

Student: Either one, you know what? I know this is going to sound really strange but I would probability would like to coach Team A. Because I wouldn't mind coaching all different ability levels, because I am just different. Whereas these two extremes, I would feel so bad for this kid that is way up here and so bad for this kid that is way down here. Sure it would be easier to coach these kids (Team B) it would be horrible, this kid would be always be playing and this kid would never be playing.

Subject C

Question 1

Researcher: In question 1c you picked a range of values for typical, why?

Student: I think, I use the mean, try to find the mean.

Researcher: What is typical for this other data set?

Student: I'll use the median, because we have two groups with the same number, so I can't just pick one.

Question 2

[Time was limited for this subject and no questions were asked here]

Question 3

Researcher: How do you tell the student that he/she is wrong?

Student: The graphs only shows the literacy rate, from the graph there is no way to find out how many countries.

Researcher: What does the third bar represents?

Student: The third bar represents 85% to 90% literacy rate for 3…

Question 4

Researcher: Suppose you are introducing the concept of the mean, how would you estimate the mean without reaching to the algorithm? Show raisin plot.

Student: Some of them would say to equate the stacks to have at the same level, and then I want them to say that the mean will be each of the 3 boxes, the high of the boxes 3,3,3…

Researcher: Is it possible to have a data set of 6 households with a mean 3.5 people?

Student: I can't have half of person, it depends on ….

Question 5

Researcher: You wrote up here that the median is 12.5 but you say down here that the median is horse, which one is it?

Student: I try to bring the number in order, sum the number which is 25 to find how many pets, then divided by 2 because is odd, and counted 12 and 12 from each end and pick the middle

Question 6

Researcher: Which one is easier to understand?....

               [Time run out and interview had to be suspended]

Subject E

Question 1

Researcher: How did you get 87?

Student: I added the times3, 3, …, and then divided by 26.

Researcher: Does that make sense to you?

Student: No, it does not. I think I did something wrong…

Researcher: I am going to show you another data set. What will be the typical time here?

Student: Add them up, 3 plus 3, plus 3,…, and divided by 21. The answer is 17.85

Question 2

Researcher: Why not method I?

Student: You could, this is just another way to find the average. It depends of the average you want, it is not wrong. As a teacher you have to teach and accept other ways. But I would prefer the method I choose.

Question 3

Researcher: What do you want the student to say the bars represent if not countries?

Student: The percent of Adult…

Researcher: How would you correct the student?

Student: The student would not able to actually say a number, because you don't know how many countries are there in Central and South America on top in your head.

Question 4

Researcher: Can you arrive to the answer without knowing the algorithm?

Student: They could go…count one from each side [crossing out an X on each side]…they can take the average of the two left over and the answer will not be 4 but very close to it.

Researcher: How about with a bigger set?

Student:	Do the same…cross out one on each side, one, one, two, two,…[cross out one X on each side and ends up with two X's under the number 31], you have two 31's, then you would not have to divide by two. The answer is 31.
	This is the median.
Researcher:	How about the mean? How can the students see the picture and estimate the mean?
Student:	It would not have to be exact?
Researcher:	No, but we want to find a good estimate.
Student:	[struggle here]
Researcher:	So the median would be easier to estimate for students than the mean?
Student:	That is what it seems to me…because I can't do it off top of my head…Maybe arrange the X's …like I would put this X down here and see how they average it out…that is the only way I can see how to do it. I would send them to the board and have them arrange the X's and make them line them up and that will be the average.
	[show work]
	Move these over here, and erase these.
	All the X's are even now, that would be three. Humm…[trying to make sense] I guess that is 3 raisins … hold on…[more thinking]…I am trying to figure out…
	You could do adding up again and then divide. You add 28, 28, and 28 or 28 times 3.
Researcher:	It is possible to have other set with the same mean?
Student:	You have one house with 2 people, two house with 3 people, you got one house with 4 people and two houses with 6 [see picture]. Then you got, count the X's, 24 people and let them write circles or something else to even it out, then you get the answer.
Researcher:	How did you come up with this numbers?
Student:	I figured up the total by multiplying 3.5 by 6, that is 21. Then divide that into six numbers.
Researcher:	What does that mean?
Student:	That would say that you have some range of number of household, a lower range and an upper range and 3.5 is in the middle.

Question 5

Researcher: What does "bird 2" means?

Student: One kid has 2 birds or two kids have one bird.

Researcher: What would be the right order to find the median?

Student: We have one dog, one goat, two fish,
               ….(dgffcocobbhhhrrrcacacacadodododododo…) and then pick the middle.

Researcher: We have the mode and the median to be a pet and the range a number.

Student: Right.

Question 6

Researcher: What do you mean by "group into range"?

Student: I though it would be easier to see how many students fall into groups of 0
               to 10 and 11 to 20, and so on.

Researcher: Why not the others?

Student: I though with the histogram would be easier to show the minutes…it is just
               a way.

Question 7

Researcher: How are they different?

Student: This student is at the level where he doesn't have to represent each tooth,
               he can just see in his head and say, ok I am going to draw a picture and he
               is going to be one person and that's gonna represent 2 and this one gonna
               be another person and it is going to represent 2 teeth that he lost. But this
               one, he may be in a lower grade…he does not have to be behind…and he
               just goes one two, one two, that is one person and another person that lost
               two tooth… the same information…it is just organized in different way.

Researcher: Would both graph answer the question what is the total number of tooth
               lost?

Student: This would have to count 2 and 2, 3 and 3, 4, 5. But this one have to
               actually add.

Researcher: What is the most number of teeth lost?

Student:	This one would have to look this one…but this one would have to look how many people this way.
Researcher:	I would prefer this student to get at this because is more advance thinking.

Question 8

Researcher:	Do the answers make sense?
Student:	Yes, I think my answers are right.
Researcher:	Because…what does the standard deviation tell you?
Student:	It is the distance away from the average.
Researcher:	Does it make sense that this one is bigger?
Student:	Huumm, yeah because these are more the same…and these come down and out, I guess, so that it become…I am not…I want to say that make sense. But I would actually do it in paper to see if it is correct.
Researcher:	So, you are saying that number represent how far each data is away from the mean. So 3.57 would mean that…
Student:	That the average will go ???….that far away.
Researcher:	Which team would you bet on?
Student:	I don't want to bet in team B just because they are taller. Team A is more like average, every one is more like in the middle.

Subject F

Question 1

Researcher: For the first question, if I give you another set of data and ask you the same question.

Student: 5 because more people took 5 minutes to travel to school. Typical means the most common number.

Question 2

Researcher: You mention that you take the mean because "it is unfair" to use the other method. Why do you say that?

Student: because if you take any of the number of weights, it is unfair.

Researcher: Why? What do you mean?

Student: Maybe the object is not that weight.

Researcher: Look careful to the numbers, do you notice something?

Student: The 15.3

Researcher: What about it?

Student: I think that is not the real weight of the object.

Researcher: Would you still consider the average (method III)?

Student: No. Maybe if not that one, the most common number.

Researcher: Why?

Student: ….can't hear!!!

Question 3

Researcher: If we want to correct this mistake, what would you say to the student?

Student: I would tell them to look at the frequency to see how many countries. There are 15 countries.

Researcher: What does each bar represent if they do not represent countries?

Student: The percent of adult female literacy.

Researcher: So, this person that said that the third bar from the right indicates 85% to 90% literacy rate, is he right?

Student: It needs to say that <u>three countries</u> have …

Question 4

Researcher: Can you find the mean in another way? Look at this data set, can you estimate the mean?

Student: I would pick the middle one, it looks like 32 for this data.

Researcher: Is it possible to have other sets of data with the same mean? (Do not use the algorithm)

Student: You can have five numbers and all those numbers are 4, then the mean is 4. Or you can change the set with 3, 4, 5 and still have mean 4.

Researcher: What does 3.5 people tell you?

Student: The average of number of people.

Researcher: Suppose someone does not know what "average" means, what would you tell that person? In other words, what does the statement "the average number of people per household is 3.5" mean?

Student: It means that the total number of people in each household divided by the number of households is 3.5.

Researcher: You are reaching for the way you got the number, for the algorithm. If someone does not know the algorithm, it would need another way to explain what "average means". How would you explain?

Student: I would tell them to look at the dictionary and find the definition.

Researcher: Ha, ha, What would you think the dictionary would say?

Student: That the average is the way to describe the mean.

Question 5

Researcher: What question was asked to generate the data?

Student: How many pets do middle school student have? (corrected from How many pets are there?)

Researcher: What does "bird 2" means in the chart?

Student: There are two people with one bird each.

Question 6

Researcher: You picked a histogram because "their datas are continuing", what do you mean by that?

Student: The data follow a continuing pattern

Researcher: As suppose to what? What would be a data that is not continuing?

Student: [silent]

Student: One of the characteristics of histogram is continuing, and time is continuous data.

Researcher: Is there is any other reason why you pick histogram?

Student: [silent]

Researcher: How much time the majority of student…?

Student: 35

Researcher: Show me how you did it?

Student:

Question 7

Researcher: How would you make one student understand the others representation?

Student:

Researcher: Which student understand better?

Student: Student 1

Researcher: Why?

Student: It pictures the number of teeth lost with each student, student 2 is difficult to understand.

Question 8

Researcher: Tell me how you got the numbers?

Student: Calculator

Researcher: Do they make sense? Why?

Student: Yes, the shape of the graphs.

Researcher: If you were going to bet on one, which one?

Student: Team B.

Subject H

Question 1

Researcher: What does typical mean to you?

Student: I PICKED THE MODE

Researcher: What about for this data?

Student: 5, that is the mode for this data.

Question 2

Researcher: You picked the average (mean) for this one, were you aware of the outlier
                and picked the mean to make every data value to count?

Student: Yes.

Question 3

Researcher: You said the student is neither right nor wrong because…

Student: The graph said is for Central and South American countries, it does not
                said which ones.

Researcher: But you do understand why the student said 7 bars?

Student: The student is counting bars.

Researcher: What would you tell the student to lead him to the right answer?

Student: I would tell him to look at the height of the graph and remember always to
                follow the height of the bar to the left to see how tall it is and that is the
                frequency.

Researcher: If you ask the student what does that mean, frequency 3, what would you
                want him to tell you?

Student: The bar indicates 85% to 90% literacy rate for 3 women.

Question 4

Researcher: Think about how would you answer these questions if you were teaching
                average for the first time to your students and they do not know the
                algorithm.

Student:	Maybe they would look at the middle, like 4 or 3 and 6 because they both have two Xs.
Researcher:	What do you think they would say for a bigger set?
Student:	They would say 32 o 33 because they are looking for the middle or some type of balance.
Researcher:	How would you convince the children that is possible to have other data set with the same mean?
Student:	You can start with the same numbers and then switch them around , like if you have 4,4,4 take away one from the first 4 and make it into 3 and then put one more and make a 5. Just playing with numbers.
Researcher:	Can you do it with the bigger data set?
Student:	Let's see…[pause] I guess you should move… I don't how to do it with data, I just know how to do it with numbers. Because you can't move an X to 39 because this represents 3 of 38, isn't?
Researcher:	What does and average of 3.5 people per household mean? How do we interpret this number?
Student:	I guess when the average…many many houses average about 3.5. I don't know what that means, I don't know how to explain what that means though. Because the kids are gonna go "what is 3.5 of a person?"

Question 5

Researcher:	You answer that the data could be generated by asking the question "What type of pets do you have in your home?" So, what does bird 2 means?
Student:	It could mean two different things. It could mean one child has two birds in the house or two different children have one bird.
Researcher:	To find the median you say that the student "would have to count through all the numbers of pets to find the middle" What do you mean by that?
Student:	You need to order 1 1 2 2 2 …and pick the middle.

Question 6

Student:	I know this is completely wrong.
Researcher:	Is it?

Student: I think so, because this is really wide and I think is kind of confusing. But at first I look at every one and though of histogram but then I thought that with a graph you can see it better. Then I try to do it and it turns out some wide.

Researcher: Can you tell me how much time the majority of students take to travel?

Student: I would say this, but like I said things are so confusing because this is so wide and that is skinny but that's taller. I guess I would say this.

Researcher: More were in between 15 and 20.

Question 7

Researcher: How are the students thinking statistically?

Student: The second student is doing like a bar graph and the first student is doing more like a stem-and-leaf.

Researcher: Do you think they can answer questions about the teeth? Like the total number of teeth lost? Do you think that one of the representations would answer this question easier than the other?

Student: I think both ways.

Researcher: How about if we want to find out how many teeth were lost the most?

Student: Either one.

Researcher: How about the rank?

Student: Student 2.

Researcher: How would you make one student understand the other?

Student:

Question 8

Researcher: Do the numbers make sense?

Student: The standard deviation is from the lowest number to the highest number how much they are changing.

Researcher: So, do it make sense that Team A has a higher standard deviation?

Student: Probably not to a child, Team A is more spread out but Team B is way up there towards the middle.

Researcher: If you were a coach, which team would you pick?

Subject I

Question 1

Researcher: For number 1, just tell me if I am right here, ummm, you say you found, OK you counted 26 then you said half of them are 13, then you counted 13 this way and then this way and then divided by two   this remains me of the procedure of….not the mean but…

Student: the median

Researcher: I want you to answer the same question here with typical, for this data set

Student: 1,2,…, 26, so also the 13th position, 1, 2, 13, so between 8 and 9 again, taraaa!. It will be 18 and a half

Researcher: So the same typical time will take for this set that that set

Student: Yeah

Question 2

Student: Cause of you don't take all of them…in other words if you don't take the mean, ummm you are leaving out, you know, the extremes the high extreme, the low extreme. You don't really,… you just know what's in the middle, which is good sometimes, but you also know the range.

Researcher: So, they are not represented if you don't take all of them into account

Student: Right

Question 3

Student: I will explain that … the bars represent just central and south america, not specific country and that when you go up the graph that's the frequency, so I don't know what we understand here, it doesn't say one thousand, one million, but that will be the frequency of how many people.

Researcher: So when he says that the third bar here indicates 85% to 90%, he is missing…he should said also that…

Student:	He left out the frequency, of what that means. 85 to 90 percent what? you know or something like that. 'case it is not literacy rate because the percent is literacy rate but you wanna know how many, so
Researcher:	That will be how many…is this case is 3…3 what?
Student:	units, whatever that is

Question 4

Researcher:	How did you figure out the 21?
Student:	All I did ummm I made everything,… I put all the numbers to 31/2 and then I changed them. I did 31/2, 31/2, , 31/2.[makes a list of six 31/2s] Then I went….and I just gave a little from here [points at the first 31/2] and put it here [the last 31/2], so in other words, I put, I took maybe 11/2 from here made this a 2 and gave 11/2 here so plus 11/2 to make this 5 and then I just distributive it, and I kept the total of this numbers to 21 so when I divided it by 6 will be 31/2 still.
Researcher:	In this case why can't just pick this as an example?
Student:	Because I know that if I have all 31/2s the mean it's gonna be 31/2, is that what you asking me?
Researcher:	Aha, would that be a good example too?
Student:	Sure,.. well, possible…no, you should not have a half a person per household, but if you are working with a different unit.
Researcher:	Does the answer make sense?
Student:	It does because some people may have 3 people in the household and some may have 4 and when you average them out…
Researcher:	In the Raisins problem how would you estimate the mean?
Student:	I would first told them … to take into account ALL of the numbers, you can't just look at one column to find uh ask them to take an educated guess of how many on average they had … and uhmm
Researcher:	What would you think they would say?
Student:	They'd probably said is somewhere something like 35 because is closer to the middle and it is a tall one, but I don't think that will be the correct answer. uhumm and after they give the answer I would start out by

explaining to them how to do a very simple way without using any algorithm, just by show them to find the median…I mean, to find the average we can find the median…so we start by crossing one out from the beginning cross one out from the end, cross another one from the beginning, cross another from the end… and go all the way down until you find the middle, whatever that is.

Researcher: What about part b of the question for this data.

Student: I can give them like rods or something to work with and ahh I can give them a scenario say…. I tell them to put …maybe 16 blocks in one stack, maybe 32 in another row and maybe 64 in another row and then I have them level out the blocks and then they find that 32 blocks will be in each column.

Question 5

Student: a. That 2 students in my class have bird.

Student: Mode would be correct because the mode is the number that occurs the most times , so you don't have to order for the mode and the range you can see what is the lowest number and the highest number is, you have to order it, I mean, unless you have a large number of data then it would be easier to order it, but is not necessary, it is only necessary for the median.

Researcher: You find that the median is "fish", what does that tell you about the pets when you said that fish is the median?

Student: That the average student in the class has fish.

Question 6

Researcher: Tell me can we answer the question?

Student: You just looking at it, do the crossing out of the number and find the median, that is the average.

Researcher: We couldn't do it with the histogram?

Student: No, you can do it with the histogram…it would look good. You could actually visualize the frequency a little bit better. But I thought this would be easier for a student to draw on its own, because with a histogram you have to find the frequency that you are going use between each number and there is a lot more stuff you have to do.

Question 7

Researcher: a. Think statistically
                 This one resembles what type of graph?

Student: A pictograph

Researcher: and this one?

Student: like a bar graph

Researcher: Now think how are they different

Student: They are very similar, if you put this one sideways it is also a bar graph, the different the pictograph is that there is only one number, but then you can have as many pictures as that many number kids are represented and then… but the graph bar every student is represented…where… will be a lot more space …this is… I personally will do a pictograph. You can see that 3 people had 7 then for the bar graph type you have to count each one up where the numbers right there for you.

Researcher: How about in terms of questions each one answers? What is the total number of teeth lost?

Student: I think the pictograph because you can see 2 people in the 2s, so that is 2 times 2, and 2 people in the 3s, so 2 times 3, and so on. while this one you have to add up because it is not lay out for you.

Researcher: What is the most common number of teeth lost?

Student: The pictograph shows it a little bit better, you can see it in the bar graph, it is very close to the seventh category it is not as visual as this one.

Researcher: How can you tell this student to do the other?

Student: The pictograph because that involves more knowledge to group them and make it more readable whereas this one they have the knowledge of how to make the graph and all the information they need to put in it, but they do not understand how to put it together yet and so they just laid it all up.

Question 8

Researcher: Do you calculation make sense?

Student: For team B we see that it is 2.28, that means that the standard deviation is smaller than team a which is 3.58. That makes sense because none of the answers are closer together for team B given a smaller deviation and further apart make a larger deviation.

Researcher: What are the units for the sd?

Student: is it inches? no…ah..I don't know what that means.
                How far apart, how frequently the grades occur?
                ….

Researcher: Which team would you bet on?

Student: Team B, because they look like a stronger team, they all fall within the
                same range whereas here some of them they fall into the lower range and
                some fall on the higher to get that mean of 75. While most of them in
                team B are at 75.

Subject J

Question 1

Researcher: What would be the typical time for this data?

Student: Still take the mean, because…like…just…I don't know. If you take the mode, what would you do if you have two most common times.

Question 2

Researcher: You picked the average because…

Student: I would use the third method, for several reasons. There is always room for error when measuring an object. How are students supposed to know that 6.15 is the most accurate weighing? In using the third method, I would make sure that the same small object would be used for each group. Just because one number may be more common than others, does not mean that is the most accurate, so I would not use the first method [the mode]. You could have a tenth group who got 6.0, what would the students then do?

Question 3

Researcher: When you say "yes and no", does that mean that the student is right in some sense?

Student: Well, by just looking at the graph he sees 7 bars, so that is right. But there could be some countries unaccounted for with some literacy rate that fall on the gap with zero frequency, so we don't know how many countries are represented.

Researcher: What would be the ideal response?

REFERENCES

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

American Statistical Association. (1991). *Guidelines for the teaching of statistics*. Alexandria, VA: Author.

Ball, D. L. (2000). Bridging practices: Intertwining content and pedagogy in teaching and learning to teach. *Journal of Teacher Education*, 51(3), 241-247.

Ball, D. L. & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiples perspectives on the teaching and learning of mathematics*. Westport, CT: Ablex.

Ball, D.L. (1999) Crossing boundaries to examine the mathematical entailed in elementary teaching. *Contemporary Mathematics*, 243, 15-36.

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

Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 433-456). New York: Macmillan.

Batanero, C. (2000). Significado y Comprension de las Medidas de Posicion Central.

Batanero, C., Godino, J. D., Green, D. R., Holmes, P., & Vallecillos, A. (1994). Errors and difficulties in understanding statistical concepts. *International Journal of Mathematics Education in Science and Technology*, 25(4), 527-547.

Berenson, S.B., Friel, S. N., & Bright, G. W. (1993, April). *Elementary teachers' fixations on graphical features to interpret statistical data*. Paper presented at the annual meeting of the American Education Research Association, Atlanta, GA.

Bloom, B. S. (1976). *Human Characteristics and School Learning*, New York: McGraw-Hill.

Bloom, B., Englehart, M. Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, Toronto: Longmans, Green.

Bright, G. W., Friel, S. N., & Berenson, S.B. (1993, April). *Elementary teachers' pedagogical knowledge about statistics.* Paper presented at the annual meeting of the American Education Research Association, Atlanta, GA.

Bright, G. W., & Friel, S. N. (1998). Graphical representations: Helping students interpret data. In S.P. Lajoie (Ed.), *Reflections on Statistics* (pp. 63-117). Mahwah, New Jersey: Lawrence Erlbaum.

Burrill, G. (1998) Statistics and probability for the middle grades: Examples from Mathematics in Context. In S.P. Lajoie (Ed.), *Reflections on Statistics* (pp. 33-59). Mahwah, New Jersey: Lawrence Erlbaum.

Burrill, G. & Hopfensperger, P. (1997) *Exploring Linear Relations: Data-Driven Mathematics*. Pearson Education.

Cobb, G. W. & D. Moore. (1997). Mathematics, Statistics, and Teaching. *American Mathematical Monthly*, 104(9), 801-823.

Cobb, G. (1992). Teaching Statistics. In L. A. Steen (Ed.), *Heeding the call for change. Suggestions for curricular action*. Washington, D. C.: Mathematical Association of America.

Cobb, P. (1999). Individual and collective Mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning*, 5-44.

Cobb, P., McClain, K., & Gravemeijer, K. (2000). Learning about statistical covariation. Manuscript.

Cobo, B. y Batanero, C. (2000). La mediana ¿Un concepto sencillo en la enseñanza secundaria? *UNO,* 23, 85-96.

Conference Board of the Mathematical Sciences (2001). *The Mathematical Education of Teachers*. Providence, Rhode Island: American Mathematical Society.

delMas, R. (2002). Statistical Literacy, Reasoning, and Learning: A Commentary. *Journal of Statistics Education*, 10(3). Retrieved June 15, 2004 from www.amstat.org/publications/jse/v10n3/delmas_discussion.html.

Even, R, & Tirosh, D. (1995). Subject-matter knowledge and knowledge about students as sources of teacher presentations of the subject-matter. *Educational Studies in Mathematics*, 29, 1-20.

Even, R. (1993). Subject-matter knowledge and pedagogical content knowledge: Prospective secondary teachers and the function concept. *Journal for Research in Mathematics Education*, 24(2), 94-116.

Fennema, E, & Franke, M.L. (1992). Teachers' knowledge and its impact. In D. A. Grouws (Ed*.), Handbook of research on mathematics teaching and learning* (pp. 147-164). New York: Macmillian.

Friel, S. N., Curcio, F. R., & 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.

Friel, S. N., Bright, G.W., Frierson, D., & Kader, G. D. (1997). A Framework for Assessing Knowledge and Learning in Statistics (K-8). In I. Gal & J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 55 – 63). IOS Press.

Friel, S. N., & Bright, G. W. (1998). Teach-Stat: A model for professional development in data analysis and statistics for teachers K-6. In S.P. Lajoie (Ed.), *Reflections on Statistics* (pp. 63-117). Mahwah, New Jersey: Lawrence Erlbaum.

Gal, I., Rothschild, K., & Wagner, D. (1989, April). *Which group is better? The development of statistical reasoning in elementary school children*. Paper presented at the meeting of the Society for Research in Child Development, Kansas City, MO.

Garfield, J. B. (2003). Assessing Statistical Reasoning. *Statistics Education Research Journal*, 2(1), 22 – 38.

Garfield J. B. (2002). The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education*, 10(3). Retrieved June 15, 2004 from www.amstat.org/publications/jse/v10n3/delmas_discussion.html.

Garfield, J. (1995) How students learn statistics. *International Statistical Review*, 63, 25-34.

Garfield J. B. (1993). An Authentic Assessment of Students' Statistical Knowledge. In N. L. Webb & A. F. Coxford (Eds.) *Assessment in the Mathematics Classroom*, 1993 Yearbook of the National Council of Teacher of Mathematics (NCTM), (pp. 187-96). Reston, VA.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: implications for research. *Journal for Research in Mathematics Education*, 19(1), 44-63.

George, E.A. (1995, October). *Procedural and conceptual understanding of the arithmetic mean: A comparison of visual and numerical approaches*. Paper presented at the annual meeting of the North American Chapter of the

International Group for the Psychology of Mathematics Education, Columbus, OH.

Gfeller, M.K., Niess, M.L, & Lederman, N.G. (1999). Preservice teachers' use of multiple representations in solving arithmetic mean problems. *School Science and Mathematics*, 99(5), 250-157.

Godino, J. & Batanero, C. (1994). Developing new theoretical tools in statistics education research.

Goodchild, S. (1988). School pupils' understanding of average. *Teaching Statistics*, 10, 77-81.

Hardiman, P., Well, A., & Pollatsek, A. (1984). Usefulness of a balance model in understanding the mean. *Journal of Educational Psychology*, 76, 793-801.

Jones, P.S. (1970). *A History of Mathematics Education in the United States and Canada*. Washington, DC: National Council of Teachers of Mathematics.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 33(4), 259-289.

Lagemann, E.C. (1996). Contested terrain: A history of education research in the United States, 1890-1990. *Educational Researcher, 26*(9), 5.

Landwehr, J. & Watkins, A. E. (1986). *Exploring Data*. Palo Alto: Dale Seymour.

Lappan, G. (2000). A vision of learning to teach for the 21st century. *School Science and Mathematics*, 100(6). 319-326.

Lappan, G., Fey, J., Fitzgerald, W., Friel, S., & Phillips, E. (2002) *Connected Mathematics. Data Analysis and Probability*. Glenview, IL: Prentice Hall.

Leon, M. & Zawojewski, J. (1993). Conceptual understanding of the arithmetic mean. Paper presented at the American Educational Research Association Annual Meeting, Atlanta, Georgia.

Lloyd, G. M., & Wilson, M. (1998). Supporting Innovation: The impact of a teacher's conceptions of functions on his implementation of a reform curriculum. *Journal for Research in Mathematics Education*, 29(3), 248-274.

Loosen, F., Lioen, M., & Lacante, M. (1985). The standard deviation: some drawbacks of an intuitive approach. *Teaching Statistics*, 7(1), 2 − 5.

Lui, H. J. (1998). *A cross-cultural study of sex differences in statistical reasoning for*

college students in Taiwan and the United States. Doctoral dissertation, University of Minnesota, Minneapolis.

Ma, L. (1999) *Knowing and teaching elementary mathematics*. Mahwah, NJ: Lawrence Erlbaum.

McLean, A. (2000). The predictive approach to teaching statistics. *Journal of Statistics Education*. 8(3).

Mevarech, Z. (1983). A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics*, 14, 415-429.

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

Moore, D. & Cobb, G. (2000). Statistics and Mathematics: Tension and Cooperation. *American Mathematical Monthly*, 107, 615-630.

Moore, D.S. (1997) New pedagogy and new content: The case of statistics. *International Statistical Review*, 65(2), 123-165.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.

National Council of Teachers of Mathematics. (2000*). Principles and Standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluations for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (1980). *An agenda for action: Recommendations for school mathematics of the 1980s*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Supervisors of Mathematics. (1977). Position paper on basic skills. *Arithmetic Teacher*, 25(l), 19-22.

Nacional Research Council (2001a). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, and B. Findell (Eds.) Mathematics Learning Study Committee, Center of Education, Division of Behavioral Social and Education. Washington, D.C.: National Academy Press.

National Research Council (2001b). *Knowing and Learning Mathematics for Teaching. Proceedings of a Workshop*. Washington, D. C. : National Academy Press.

Noss, R., Pozzi, S., & Hoyles, C. (1999). Touching epistemologies: meanings of averages and variation in nursing practice. *Educational Studies in Mathematics*, 40, 25-51.

Pollatsek, A., Lima, S., & Well, A. D. (1981). Concept or computation: Students' understanding of the mean. *Educational Studies in Mathematics*, 12, 191-204.

Peterson, P. L. (1988). Teachers' and students' cognitional knowledge for classroom teaching and learning. *Educational Researcher*, 17(5), 5 – 14.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3-14.

Porter, A. C., Kirst, M. W., Osthoff, E. J., Smithson, J. S., & Schneider, S. A. (1993). *Reform up close: An analysis of high school mathematics and science classrooms* (Final Report to the National Science Foundation on Grant No. SPA-8953446 to the Consortium for Policy Research in Education). Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research.

Porter, A. C., & Smithson, J. L. (2001). *Defining, developing, and using curriculum indicators*. Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.

Putman, R. & Borko, H. (2000). What Do New Views of Knowledge and Thinking Have to Say About Research on Teaching Learning? *Educational Researcher*, 29(1), 4 – 15.

Putnam, R., Heaton, R. Prawat, R., & Remillard, J. (1992). Teaching mathematics for understanding: discussing case studies of four fifth-grade teachers. *The Elementary School Journal*, 93(2), 213-229.

Russell, S.J., Goldsmith, L.L., Weinberg, A. S., & Mokros, J.R. (1990, April). *What's typical? Teachers' descriptions of data*. Paper presented at the annual meeting of the Americal Educational Research Association. Boston, MA.

Scheaffer, R. L, Watkins, A. E., Landwehr, J.M. (1998) What every high-school graduate should know about statistics. In S.P. Lajoie (Ed.), *Reflections on Statistics* (pp. 3-31). Mahwah, New Jersey: Lawrence Erlbaum.

Senk, S., Viktora, S., Usiskin, S., Ahbel, N., Highstone, V., Witonsky, D., Rubenstein, R., Schultz, J., Hackworth, M., McConnell, J., Aksoy, D., Flanders, J., & Kissane, B. (1998). *Function, Statistics, and Trigonometry*. The University of Chicago School Mathematics Project. Second Edition. Glenview, IL. Scott Foresman Addison Wesley.

Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 205 – 237). Dordrecht, Netherlands:Kluwer.

Shaughnessy, J.M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed*.), Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: Macmillian.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.

Strauss, S., & Bichler, E. (1988). The development of children's concepts of the arithmetic average. *Journal for Research in Mathematics Education*, 19, 64-80.

Stylianides, A. J. & Ball, D. L. (2004). *Studying the mathematical knowledge needed for teaching: The case of teachers' knowledge of reasoning and proof.* Paper prepared for the 2004 Annual Meeting of the American Educational Research Association, San Diego, CA, April 14, 2004.

Swafford, J., Jones, G., Thornton, C, Stumo, S., & Miller, D. (1999). The impact on instructional practice of a teacher change model. *Journal of Research and Development in Education*. 32(2), 69-82.

Thompson D. & Senk S. (1998). Using rubrics in high school. *The Mathematics Teacher*, 91(9), 786 – 793.

Thompson D. & Senk S. (1993). Assessing Reasoning and Proof in High School. In N. L. Webb & A. F. Coxford (Eds.) *Assessment in the Mathematics Classroom*, 1993 Yearbook of the National Council of Teacher of Mathematics (NCTM), (pp. 167-76). Reston, VA.

Tirosh, D. (2000). Enhancing prospective teachers' knowledge of children's conceptions: The case of division of fractions. *Journal for Research in Mathematics Education*, 31, 5-25.

Vacc, N. & Bright, G. (1999). Elementary Preservice Teachers' Changing Beliefs and Instructional Use of Children's Mathematical Thinking. *Journal for Research in Mathematics Education*, 30(1), 89 – 110.

Valverde, G., Bianchi, L., Wolfe, R., Schmidt, W., & Houang, R. (2002). *According to the Book: Using TIMSS to Investigate the Translation of Policy into Practice Through the World of Textbooks*. Kluwer Academic Publishers: Dordrecht, The Netherlands.

Watson, J. and Moritz, J. (1999). The beginning of statistical inference comparing two data sets. *Educational Studies in Mathematics*, 37, 145-168.

Watson, J. & Moritz, J. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31(1), 44-70.

Watson, J. M. (1997). Assessing Statistical Thinking Using the Media. In I. Gal & J.B.
Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 107-121).
IOS Press.

Wilson, S., Floden, R., & Ferrini-Mundy (2001). *Teacher Preparation Research:
Current Knowledge, Gaps, and Recommendations*. Center for the Study of
Teaching and Policy: University of Washington.