

DEVELOPMENT OF A HIGH SCHOOL STATISTICAL THINKING
FRAMEWORK

Randall E. Groth

238 pages

May 2003

Describing, organizing and reducing, representing, analyzing, and collecting data are fundamental statistical thinking processes. In this dissertation, I identify and describe levels of high school students' thinking within each process.

APPROVED:

Date Cynthia W. Langrall, Co-Chair

Date Edward S. Mooney, Co-Chair

Date Beverly J. Hartter

Date Sharon S. McCrone

DEVELOPMENT OF A HIGH SCHOOL STATISTICAL THINKING
FRAMEWORK

Randall E. Groth

238 pages

May 2003

The study sought to describe the statistical thinking of high school students. The two research questions guiding the study were: (i) What are the defining characteristics of different patterns of high school students' statistical thinking within the processes of describing, organizing and reducing, representing, analyzing, and collecting data? (ii) What levels of statistical thinking can be associated with each of the patterns? In order to answer the two research questions, high school students of various grade levels and mathematical backgrounds and recent high school graduates were asked to solve statistical thinking tasks during clinical interview sessions. The cognitive model described by Biggs and Collis (1982, 1991) was applied in differentiating among patterns of sophistication in the students' responses to the interview tasks. The study identified and characterized levels of thinking which provide the basis of a framework useful for advising instruction, curriculum development, and further research in the area of high school statistics.

APPROVED:

Date Cynthia W. Langrall, Co-Chair

Date Edward S. Mooney, Co-Chair

Date Beverly J. Hartter

Date Sharon S. McCrone

DEVELOPMENT OF A HIGH SCHOOL STATISTICAL THINKING
FRAMEWORK

RANDALL E. GROTH

A Dissertation Submitted in Partial
Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

Department of Mathematics

ILLINOIS STATE UNIVERSITY

2003

© 2003 Randall E. Groth

DISSERTATION APPROVED:

Date Cynthia W. Langrall, Co-Chair

Date Edward S. Mooney, Co-Chair

Date Beverly J. Hartter

Date Sharon S. McCrone

ACKNOWLEDGEMENTS

I wish to thank the faculty associated with the Ph.D. program in Mathematics Education at Illinois State University. It has been a pleasure to study under such a talented and diverse group of faculty members the past few years.

I wish to especially thank the faculty who served on my dissertation committee: Dr. Cynthia Langrall, Dr. Edward Mooney, Dr. Beverly Hartter, and Dr. Sharon McCrone. The effort each one of you has put into reading and commenting upon my work has helped extraordinarily. I am fortunate and grateful to have had the opportunity to work with each one of you.

I also am grateful to my friends and family for their support. I would not have been able to accomplish this goal without them. My fellow graduate students have provided numerous thought-provoking discussions. My family and my fiancée Jennifer have provided valuable encouragement throughout the process.

Finally, and most importantly, I thank God for the numerous blessings He has granted that have made this work possible.

R.E.G.

CONTENTS

	Page
ACKNOWLEDGEMENTS	i
CONTENTS	ii
FIGURES	vii
CHAPTER	
I. THE PROBLEM AND ITS BACKGROUND	1
Background	1
Research Questions	4
Statistical Thinking Processes	4
Describing Data	5
Organizing and Reducing Data	5
Representing Data	6
Analyzing Data	7
Collecting Data	7
Theoretical Perspective	8
The Biggs and Collis Cognitive Model	9
The Formal Mode and Statistical Thinking	11
Significance of the Study	14
II. REVIEW OF RELATED LITERATURE	16
Studies Further Explicating the Biggs and Collis Cognitive Model	16
Empirical Research of Statistical Thinking Processes	20
Describing Data	21

The Complexity of the Process of Describing Data	21
Levels of Sophistication in Describing Data	22
Students' Proficiency in Reading Different Types of Graphs	23
The Importance of Graph Reading Skills for Data Analysis	23
Development of Graph Reading Skills through Instruction	24
Summary	25
Organizing and Reducing Data	26
Understanding of the Arithmetic Mean	26
Understanding of Measures of Center in General	28
Measures of Spread and Measures of Center in Concert	31
Organization of Raw Data Sets	34
Summary	35
Representing Data	36
Representing Data without the Assistance of Computers	36
Representing Data with the Assistance of Computers	40
Summary	42
Analyzing Data	42
Informal Data Analysis and Individual Cognition	43
Informal Data Analysis in the Context of Social Interaction	47
Research Concerning Understanding of Formal Inference and its Foundations	50
Summary	55
Collecting Data	55
Students' Knowledge of Data Collection Methods	56
Students' Abilities to Critique Samples	57
Developing Understanding of Data Collection through Instruction	60
Summary	61
Relationship to this Study	62

III. RESEARCH DESIGN AND METHODOLOGY	63
Study Design	63
Participants	66
Interview Protocol	69
Procedure	74
Data Analysis	76
IV. RESULTS	78
Patterns of Response to Statistical Thinking Tasks	78
Describing Data	79
Organizing and Reducing Data	83
Subprocess 1: Using Measures of Center	84
Subprocess 2: Using Measures of Spread	89
Subprocess 3: Recognizing the Effects of Data Transformation upon Center and Spread	94
Subprocess 4: Organizing Raw Sets of Data	97
Representing Data	100
Analyzing Data	105
Subprocess 1: Comparing Univariate Data Sets	106
Subprocess 2: Analyzing Sample Means	110
Subprocess 3: Identifying Atypical Points in a Tabular Data Set	114
Subprocess 4: Making Multiplicative Comparisons	118
Subprocess 5: Identifying Atypical Points in a Graphical Bivariate Data Set	121
Subprocess 6: Interpolating within Bivariate Data	125
Subprocess 7: Extrapolating from Bivariate Data	127
Subprocess 8: Analyzing Bivariate Relationships	131
Collecting Data	135
Subprocess 1: Designing a Non-Experimental Study	136
Subprocess 2: Designing an Experimental Study	141
Subprocess 3: Critiquing a Study	145
Summary	147
Relationship to the Theoretical Perspective	148

Overview	148
Concrete Symbolic and Formal Modes	150
The Concrete Symbolic Mode and Statistical Thinking	150
The Formal Mode and Statistical Thinking	151
Levels and Modes within Patterns of Response	152
Describing Data	152
Organizing and Reducing Data	153
Representing Data	158
Analyzing Data	160
Collecting Data	170
Summary	175
V. REFLECTIONS ON THE STUDY	177
Summary of the Study	177
Describing Data	178
Organizing and Reducing Data	179
Representing Data	182
Analyzing Data	183
Collecting Data	188
Critique of the Study	188
Implications	191
Implications for Researchers	192
Implications for Teachers	195
Closing Remarks	197
REFERENCES	199
APPENDIX A: Background Descriptions and Anecdotal Information for Study Participants	213
APPENDIX B: Statistical Thinking Tasks	219
APPENDIX C: Alignment of Statistical Thinking Tasks with Statistical Thinking Processes	234

APPENDIX D:	Alignment of Statistical Thinking Tasks with NCTM Principles and Standards for School Mathematics	235
APPENDIX E:	Tasks Designed to Allow Students to Demonstrate Knowledge of Some of the Topics Included on the 2001 AP Statistics Syllabus but not Included in the NCTM Principles and Standards for School Mathematics	237
APPENDIX F:	Student Questionnaire	238

FIGURES

Figure	Page
1. Statistical thinking task involving the analysis of data	12
2. Students who had recently completed a semester-long high school statistics course	68
3. Students who had recently completed a year-long high school statistics course	68
4. Students still enrolled in high school at the time of the study	68
5. Describing statistical graphs	81
6. Using measures of center	85
7. Using measures of spread	90
8. Recognizing the effects of data transformation upon center and spread	95
9. Organizing raw sets of data	98
10. Representing data	101
11. Comparing univariate data sets	107
12. Analyzing sample means	111
13. Identifying atypical points in a tabular data set	115
14. Making multiplicative comparisons	119
15. Identifying atypical points in a graphical bivariate data set	122
16. Interpolating within bivariate data	125
17. Extrapolating from bivariate data	128

18. Analyzing bivariate relationships	132
19. Designing a non-experimental study	137
20. Designing an experimental study	142
21. Critiquing a study	145
22. Relationship between the theoretical model and patterns of response for describing data	153
23. Relationship between the theoretical model and patterns of response for organizing and reducing data	154
24. Relationship between the theoretical model and patterns of response for representing data	159
25. Relationship between the theoretical model and patterns of response for analyzing data	162
26. Relationship between the theoretical model and patterns of response for collecting data	171

CHAPTER I

THE PROBLEM AND ITS BACKGROUND

Background

It is becoming more widely recognized among the mathematics education community that in today's data-driven society, no student should leave high school without engaging in the study of statistics. It is no longer adequate for high school students to take a sequence of courses designed only to prepare them for the study of calculus. Statistical reasoning is essential for all students, regardless of what occupation they may choose to pursue (Gal & Garfield, 1997). Statistics play a key role in shaping policy in a democratic society, so statistical literacy is essential for all citizens in order to keep a democratic government strong (Wallman, 1993). Recognizing the key role statistics play in modern society, the National Council of Teachers of Mathematics (NCTM) (2000) recommended that "by the end of high school students have a sound knowledge of elementary statistics" (p. 48).

Several professional organizations have joined NCTM (1989, 2000) in the call for a more comprehensive treatment of statistics at the secondary level. The American Statistical Association (1991), for one, recommended that high school students engage in exploring data, using techniques of formal inference, planning studies, and analyzing how statistics are used in society. The American Association for the Advancement of Science (1993) echoed some of the same recommendations and emphasized that high school

graduates need to be able to critically analyze the design of studies in which statistical techniques are used. The need for comprehensive statistical education at all grade levels has been recognized not only in the United States, but in the international community as well (e.g., School Curriculum and Assessment Authority & Curriculum and Assessment Authority for Wales, 1996; Australian Education Council, 1994).

There is some evidence to suggest that the calls for expanded treatment of the subject of statistics at the secondary level have not gone unheeded. On the student questionnaire given with the 1990 National Assessment of Educational Progress (NAEP), 87.6 percent of all twelfth-grade students indicated that they had not taken probability or statistics, but by the 1996 NAEP that number had dropped to 79.2 percent (Shaugnessy & Zawojewski, 1999). In the late 1990's, reform curricula funded by the National Science Foundation (NSF) emerged at the secondary level integrating the teaching of statistics with the teaching of traditional topics such as algebra and geometry (Hirsch, Coxford, Fey, & Schoen, 1998; Alper, Fraser, Fendel, & Resek, 1998). Even traditional algebra textbooks have begun to include statistical topics along with algebraic topics (e.g., Collins, 1998). In 1997, the College Entrance Examination Board administered its first Advanced Placement Statistics Examination to 7,600 students (Garfield & Chance, 2000). By 2001, the number of students taking the exam had grown to 41,609 (College Board, 2001a). The subject of elementary statistics appears to be taking root at the secondary level.

As the study of statistics becomes more common at the high school level, it is imperative that the research base for advising curriculum design, instruction, and further research be expanded. Extensive empirical research has been done within topics

traditionally included in secondary school curricula, such as algebra and functions (Chazan, 2000; Vinner, 1985; Sfard & Linchevski, 1994; Tall, 1992), geometry (van Hiele, 1986; Usiskin, 1982; Fuys, Geddes, & Tischler, 1988; Burger & Shaughnessy, 1986), and calculus (Thompson, 1994; Ferrini-Mundy & Guadard, 1992; Heid, 1988; Orton, 1983). Since statistics is a relatively new topic in secondary school curricula, research in this area is not as well developed (Lajoie & Romberg, 1998). In fact, it is only within the past decade that research studies focusing upon the teaching and learning of statistics at any level have begun to accumulate (Shaughnessy, Garfield, & Greer, 1996). Therefore, the research base for advising statistics instruction at the high school level is only in the beginning stages of construction.

One powerful way to build the research base in the area of high school statistics is to conduct research focused upon building models that describe students' statistical thinking. Cobb et al. (1991) and Resnick (1983) contended that research-based frameworks describing students' thinking are vital for advising curriculum and instruction. Fennema and Franke (1992) pointed out that research-based knowledge of students' thinking does, in fact, serve to improve teachers' instruction. Jones et al. (2001) demonstrated the effectiveness of providing teachers with knowledge of students' thinking in the area of statistics by successfully using a research-based statistical thinking framework to advise instruction at the elementary school level. The present study focused upon building a comparable framework with the potential to advise statistics instruction at the high school level.

The high school framework constructed as a result of the present study describes various levels of students' cognition across five different statistical thinking processes. A process is a cognitive operation that incorporates the organizing, coding, and interpreting of information (Reber, 1995). In conceptualizing the study, describing, organizing and reducing, representing, analyzing, and collecting data were identified as the five statistical thinking processes to be investigated. Within each statistical thinking process, descriptions of the characteristics of different levels of thinking were sought. This conceptual structure was derived from the research of Jones et al. (2000) and Mooney (2002), who constructed frameworks to describe the statistical thinking of elementary school and middle school students, respectively. Each of those studies, like this one, sought to describe students' cognitive levels across statistical thinking processes.

Research Questions

Two main research questions guided the construction of the high school statistical thinking framework:

1. What are the defining characteristics of different patterns of high school students' statistical thinking within the processes of describing, organizing and reducing, representing, analyzing, and collecting data?
2. What levels of statistical thinking can be associated with each of the patterns?

Statistical Thinking Processes

Describing, organizing and reducing, representing, analyzing, and collecting data are five non-disjoint statistical thinking processes. This section serves to identify some of the defining features of each of the processes and the relationships they have to each other.

Reports written by the National Research Council (2001) and Mooney (2002) were drawn upon to help define the first four processes, and an essay by Scheaffer, Watkins, and Landwehr (1998) was used to help define the fifth.

Describing Data

Describing data “involves reading displays of data (e.g., tables, lists, graphs); that is, finding information explicitly stated in the display, recognizing graphical conventions, and making direct connections between the original data and the display” (National Research Council, 2001, p. 289). This process forms an essential part of the foundation of statistical thinking. Curcio (1987) emphasized this fact by pointing out that the ability to “read the data” is prerequisite to the abilities to “read between the data” and “read beyond the data.” Students need to be able to “read the data” in order to engage in data analysis. Wainer (1992), in discussing how to assess students’ ability to analyze statistical graphs, emphasized this by pointing to the necessity of measuring ability to read information explicitly from graphs before investigating ability to answer higher level analysis questions involving comparing graphs and making inferences or predictions from them. Data description skills are indispensable parts of the repertoire needed to engage in analyzing data.

Organizing and Reducing Data

Another important part of the overall foundation for statistical thinking is the process of organizing and reducing data. This process “incorporates mental actions such as ordering, grouping, and summarizing. Data reduction also includes the use of representative measures of center (often termed measures of central tendency) such as

mean, mode, or median, and measures of spread such as range or standard deviation” (National Research Council, 2001, p. 289). Students who can flexibly group data can later examine and reflect upon their groupings in order to find trends and patterns within the data. Those proficient in using measures of center and spread to describe data sets are able to convey important information about some of the defining features of the data sets using just a few summary statistics. These defining features can also be used to make comparisons among different sets of data. Therefore, organizing and reducing data is an essential part of analyzing data.

Representing Data

Representing data is the third of the foundational statistical thinking processes investigated in this study. Representing data “involves displaying data in a graphical form” (Mooney, 2002, p. 27). This process includes being able to construct a data display for a given data set and constructing alternate displays for a data set. Students must be able to organize data and understand conventions like labeling and scaling in order to construct data representations (National Research Council, 2001). It is important to note that the construction of representations of data generally should not be an end in and of itself (Bright & Friel, 1998). Rather, the ability to represent data is important due to the fact that once visual representations have been constructed, they can be used to gain insight about the nature of the set of data. The representations can also be used to convey the results of data analysis to others.

Analyzing Data

The three foundational processes of describing, organizing and reducing, and representing data all come into play as students engage in the process of analyzing data. Analyzing data involves identifying trends and making inferences or predictions from charts, tables, or graphs (Mooney, 2002). It includes making comparisons within data sets or data displays, making comparisons between data sets or data displays, and making inferences from a given data set or data display. Curcio's (1987) categories of "reading between the data" and "reading beyond the data" are parts of this process, since the former includes making comparisons within and among data sets, and the latter includes making inferences from data sets. The nature of informal data analysis is captured in Tukey's idea of "Exploratory Data Analysis" (EDA), since "the goal of EDA is to see what the data in hand say, on the analogy of an explorer entering unknown lands" (Cobb & Moore, 1997, p. 807). Students who have engaged in EDA can go on to learn how to meaningfully employ methods of formal inference in the process of analyzing data (Cobb & Moore, 1997). Analyzing data is perhaps the most complex of the statistical thinking processes described thus far.

Collecting Data

The final statistical thinking process investigated in the study is collecting data. This process includes planning, conducting, and critiquing surveys, experiments, and observational studies (Schaeffer, Watkins, & Landwehr, 1998). The collection of data is one of the central components of the practice of statistics (Moore, 1990; Cobb & Moore, 1997; Wild & Pffankuch, 1999). Understanding data collection processes is an important

component for effective citizenship in a democracy (Wallman, 1993). Citizens are constantly presented the results of surveys, experiments, and observational studies. In order to be able to interpret the results of these studies, it is vital to have some understanding of the methods used in planning and conducting the studies. While many of the issues involved in collecting data are of a non-mathematical nature, the investigation of this process was included in the study in order to help provide a more complete portrait of students' statistical thinking.

Theoretical Perspective

The Neo-Piagetian model of development described by Biggs and Collis (1982) was used as a model for identifying levels of statistical thought in this study. The choice to use this model was made because it has been used effectively by other researchers to help identify various levels of sophistication in statistical thinking. The statistical thinking developmental framework for elementary school students formulated by Jones et al. (2000) and the Middle School Students Statistical Thinking (M3ST) framework (Mooney, 2002) are both based upon Biggs and Collis' model. Watson, Moritz, and colleagues have conducted several studies in which they applied the Biggs and Collis model to describe the relative sophistication of students' responses to statistical thinking tasks (e.g., Watson & Collis, 1994; Watson, Collis, Callingham, & Moritz, 1995; Watson & Moritz, 1999a; Watson & Moritz, 1999b; Watson & Moritz, 2000a; Watson & Moritz, 2000b). Since Biggs and Collis' model helped to construct the levels within the high school statistical thinking framework, in this section I will describe the model itself in some detail and the parts of the model that are especially relevant for describing the

statistical thinking of high school students.

The Biggs and Collis Cognitive Model

According to Biggs and Collis (1982), human development follows “modes” which are similar to the “stages” of development described by Piaget (1983). Biggs and Collis’ (1991) five modes of development are: sensorimotor (from birth), ikonic (from around 18 months), concrete symbolic (from around 6 years), formal (from around 14 years), and post-formal (from about 20 years). A primary difference between the “modes” of Biggs and Collis and the “stages” of Piaget is that the Biggs and Collis model includes a “post-formal” mode of development after the “formal” mode. Another difference is that Biggs and Collis emphasize that humans have the potential for “multimodal” functioning, in that they often use thinking characteristic of a lower mode of development in order to make sense of tasks that require functioning characteristic of a higher mode of development (Biggs & Collis, 1991). Finally, Biggs and Collis (1982) assert that *decalages*, which are instances of uneven functioning across modes by an individual on tasks from different subject areas, are more frequent than Piaget’s work implies (The research efforts of Jones et al. (2000) and Mooney (2002) indicate that instances of uneven functioning by a student may also occur across statistical thinking processes). Overall, however, Biggs and Collis’ (1991) “modes” of human development are quite similar in sequence and structure to Piaget’s stages. As humans attain each mode, their thought processes are qualitatively different from those exhibited during previous modes of development.

According to Biggs and Collis (1991), modal acquisition tells only part of the story

of human cognitive development. Biggs and Collis' theory of development is again similar to that of Piaget (1983) in that it posits the existence of three distinct levels of development within each of the modes. The unistructural level is the first level of development within any given mode. Within this level, responses to tasks are incomplete because students simply seize upon the first relevant aspect of the given task which comes to mind. The multistructural level is the next highest level of functioning within a given mode. Here, students consider and seize upon various aspects of the task, but their responses are incomplete because they do not see the relationships among those various aspects. They see relationships among the aspects of a task only after reaching the relational level, which is the highest level of functioning within any given mode. These levels of development are based upon studies across several different content areas indicating that quantitative stages of learning occur before learning changes qualitatively (Biggs, 1999). The difference between the unistructural and multistructural levels is a quantitative one (dictated by the number of relevant aspects incorporated) whereas the difference between the multistructural and relational levels is a qualitative one (dictated by the quality of the relationship existing among the aspects incorporated).

Statistical thinking research using the Biggs and Collis model has focused upon describing students' statistical thinking within the ikonic and concrete symbolic modes. The M3ST framework (Mooney, 2002) and the Jones et al. (2000) elementary school framework both describe students' functioning within these two modes of development. The first level of thinking described in each framework captures the essence of relational thinking within the ikonic mode, and the remaining three levels described in each are

isomorphic to the unistructural, multistructural, and relational levels within the concrete symbolic mode. Watson, Moritz, and colleagues have focused upon describing statistical thinking within the same two modes of development (e.g., Watson & Collis, 1994; Watson, Collis, Callingham, & Moritz, 1995; Watson & Moritz, 1999a; Watson & Moritz, 1999b; Watson & Moritz, 2000a; Watson & Moritz, 2000b). Because of this body of research, we are beginning to have a better understanding of the characteristics of statistical thinking within the ikonic and concrete symbolic modes.

Researchers have yet to give a comprehensive description of what characterizes the various aspects of statistical thinking of students within the formal mode of development. Assuming that the timeline for modal acquisition reported by Biggs and Collis is fairly accurate, one would expect some high school students to function predominantly at the concrete symbolic mode of development, while others would tend to function at the formal mode of development. In the next subsection, I will describe some of the key characteristics of the formal mode of development, and then discuss how formal mode thinking might manifest itself within the context of statistics.

The Formal Mode and Statistical Thinking

The ability to deal with more abstract concepts separates the formal mode thinker from the concrete symbolic thinker. In Piaget's stage model, which is similar in structure to Biggs and Collis' developmental model, hypotheticodeductive reasoning marks the transition to formal operational thought (Blackburn & Papalia, 1992). Blackburn and Papalia (1992) explained that "the elements of such thought are propositions – conceptual rather than physical entities" (p. 147). In other words, hypotheticodeductive thinkers are

able to go beyond reasoning about what is concrete and readily accessible to reasoning about abstract concepts and hypotheses. As such, formal mode thinkers have the ability to examine the underlying structure of a subject and generate solutions to problems based upon that structure (Biggs & Collis, 1991). Piagetian tasks involving combinatorial and proportional reasoning, which require hypotheticodeductive reasoning, are two types of tasks that traditionally have been used to differentiate concrete-symbolic thinkers from formal thinkers (Blackburn & Papalia, 1992). Students functioning at the level of formal operations are able to: write essays about hypothetical situations, design experiments to answer questions, justify various opposing positions with logical arguments, and examine broad concepts rather than just narrow sets of given facts (Woolfolk, 1993).

As an example of how formal mode thinking may be differentiated from concrete symbolic thinking within the context of statistics, consider the task in figure 1.

Department Store Worker Salaries															
Suppose that in 1975, a newspaper reporter took a random sample of 15 department stores from the state of Illinois. For each department store he sampled, he found out how much the highest paid man and the highest paid woman in the department store were paid per hour. The table below shows the results of his survey:															
Salaries (in dollars per hour) of the people in the reporter's sample															
Store	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Women	12.5	9.0	11.0	9.5	2.5	12.0	3.0	4.0	4.0	2.5	6.0	8.5	4.5	3.0	10.0
Men	17.5	15.0	20.0	20.0	20.0	19.0	20.0	18.0	5.5	6.0	10.0	16.5	12.5	10.0	7.0
After doing this survey, the reporter wrote, "There is no difference between the average hourly salary of the highest-paid male employees of department stores in Illinois and the average salary of their female counterparts." Do the numbers from his survey support this statement? Why or why not?															

Figure 1. Statistical thinking task involving the analysis of data

When this task was given to a group of high school students (Groth, 2002a), including two AP Statistics students who had begun the study of formal inference, they did not think beyond the data listed in the table in order to answer the question. Most calculated a mean salary for the males and a mean salary for the females and compared the two. No hints of thinking within the formal mode of development were evident in their responses to the question, since they never went beyond the data at hand. Thinking only about the data set at hand and not thinking about other possible data sets that could be drawn from the population placed all of the responses of students in the Groth (2002a) study within the concrete symbolic mode of development.

Even though only concrete symbolic responses were obtained in response to the question in figure 1, in conceptualizing this study I conjectured that such a question has the potential to reveal a variety of levels of thinking within the formal mode. In Biggs and Collis' (1991) terms, students at a unistructural level of development in the formal mode may recognize that the sample is only one out of a large number of possible samples, and recognize the need to find out how likely the sample is to be observed by chance within the structure of a hypothetical distribution of all possible samples. Students at a multistructural level of development within the formal mode may have knowledge of a number of formal inferential methods and recognize the need to use one of them in answering the question. However, they may have an underdeveloped understanding of those methods, and might apply the wrong one in answering the question. Students at the multistructural level, for instance, might incorrectly use a two-sample t-test to compare the mean salaries in this situation. This type of mistake actually occurs quite frequently

on AP Statistics examinations (College Board, 2002). Students functioning at the relational level would be expected to use the correct formal inferential procedure (a matched pairs t-test) to analyze the data, and justify fully why the formal inferential procedure is appropriate in this situation. Hence, the Biggs and Collis model, even though previously used by researchers only to describe levels of thinking within the ikonic and concrete symbolic modes, can also provide a framework for describing levels of statistical thinking within the formal mode.

Significance of the Study

This study was unique in its extensive focus upon high school students. As will be seen in the next chapter, several statistical thinking studies have included students of high school age, but have not focused exclusively upon them. The focus upon high school students was important for two reasons. First, little research currently exists to advise the teaching of statistics to high school students. This study helps to remedy that problem. Second, available research gives few clues about the characteristics of statistical thinking within the formal mode of development. This study plays a role in filling that theoretical gap. The focus upon high school students should, therefore, appeal both to teachers seeking to improve instruction in statistics and to researchers wishing to continue to detail the characteristics of thinking within the formal mode of development.

It is important to note that no framework of this nature is ever “complete.” There is room for further investigation of processes and levels of high school students’ statistical thinking which are not included in the framework constructed by this study.

Incompleteness is characteristic of all studies, since no form of research can ever claim to

give access to an external, unchanging, objectively existing reality (Smith & Deemer, 2000). Therefore, while the framework constructed by this study does not claim to expose a complete and unchanging reality about all aspects of secondary students' statistical thinking, it does contain insights about students' thinking that can help advise teaching, research, and curriculum development in the area of high school statistics.

CHAPTER II

REVIEW OF RELATED RESEARCH

In this chapter, I take up the task of analyzing and summarizing the literature that helped guide the development of the high school students' statistical thinking framework. One goal of the chapter is to help to clarify and define the scope of my study and the lens through which data were viewed and analyzed. Another goal is to help to situate this study within the larger body of statistical thinking research. My discussion is divided into two main sections. In the first section, I discuss empirical research studies that have further explicated aspects of the Biggs and Collis (1982, 1991) developmental model. In the second section, I discuss the research literature pertaining to each of the five statistical thinking processes that comprise the framework: describing, organizing and reducing, representing, analyzing, and collecting data.

Studies Further Explicating the Biggs and Collis Cognitive Model

As discussed in chapter 1, the general cognitive model described by Biggs and Collis (1982, 1991) was used to identify levels of sophistication in high school students' statistical thinking. The model posits that human development follows five modes, which are somewhat similar in structure to Piagetian stages. Within each mode, three levels of development exist: unistructural, multistructural, and relational. At the unistructural level students demonstrate knowledge of only one relevant aspect of a task. At the multistructural level, students are aware of several of the components necessary to solve a

problem, but cannot integrate them properly to solve the problem. Upon reaching the relational level, students finally see the relationships among the components needed to solve the problem. In the first chapter, I discussed the Biggs and Collis model in detail, along with an explanation of the modes of development I expected to observe in the present study. In this section, I will discuss how the model has been used and refined in empirical research focused upon statistical thinking.

Jones et al. (2000) and Mooney (2002) both used Biggs and Collis' developmental model to identify four distinct levels of statistical thinking. Clinical interviews were used in each study to form frameworks for levels of statistical thinking across four data analysis processes: describing, organizing and reducing, representing, and analyzing and interpreting data. Responses to clinical interviews were assigned to one of four categories: idiosyncratic, transitional, quantitative, or analytical. These categories were isomorphic to the levels of development within and preceding the concrete symbolic mode. The idiosyncratic category was the same in structure as the relational level of the ikonic mode, and the remaining three categories, respectively, were the same in structure as the unistructural, multistructural, and relational levels within the concrete symbolic mode. Both studies represent fairly straightforward applications of Biggs and Collis' model for the purpose of identifying levels of sophistication in statistical thinking.

While Jones et al. (2000) and Mooney (2002) each identified only one unistructural – multistructural – relational (UMR) cycle within the concrete symbolic mode, other researchers have identified more than one. Watson, Collis, Callingham, and Moritz (1995) documented the occurrence of more than one UMR cycle in the area of statistical

thinking. In their study, two UMR cycles within the concrete symbolic mode emerged as sixth-grade students engaged in analyzing data cards containing various bits of information about individual people. The first UMR cycle climaxed in students “interpreting the information on the cards in an aggregated rather than an individual sense” (p. 254). The second UMR cycle saw students move from realizing that they needed to justify causal claims made during data analysis to using statistics and graphical displays to justify the claims. Other studies, the results of which will be further discussed in the next section, have confirmed that there can exist two (Watson & Moritz, 1999a) or even three (Reading & Pegg, 1996) UMR cycles within the concrete symbolic mode. In theory, there seems to be no limit to the number of UMR cycles which may occur in any given mode of development (Pegg & Davey, 1998).

In addition to identifying several UMR cycles, empirical research has identified instances of multimodal functioning in statistical thinking. Multimodal functioning occurs when students use thinking characteristic of a lower mode of development to provide support for solving a task requiring thinking at a higher mode (Biggs & Collis, 1991). Watson and Collis (1994) described how students used ikonic support to help compare the test scores for two classes. Visual strategies in comparing two graphs (ikonic) were used by some students to help support strategies based upon totals and averages (concrete symbolic). In the same study, ikonic strategies were also used to support concrete symbolic thinking in tasks involving interpreting a bar chart and determining the “fairness” of two dice. Watson, Collis, Callingham, and Moritz (1995) documented ikonic support for students’ concrete symbolic thinking within the context of

a data card analysis activity. In the study, they found that imaginative speculation about causes of data values while examining cards supported the sorting of cards in a systematic fashion. While these studies document only ikonic support for concrete symbolic thinking, it is important to note that it is also possible to have support at any given mode of thinking from any one or more of the modes below it (Pegg & Davey, 1998).

Statistical thinking research also suggests that individual students exhibit decalages across statistical thinking processes. That is, a student who tends to function at a given level within one of the statistical thinking processes does not necessarily function at that same level for all statistical thinking processes. For example, some of the elementary school students in the Jones et al. (2000) study exhibited lower levels of thinking for the process of analyzing data than for other statistical thinking processes. Mooney (2002) found some inconsistency in the performance of middle school students across statistical thinking processes. Hence, a richer picture of an individual student's statistical thinking is obtained when the student is asked to solve tasks incorporating different statistical thinking processes. A student who exhibits a particular level of thinking on tasks associated with one statistical thinking process will not necessarily exhibit that same level of thinking across all processes.

Empirical research, then, has helped to clarify how Biggs and Collis' model of cognitive development may be applied to identifying levels of sophistication in students' statistical thinking. Some studies have shown that only one UMR cycle may be readily identifiable from empirical data (Jones et al., 2000; Mooney, 2002). Other studies have

shown that it is sometimes possible to identify more than one UMR cycle within students' thinking (Watson, Collis, Callingham, & Moritz, 1995). Research also has demonstrated that statistical thinking at a high mode of development is sometimes supported by thinking characteristic of lower modes (Watson & Collis, 1994). Finally, it has been demonstrated that individual students may exhibit different levels of sophistication in thinking across statistical processes (Jones et al., 2000; Mooney, 2002). The manner in which levels of statistical thinking were identified and described in each of these studies helped in the process of identifying and describing levels of thinking in the present study.

Empirical Research of Statistical Thinking Processes

This study investigated high school students' statistical thinking across five processes: describing, organizing and reducing, representing, analyzing, and collecting data. Study-specific definitions were assigned to each process in chapter 1. To recapitulate briefly, describing data involves the explicit reading of data presented in tables, charts, or graphs. Organizing and reducing data is the process of arranging, categorizing, or consolidating a given set of data into summary form. Representing data entails displaying a given set of data by using graphs. Analyzing data involves identifying trends and making inferences or predictions from a data display or set, using formal inferential methods when appropriate. The process of collecting data involves planning, conducting, and critiquing surveys, experiments, and observational studies. In this section, I describe empirical research of each of the five processes.

Describing Data

Several aspects of students' thinking within the process of describing data have been studied by researchers. The research literature illustrates the complexity of the process of describing data. It provides a framework detailing the different levels of sophistication in thinking that can be identified as students become more proficient in describing data. It makes us aware of the statistical representations students generally are able to read successfully and those that tend to cause difficulty. In addition, research illustrates the importance of this statistical thinking process by showing that students must have strong graph reading skills in order to engage in meaningful data analysis. Finally, it shows that students' abilities to read data from different types of graphs, even some of those generally considered "difficult" to read, can be developed through instruction.

The Complexity of the Process of Describing Data

Friel, Curcio, and Bright (2001) described the complex nature of the process of reading graphs. They defined "graph comprehension" as "the ability of graph readers to derive meaning from graphs created by others or by themselves" (p. 145). Through a review of literature from several different disciplines, they identified 4 factors that affect graph comprehension. The first factor is the purposes for which graphs are used. Graphs used within the context of data analysis are likely to carry more robust meaning than those presented in conventional textbook instruction. The second factor is the characteristics of the tasks in which graphs appear. Specifically, research has shown that graphs set within a context will be read differently than those that are context-free. Thirdly, the characteristics of the data set represented must also be considered. They

state, “The spread and variation within a data set, the type of data, the size of a data set, and the way a representation provides structure for data (i.e., graph complexity) can influence graph comprehension” (p. 141). The fourth factor influencing graph comprehension was identified as “learner characteristics.” Logical and abstract thinking abilities, familiarity with different contexts for graphs, and mathematical knowledge and experience all come into play as students attempt to read graphs. Friel, Curcio, and Bright (2001) concluded that “making sense of graphs appears to be more complex than once thought” (p. 151).

Levels of Sophistication in Describing Data

In constructing the Middle School Students Statistical Thinking (M3ST) framework, Mooney (2002) documented four levels of sophistication in middle school students’ thinking about tasks involving “describing data.” Interview tasks were designed to investigate the abilities of 12 target students to read data from a two-column table, bar graphs, a scatter plot, and a line graph. At the lowest level of thinking, students demonstrated little awareness of the display features, recognized the same data represented in different displays only by using subjective judgments, and evaluated the effectiveness of displays based on irrelevant features or reasons. Students at the lowest level were not able to identify the units of data values or else misinterpreted them. At the highest level of thinking, students were able to correctly read all display features of graphs, and were able to pick out irrelevant or cosmetic features. They recognized the same data in different displays by using numerical rather than subjective reasoning. They also evaluated the effectiveness of different data displays by focusing upon relevant

display features. Mooney's M3ST framework reveals the existence of various levels of thinking in response to tasks asking students to read statistical displays.

Students' Proficiency in Reading Different Types of Graphs

The Seventh Mathematics Assessment of the National Assessment of Educational Progress (NAEP) provided information about how well students in the U.S., in general, are able to read statistical graphs (Zawojewski & Shaughnessy, 2000). These NAEP data indicated that students did well on reading information from pictographs in grade 4, reading stem-and-leaf plots in grade 8, and line graphs in grade 12. However, students in grade 12 did poorly in reading box plots, with nearly two-thirds of student test-takers incorrectly answering a question requiring them to read from such a display. Box plots and stem and leaf plots had not previously been included on the NAEP, but as the results show, students were better able to make sense of stem and leaf plots than box plots. Zawojewski and Shaughnessy (2000) conjectured that difficulties in reading box plots may be due to the fact that they are a more complex type of display than stem and leaf plots. While stem-and-leaf plots explicitly show individual data points in their graphical displays, box plots tend to hide some of the data by presenting it in a more reduced form. Statistical displays incorporating higher levels of data reduction appear to be difficult for many students to read.

The Importance of Graph Reading Skills for Data Analysis

It appears that the inability to read some types of graphical displays has a negative impact on secondary students who attempt to engage in data analysis. Biehler (2001) documented several difficulties twelfth-grade students had in reading data displays while

engaged in data analysis tasks during clinical interviews. Some students were unable to read information presented in a frequency bar graph. Some had difficulty interpreting the meaning of the labels on the horizontal axis of histograms. Also, describing the same data presented in the form of both a box plot and a histogram proved to be conceptually challenging for students. Different aspects of the process of describing data present cognitive hurdles for many students to overcome before they are able to meaningfully engage in statistical data analysis.

Development of Graph Reading Skills through Instruction

Bright and Friel (1998) conducted a study to learn how instruction influences the abilities of students in grades 6, 7, and 8 to read information from statistical graphs and make connections between the graphs. They reported on the results of clinical interview sessions with four eighth-grade students who had just completed a unit emphasizing connections between different types of statistical graphs. They found that while the students tended to confuse the meanings of the x- and y-axes before the instructional unit, they did not do so after completing the unit. They also reported that line plots, with x's representing individual data points, were much easier for the students to read than bar graphs, which present the data in a more reduced form. Also, while students demonstrated the ability to read both stem-and-leaf plots and histograms, they had difficulty seeing connections between the two representations. The researchers concluded that it is important for students to be given experiences that allow them to increase their ability to read graphical displays before they are asked to make comparisons between data represented in graphs and make inferences from graphical displays.

Friel (1998) reported on eighth-grade students' ability to read information from box plots after they had completed an instructional unit focused on topics related to sampling and using box plots to represent data. She reported on the results of clinical interviews with six students at the end of the instructional unit. She found that the students she interviewed had no problems with reading data from box plots. When shown a box plot comparing quiz scores for students in three different class periods and asked, "What are the minimum and maximum values and the median for first period" (p. 367), students experienced no difficulty. All students were able to answer correctly. In addition, some students even identified the lower and upper quartiles, even though they were not explicitly asked to do so. All of the students she interviewed demonstrated competence in reading information directly from box plots. Her findings suggest that the reason high school students generally have difficulty in reading box plots may be due to the lack of instructional attention currently given to the topic at the middle and high school levels.

Summary

Describing data is a complex statistical thinking process. There are several different levels of sophistication in thinking within the process. Certain types of graphical displays, such as those that incorporate data reduction, can prove to be especially difficult for students to read. An inability to read several different types of graphical displays can impair high school students from performing data analysis tasks. Through instruction, however, it is possible for students to begin to see connections among statistical graphs and read complex displays such as box plots.

Organizing and Reducing Data

The body of research documenting students' ability to organize and reduce data is even more abundant than the body of research about describing data. This is due to the fact that several researchers have investigated students' understandings of measures of center. Some of these researchers have focused solely upon students' understanding of the arithmetic mean and identifying difficulties students have with the concept. Others have taken a broader perspective, investigating students' use of measures of center in general and describing how understanding of these measures tends to develop. Our understanding of how students use measures of spread is also growing, since some researchers have investigated students' use of measures of spread along with their use of measures of center. Research concerning students' ability to organize "raw" sets of data is sparse (Mooney, 2002), but is beginning to accumulate as seen in some of the studies to be discussed.

Understanding of the Arithmetic Mean

Pollatsek, Lima, and Well (1981) conducted one of the first empirical research studies aimed at investigating students' understanding of the arithmetic mean. They interviewed 37 psychology undergraduates at the beginning of an introductory statistics course. During clinical interview sessions, they asked students to solve problems requiring the computation of means and weighted means. Many of the students interviewed computed unweighted means in situations where it would have been correct to compute weighted means. One student correctly calculated a weighted mean in one problem context, but failed to do so in a different context. Pollatsek, Lima, and Well

concluded that students need to have three types of knowledge about the mean: functional, so that the mean is understood as a real-world concept; computational, including being able to work the algorithm in reverse; and analog, suggesting that teachers help students develop a mental image of the mean as the balance point for any given set of data.

Strauss and Bichler (1988) were among the first to investigate younger students' understanding of the arithmetic mean. They conducted clinical interviews with 80 students, 20 each of age 8, 10, 12, and 14. During the clinical interviews, students solved problems aimed at uncovering their understandings of seven different properties of the arithmetic mean. They found that almost all students interviewed were aware that the mean is influenced by values other than the mean, and that the mean does not necessarily equal one of the values in the data set from which it arises. Over 80% of the 12- and 14-year olds were aware that the mean is located between the extreme values of a data set. Virtually all students were not aware that the sum of the deviations from the arithmetic mean is zero. Some students also had difficulty understanding that when an average is calculated, any zero values must also be summed, and that the average value is representative of the values in a set of data. The findings of Strauss and Bichler show that it is possible for students to have some understanding of the arithmetic mean without grasping all of its properties completely.

Cai (2000) compared students' understanding of the arithmetic averaging algorithm in two countries: the U.S. and China. 232 sixth-grade students from Milwaukee, Wisconsin were compared to 311 sixth-grade students living in a demographically similar

city in China. Students were given 10 written tasks requiring the application of the arithmetic mean. Some problems required simple calculations, while others were more complex and required students to perform the algorithm in reverse. Chinese students were more successful than American students in obtaining correct answers to the items. Both sets of students were more successful with items requiring simple computation using the algorithm than with items requiring a more conceptual understanding. Chinese students were more successful with problems requiring the execution of the algorithm in reverse, mostly due to the fact that they were more familiar with using algebraic representations to solve problems. Based upon the data collected, Cai stated that allowing children to use invented strategies might help them come to understand the arithmetic mean and its applications, and that U.S. students should be taught to use both arithmetic and algebraic representations to solve problems.

Understanding of Measures of Center in General

Mokros and Russell (1995) studied how students construct representative measures for data sets along with studying their understanding of the arithmetic mean. In the study, they interviewed 21 students, seven each from fourth, sixth, and eighth grades. Each of the students in the study had been taught the algorithm for computing the arithmetic mean during their mathematics classes. During the clinical interviews, students were asked to solve seven open-ended problems concerning the notion of average.

Mokros and Russell (1995) identified five predominant approaches to solving the problems: average as the mode, an algorithm, reasonable, the midpoint, and the balance point. Students who conceptualized average as the mode tended to come up with

idiosyncratic strategies to solve the problems. Students who conceptualized averages as algorithms tended to misapply the algorithm in solving problems, and common sense was often sacrificed in the use of the algorithm. Those classified as seeing the average as a “reasonable” value had a sense that the average should be representative of the data set from which it arises, but did not have a well-articulated mathematical definition for average. Students in the “midpoint” category also had a sense that the average should be representative, but for the most part were not aware of the fact that they were using the formal measure of center called the median to answer the problems. Students classified in the “balance point” category tended to use sound intuitive balancing strategies along with a misapplied algorithm for calculating the mean. Mokros and Russell concluded that students need to be taught to see data sets as entities rather than as disparate points, and that strictly algorithmic approaches to teaching the mean can prevent some students from seeing the mean as a representative measure.

NAEP data confirm that students frequently make poor choices in selecting measures of center to describe data sets (Zawojewski & Shaugnessy, 2000). When asked to choose between the mean and the median for describing sets of data, students predominantly chose the mean, even when outliers within the sets of data made the mean less indicative of center than the median. On a test item containing a data set structured so that the median would be a better indicator of center than the mean, only 4 percent of twelfth-graders correctly chose the median to summarize the data set and satisfactorily explained why it was the appropriate measure of center to use. The lack of use of the median could be explained by lack of familiarity with the statistic. On another item, only 33 percent of

twelfth graders were able to find the median when given a set of unordered data. It seems that since most students are familiar with the mean, and not the median, they think that the mean is the only measure of center that may be used to summarize data sets.

Watson and Moritz (1999b) demonstrated the usefulness of Biggs and Collis' model for identifying different levels of understanding of problems involving measures of center. Their study gives insight about students' understanding of concepts of average and how they use them as representative measures. Written tests were administered to classes of students in grades 3 through 11. They found that response levels (all within the concrete symbolic mode) generally improved with grade level. At the unistructural level, students demonstrated single, colloquial meanings for the concept of average and its application. At the multistructural level, students demonstrated some understanding of averaging within the context of the problem given, but sometimes did not demonstrate a sense of representativeness, or else were not able to obtain a correct numerical answer for the type of average they had chosen to use in response to the question. At the relational level, students coordinated the idea that a measure of center is representative of a data set with the correct computation of the measure of center. Even at the relational level, however, some students did not recognize the influence that outliers had upon the computation of the mean. As a result of a low percentage of students in the study being classified at the relational level, Watson and Moritz (1999b) set it as a goal "to help students build an early concept of average which is multifaceted and based on empirical data sets, followed by differentiation of how different measures of average represent data sets and examples from social settings" (p. 37).

Watson and Moritz (1999a) further detailed the structure of the different levels of understanding of average, and also documented the longitudinal development in regard to understanding of the concept of average using data from 94 students in grades 3 to 9. At the “preaverage” level of understanding, students formulated imaginative stories in order to answer questions, and sometimes did not even see averages as numerical measures. Students at the next level, who had a single colloquial usage for average, simply had some idea of what an average is from out of school personal experience. Students with multiple structures for average had more than one single colloquial idea, and described averages using terms like “most,” “middle,” “in between,” or the “add and divide” algorithm. At the next level, “representation with average,” students mentioned the relationship between the average measure and the data set itself, and as a result often checked answers for reasonableness. At higher levels, students saw the need to reverse the algorithm for calculating the arithmetic mean in problems, and were able to do so successfully in one or more contexts. Longitudinal data showed that after 3 or 4 years, all students tended to develop an appreciation of the representative nature of averages. It also showed that students tended to progress to higher levels of understanding as they got older, rather than remaining at the same level or dropping levels.

Measures of Spread and Measures of Center in Concert

Mevarech (1983) investigated college students’ ability to calculate variances and arithmetic means. In the first phase of the study, the participants were 56 college freshmen who had taken an introductory statistics course and 47 sophomores who had experienced two introductory courses. When presented with a questionnaire containing

problems that had been solved either correctly or incorrectly, students revealed misconceptions they held about calculating means and variances. Several students performed erroneous operations that involved adding weighted averages and adding variances. In the second phase of the study, 139 freshmen majoring in education were given the same questionnaire before taking an introductory college statistics course. Then, one group of the freshmen was taught via a conventional lecture/discussion strategy, while the other group was given instruction in which they were given frequent feedback about performance and were engaged in activities to correct misconceptions. At the end of the course, the second group had a higher mean score on problems involving computation of means and variances than the first group. Mevarech stated that this was evidence that students need to be engaged in activities aimed at dispelling misconceptions about calculating means and variances.

Reading and Pegg (1996) explored a slightly different aspect of students' understanding of data reduction using measures of center and spread. Rather than investigating students' proficiency in calculating formal measures of center and spread, they asked them to find the typical values in one data set presented numerically and one data set presented graphically. Their study included 180 students, 30 each from grades 7 through 12. In coding different levels of sophistication in interview responses, they used Biggs and Collis' (1982, 1991) model of cognitive development along with the notion that there can be several cycles of levels of development within any given mode. Using this perspective, they identified 9 levels of sophistication in students' responses. The first 3 levels represented a UMR cycle within the ikonic mode, while the last 6 levels

represented 2 UMR cycles within the concrete symbolic mode. There was a steady upward progression in levels of response with the ages of the students. At the lowest level of development, students' responses indicated that they did not understand the requirements of the question. At the highest level of development, students responded to the questions using both a measure of center and a measure of spread. Only 4 students' responses were classified at the highest level for the first task, and only 1 response was classified at the highest level for the second item. The modal level was the seventh, where students completed the problems by responding with one simple statistic, usually a measure of central tendency. These results suggest that students are far more likely to use measures of center than measures of spread when asked to describe data sets.

Mooney (2002) also detailed levels of student understanding of tasks requiring the use of measures of center and spread in the M3ST framework. Students responding to Mooney's tasks at the idiosyncratic level of thinking could not describe sets of data using measures of center or spread. Students at the transitional level described the center and spread of data using invented measures that are partially valid. Among the responses at the quantitative level, the center and spread of data were described using either flawed procedures or correct and valid invented measures. At the highest level in the M3ST framework, the analytical level, students demonstrated the ability to describe data using correct and valid measures of center and spread. These levels of understanding described in the M3ST framework add to our understanding of the levels through which students may progress as they become more competent in using measures of center and spread to reduce data.

Groth (2002b) investigated the possibility of extending the structure of the M3ST framework for the purpose of characterizing the thinking of high school students in regard to reducing data using measures of center and spread. Four students from an AP Statistics course participated in clinical interviews shortly after a unit in which they studied several measures of central tendency and spread. Many of the responses to questions administered during the clinical interviews could be categorized using the existing M3ST framework. In terms of the M3ST framework, none of the students interviewed exhibited higher than an analytical level of understanding of measures of spread. In fact, some students had an algorithmic conception of measures of spread, especially in regard to standard deviations. However, a higher level of thinking than “analytical” in regard to the use of measures of center was identified. Rather than using just one valid and correct measure of center to describe sets of data, one student interviewed was able to consistently use several different measures of center to describe “typical” values in sets of data. The study provided evidence that it would be feasible in this study to use a conceptual and theoretical structure similar to that of the M3ST framework in order to describe high school students’ understanding of measures of center and spread.

Organization of Raw Data Sets

Research aimed at refining and extending the structure of the M3ST framework has also helped provide a clearer picture of the levels of sophistication through which students progress in organizing data. Mooney, Langrall, Hofbauer, and Johnson (2001) conducted clinical interviews with middle school students in which they asked the

students to organize raw sets of data. Four levels of thinking about these tasks were identified as a result of the study. Students at the idiosyncratic level, the first of the four levels, did not attempt to group or order data. At the next level, transitional, students would group or order data, but not do so in a summative form. They did group or order data in summative form upon reaching the quantitative level. At the highest level of sophistication in the framework, students would group or order the data in summative form, and would create original categories, sometimes not explicitly represented by given data points (e.g., students at the highest level, in response to one task, grouped different types of fish under the original category “fish”). The study served to modify the descriptors concerning data organization published in the original M3ST framework (Mooney, 2002).

Summary

Researchers are still working on painting a complete picture of students’ thinking while organizing and reducing data. Research on students’ use of measures of center and spread is fairly abundant, while research detailing students’ abilities to group and order data has just begun to accumulate. Currently, we know that many students have significant conceptual gaps in their understanding of measures of center and spread. Research shows that students tend to have inflexible, shallow understandings of the algorithms for calculating means and variances. Some students’ understanding of means and variances is so purely algorithmic that they do not realize that the measures help to summarize data sets. They tend to use arithmetic means to summarize data sets even when it is not appropriate, and tend not use measures of spread when it would make sense

to do so. These problems are not universal, however, since the research does indicate that some students do operate at high levels of sophistication in thinking in regard to data reduction, making proper use of both measures of center and spread to summarize data.

Representing Data

Many currently available computer software programs have the ability to produce multiple graphical representations of raw data. Researchers have studied students' abilities to represent data both with and without the assistance of computers. Studies investigating students' abilities to represent data graphically without the assistance of computers have dealt with several topics. These topics include production of line graphs, the types of errors made in producing graphs of verbally described data sets, the ability to complete partially constructed graphs, and the developmental levels which occur along the way to becoming more competent in representing data. The computer's role in helping students represent data has only recently been considered by researchers. In particular, researchers have begun to describe how students function within a computer environment that assists with the representation of statistical data. They have also discussed how computer environments might be modified to help students become more proficient in representing data.

Representing Data without the Assistance of Computers

Padilla, McKenzie, & Shaw (1986) investigated students' abilities to represent data using line graphs. In their study, 625 students in grades 7-12 completed a multiple-choice instrument assessing their competence in subskills related to producing line graphs. Students correctly answered about half of the items. Average scores per grade level on

the instrument showed a general pattern of increase in grades 9-12, with the exception of a slight decrease in performance in grade 11. Padilla, McKenzie, and Shaw also reported performance on items related to specific subskills related to producing line graphs. Eighty-four percent of the students in the study could plot points and determine the x- and y-coordinates of any given point. However, only 32% could properly scale axes. Only 26% demonstrated the ability to use a best fit line. Forty-six percent were able to assign variables to axes. Fifty-seven percent were able to interpolate and extrapolate from bivariate data. Forty-nine percent could describe relationships between two variables in a line graph, and 47% demonstrated the ability to interrelate graphs. On the basis of the results of the study, the researchers recommended that teachers pay more explicit attention to helping students develop the subskills needed to produce line graphs.

Berg and Phillips (1994) also examined students' ability to construct line graphs. In their study, 20 seventh-graders, 21 ninth-graders, and 31 eleventh-graders each participated in two rounds of clinical interviews. In the first round of interviews, the researchers examined students' ability to construct and interpret line graphs set in several different contexts. In the second round of interviews, students were given Piagetian tasks related to locating points in three dimensions, multiplicative seriation, multiplicative measurement, and proportional reasoning. The results of the first round of interviews showed that fewer than 1/3 of the eleventh-graders were able to use age vs. weight data to correctly construct a line graph. Much of the performance related to constructing graphs seemed to be uniform throughout the grade levels, with the only differences existing in ordering or scaling axes and drawing distance vs. time graphs. Students who passed the

tasks presented in the second round of interviews did much better on the graphing interviews than those who did not. In essence, higher levels of cognitive development (in the Piagetian sense) seemed to be positively related to the ability to construct and interpret line graphs. Berg and Phillips (1994) concluded, “The evidence from this study suggests that these logical thinking abilities may indeed be the necessary tools that enable a person to construct and interpret graphs. The evidence also indicates that cognitive development is worth consideration when a researcher is attempting to determine the origins of graphing difficulties” (p. 331).

Mevarech and Kramarsky (1997) studied students’ ability to construct graphs of data stated in the form of a story. Ninety-two eighth-grade students were asked to construct graphs to illustrate the claims of three different students who each made different claims about how the amount of time they spent studying influenced their grades. Students were asked to construct the graphs once before instruction, and once after a formal unit on graphing in which they were introduced to bar graphs, histograms, line graphs, pictograms, pie diagrams, and linear graphs. The students who constructed correct graphs all chose to use line graphs. Students who constructed incorrect graphs displayed three types of alternative conceptions. One group of them chose to construct only one point that would have appeared on a correct line graph, or one bar that would have appeared in a correct histogram. A second group dealt with the covariation in the situation by constructing two graphs rather than displaying the variables of “grade” and “time spent studying” on the same graph. A third group seemed to think that each situation should be represented by a line graph, and that line graphs are always increasing. Accordingly, they

adjusted the points and/or scales in their representations so that the result of their efforts was a line graph which appeared to be increasing. Students who held alternative conceptions tended to be quite stable in using the same alternative conception to construct each graph they were asked to produce. Students' ability to produce correct graphs seemed to improve only slightly after the instructional unit. Mevarech and Kramarsky (1997) concluded, "this study indicates that the transition from verbal descriptions to graphic representations is associated with various kinds of alternative conceptions that were robust in resistance to traditional instruction about graphing" (p. 261).

Data from the NAEP provide some information about U.S. students' abilities to complete partially constructed graphs (Zawojewski & Shaughnessy, 2000). Some of the items on the examination required students to complete a partially constructed graphical display using data presented in a table. Both eighth- and twelfth-grade students performed well on these tasks. 77 percent of eighth-graders and 90 percent of twelfth-graders were able to successfully complete the displays. Unfortunately, no information is available about the specific types of graphs the students were asked to complete, since the graphical construction items were classified as secure. Grade 4 students showed strong performance on items requiring the completion of pictographs and bar graphs requiring only the use of additive reasoning, with over half responding correctly. Their performance was weaker on an item requiring some proportional reasoning in the construction of a bar graph, with only 36 percent answering correctly.

Mooney (2002) documented levels of sophistication in understanding in middle school students for the process of representing data using several different types of graphs. He found that students using idiosyncratic reasoning either did not construct displays for data, or else constructed displays that are both incomplete and unrepresentative of the data. They also did not complete partially constructed atypical displays. At the transitional level, students constructed data displays that were partially correct, and they made some progress in completing partially constructed atypical displays. Upon reaching the quantitative level, students were able to construct complete and representative data displays, perhaps with only a few flaws. They also successfully completed partially constructed atypical displays. At the highest level of reasoning, analytical, students were able to construct complete and representative displays while taking the characteristics of the data set and its context into account.

Representing Data with the Assistance of Computers

Although numerous existing computer software programs allow students to produce several representations of statistical data quite quickly (Ben-Zvi, 2000), existing technology does not help them decide which type of representation is most appropriate for any given situation. Biehler (2001) noted some of secondary students' poor choices in deciding upon representations for data when using a computer software package with the capability to produce several different types of graphs. Many of the students in his study did not take advantage of the software's ability to produce multiple graphs of data. The software also had the ability to produce summary statistics, so some students were content to use the summary statistics to provide pictures of data sets and did not utilize

the software's graphical capabilities. Many students in the study did not use box plots to represent data in situations where it would have been appropriate, seemingly not appreciating the fact that "box plots can be used to see properties of distributions such as symmetry and skewness that cannot be well-defined in empirical distributions" (p. 181). From Biehler's study, we learn that technology with the capability to produce multiple graphical representations for statistical data does not necessarily help students develop more sophisticated thinking about representing data.

Friel, Curcio, and Bright (2001) provide some insight about why students such as those in Biehler's (2001) study do not develop more sophisticated thinking about representing data by using technology. In order to develop more sophisticated thinking within this process, one must develop both graph comprehension and graph sense. Graph comprehension "includes a consideration of what is involved in constructing graphs as tools for structuring data and, more important, what is the optimal choice for a graph in a given situation" (p. 145). They state that graph sense "develops gradually as a result of one's creating graphs and using already designed graphs in a variety of problem contexts that require making sense of data" (p. 145). While technology may ultimately be shown to play an important role in helping represent data to aid in developing graph comprehension, it is noted that "often, when students move to use technology such as computer graphing programs, they demonstrate a lack of understanding of the relationships among the graph, the type of data, the purpose of analysis, and the judgment task" (p. 151). Hence, graph comprehension and graph sense are not given a chance to develop. As a result, Friel, Curcio, and Bright (2001) recommend that researchers

investigate the possibility of helping students use technology that allows them to invent or reinvent their own graphical representations for displaying data.

Summary

Research provides important information about students' abilities to represent data. Students have difficulty with many of the subskills needed to produce graphs (e.g., scaling axes). Their ability to produce graphs seems to improve as they become more cognitively mature. Producing graphs of verbally described data sets presents a special set of difficulties, as does producing graphs whose construction requires proportional reasoning. Technology can help students develop sophisticated understandings of different data representations. However, the use of statistical software packages does not, by itself, guarantee the development of more sophisticated thinking about representing data.

Analyzing Data

As one of the most complex statistical thinking processes, analyzing data has received a great deal of attention from the research community. At least three strands in this line of research are readily apparent. The first strand is that which has focused upon an investigation of individual cognition as students engage in informal data analysis. The second strand can be characterized by a focus upon the dynamics of the classroom social setting during informal data analysis. The third strand is that which deals with investigating students' notions about formal inference and its intuitive foundations. My discussion of literature surrounding the process of analyzing data will be organized around the three strands.

Informal Data Analysis and Individual Cognition

Batenero, Estepa, Godino, and Green (1996) investigated students' preconceptions about statistical association. In the study, 213 students ages 17-18 completed a questionnaire in which they were asked to analyze data presented in 2×2 , 2×3 , and 3×3 contingency tables. The researchers found that three different factors contributed to the difficulty of the tasks: the size of the table, the students' own personal theories about the context of the problem, and some students' inability to deal with the concept of inverse association. The researchers also identified two correct (or partially correct) strategies for analyzing the data in the tables, and three incorrect strategies. One correct strategy involved viewing the rows and/or columns in the table as distributions, and comparing them as such. A second correct strategy involved examining and comparing the frequencies of cases for and against the variables of interest. One group of students using incorrect strategies felt that there could be no dependency between variables presented in a table unless some of the cells contained zero values. A second group using an incorrect strategy confused an inverse association between variables with independence between variables. The third group with an incorrect strategy used only part of the information presented in the tables to make a decision about association between variables, sometimes using as little as one cell of the table to come to a conclusion. Batenero, Estepa, Godino, and Green (1996) concluded, "All of these findings show the complexity of a topic that is simple in appearance" (p. 167).

Friel (1998) studied students' ability to analyze data presented in the form of box plots. After an instructional unit including box plots, the eighth-graders in her study had

little trouble in responding to questions about box plots that required them to “read between the data.” They all demonstrated the ability to successfully analyze data between the quartiles of each box plot and between the extreme values. Almost all of them also successfully identified “outliers” in the data, with only one student unable to recall the concept of outliers. In questions requiring students’ to “read beyond the data,” students demonstrated the ability to compare the relevant features of box plots, such as medians, quartiles, and ranges. Some students also discussed the possible distributions of data within each box and used features such as clustering and spread in their analysis. Overall, the students studied demonstrated the ability to use box plots as tools for statistical data analysis.

Watson and Moritz (1999a) examined students’ abilities to make comparisons between two data sets presented in forms similar to bar graphs and line plots. In clinical interviews with 88 students in grades 3-9, they presented one task where the number of items in each data set was the same, and one task where the numbers of items in each set differed. Some students used simple visual strategies to compare the data sets, some used numerical strategies, and still others used visual and numerical strategies in concert. Using Biggs and Collis (1982, 1991) model of cognitive development, two UMR cycles within the concrete symbolic mode of development were identified. At the highest level of sophistication within the first cycle of development, students demonstrated the ability to draw correct conclusions, but only when the two data sets being compared were of the same size. At the highest level of sophistication in the second UMR cycle, students employed proportional reasoning in some manner in order to compare data sets of

unequal sizes. Some of the students who did not initially recognize problems inherent in comparing two groups of unequal size were later pointed to these problems by the researchers, and some of these students answered at a higher level after being given this information.

Mooney (2002) also identified various levels of sophistication among middle school students in response to data analysis tasks. In the formulation of the M3ST framework, four levels of response within the analysis process were identified. Students at the idiosyncratic level were not able to make comparisons within or between data sets. Any inferences that were made were not soundly based on the data at hand and its context. At the transitional level, students demonstrated the ability to make a single correct comparison between and within data sets. Inferences that were made were based on the data at hand, but may be only partially reasonable. Upon reaching the quantitative level of thinking, students made either local or global comparisons within and between data sets. Inferences were reasonable and were based upon the data and its context. At the highest level of thinking identified, the analytical level, both local and global comparisons were made within and between data sets. Reasonable inferences were made by considering the data, its context, and multiple perspectives.

The analyzing and interpreting data construct within the M3ST framework was extended by the research of Mooney, Langrall, Hofbauer, and Johnson (2001). Clinical interview items in this study allowed students to demonstrate the attainment of multiplicative reasoning in analyzing data sets. Students at the idiosyncratic level did not use relative thinking in the tasks where it was applicable. Transitional level students used

relative thinking, but did not numerically describe the data. At the quantitative level, students used relative thinking and numerical strategies to make comparisons. However, when they did so, they considered only part of the data set, ignoring some parts that contained relevant information. Upon reaching the analytical level, relative thinking and numerical strategies were used in order to compare all relevant features of data sets. The descriptors generated in this study can be merged directly into the overall structure of the M3ST framework.

Further information regarding students' individual cognition while engaged in informal data analysis tasks is provided by NAEP data (Zawojewski & Shaughnessy, 2000). In general, US students did quite poorly on tasks related to data analysis. Fewer than half demonstrated the ability to interpret graphs, make decisions from data presented in graphs, and explain reasoning behind any interpretations and decisions that were made. Less than 25% of all twelfth-graders responded correctly to an item requiring them to analyze a scatter plot and make a decision based upon the data presented in it. Likewise, less than 25% were successful with an item requiring the analysis of a histogram. The item required students to use proportional reasoning to read the histogram, and then to make predictions beyond the given histogram to another population. NAEP data seem to indicate that along with being the most structurally complex of the statistical thinking processes discussed so far, analyzing data is also the most challenging of the processes for students to master.

Informal Data Analysis in the Context of Social Interaction

As part of their study documenting two UMR cycles of levels of sophistication in thinking present among sixth-graders analyzing data cards, Watson, Collis, Callingham, and Moritz (1995) discussed the quality of responses to the data cards tasks that were given by students working in groups. Group responses to tasks were generally more sophisticated than the individual responses given during the clinical interviews that were used to identify and describe levels of sophistication in thinking in analyzing the cards. None of the group work produced was classified at the lowest level of thinking identified during clinical interviews. The lowest level of thinking evident in group responses was multistructural within the first cycle of development. While several did not exceed that level of response, evidence of thinking beyond the level were documented. Groups that did show evidence of progressing to higher levels of thinking were those who considered multiple aspects of the data cards in their analysis, and those who used means to compare groups of people on different characteristics. The researchers concluded that the levels of thinking established and used in the study may provide a useful means for evaluating the quality of work students produce while interacting in groups during data analysis activities.

Cobb (1999) provided a sociocultural perspective on the development of the statistical thinking process of analyzing data by discussing the results of a 10-week teaching experiment conducted with a class of seventh-graders. The goal of the teaching experiment was to get students to see data sets as entities, or distributions, rather than just sets of individual points. Cobb also wanted students to develop multiplicative reasoning

abilities in comparing sets of data. During the teaching experiment, students worked in pairs, using computer mini-tools to analyze data sets provided by the researchers. After working in pairs, they used an overhead projection unit to share the results of their statistical data analysis with the entire class. As students used the first mini-tool to analyze data, the classroom discourse indicated that students viewed data sets as collections of data rather than as distributions. After some time, the researchers introduced a second mini-tool designed to help students see data sets as distributions, allowing them to partition data sets to aid in the development of multiplicative reasoning. After using the second mini-tool, classroom discourse changed to indicate that students had begun to see data sets as distributions, and they began to realize that an important purpose of statistical data analysis was to compare distributions. Cobb's (1999) study shows how students can gradually come to understand concepts underlying informal data analysis within a classroom context.

McClain and Cobb (2001) further elaborated on the classroom context and the sociomathematical norms which emerge during the course of a teaching experiment with the same goals and computer mini-tools discussed in Cobb (1999). McClain and Cobb (2001) noted that students had several questions about how the data sets they were to analyze were produced. The researchers would talk through the data production process with students in order to help them become more familiar with the data they were to analyze. Also, during whole-class discussions, notions about what were to be considered "important" parts of distributions emerged. Students began to see ranges, medians, maximum and minimum values, and the number of data points on each side of a value as

important characteristics of data sets. During class, students developed multiplicative reasoning by using the second computer mini-tool to begin to reason proportionally about intervals of sets of data. Students were not introduced to a number of specific types of graphs, but rather were exposed to social processes which constitute the essence of statistical thinking. In this process, they developed understandings of statistical measures as representative of a data set rather than as algorithms. They also developed multiplicative reasoning and came to see data sets as entities rather than as collections of individual points.

A study conducted by Ben-Zvi and Arcavi (2001) helps illustrate the necessary components in becoming proficient in computer-supported exploratory data analysis (EDA). During the study, the interaction of two seventh-grade students was videotaped as they engaged in statistical data analysis tasks. Ben-Zvi and Arcavi noticed and documented several of the changes that happened as the students became more proficient in the process of data analysis. They began to become more adept at focusing on relevant aspects of data sets rather than irrelevant ones. They demonstrated the ability to look locally at tables and graphs, but also began to see them more globally as the study progressed. An indicator of this was that they would transfer between viewing data locally and globally by setting appropriate scales on the graphs they worked with. Eventually, the students needed to realize that statistical data does not behave as deterministically as mathematical objects encountered in subjects such as algebra. They also gradually became aware of the fact that the lines used in statistical graphs are sometimes “meaningless,” in the sense that they may just be used to show trend. Each of

these competencies were developed within the context of collective informal statistical data analysis.

Whereas Ben-Zvi and Arcavi (2001) documented the components necessary for EDA which students developed, Biehler's (2001) study documented some of the difficulties and cognitive obstacles high school students encountered within the context of computer-supported informal data analysis. Biehler (2001) conducted clinical interviews with pairs of high school students engaged in data analysis tasks. Students experienced a number of difficulties with the tasks they were presented. Some students had difficulty overcoming the conception that cause and effect can be established as soundly in data analysis as it is in mathematics. They also had difficulty understanding that the reporting of exact numerical results is valued differently in mathematics than it is in a data analysis setting. Some students did not take into account that numerous variables could impact and influence a given variable. Students often did not take sample size into account in interpreting summary statistics. Biehler also found that students experienced difficulty when "common sense" notions about context conflict with statistical evidence. Biehler's (2001) study reveals a number of difficulties with data analysis tasks which need to be taken into consideration by researchers, teachers, and curriculum developers.

Research Concerning Understanding of Formal Inference and its Foundations

A study by Confrey and Makar (2001) illustrates how informal data analysis activities can lead students to see the need for formal inferential methods. The research was conducted during a series of workshops and a summer institute aimed at helping teachers develop an understanding of statistical concepts and the process of data analysis.

Teachers were initially engaged in data analysis by examining data from a very familiar context: students' scores on a statewide standardized test. Data analysis was supported by the computer software program Fathom. In the initial sessions, teachers tended to focus upon examining individual data points from these data sets rather than trying to identify trends in the data. It wasn't until later sessions that teachers' data exploration activities led them to see the need for formal inferential techniques. This shift in thinking occurred when the researchers presented the teachers with the task of determining the "fairness" of a set of dice. The need to determine whether or not the dice were loaded helped lead the teachers to see the need for concepts from formal inference, such as Z scores and t tests. Ultimately, teachers also came up with interesting statistical statements about the standardized test data they had been working with. In engaging in the workshop, teachers became more statistically proficient as they gained information about students' performance on statewide standardized tests.

Along with experiences in informal data analysis, an understanding of the concept of sampling distribution is vital if students are to understand and meaningfully apply techniques of formal inference. Chance, Garfield, and Del Mas (2001) investigated undergraduates' understanding of sampling distributions within the context of a study of the effectiveness of a software program designed to simulate the sampling process and form sampling distribution graphs. During the study, five different levels of sophistication in undergraduates' thinking about sampling distributions were identified. At level 1, which they called idiosyncratic reasoning, students simply had knowledge of words and symbols related to sampling distributions, but had no real conceptual

knowledge of them. At level 2, verbal reasoning, students could select correct definitions for sampling distributions and the central limit theorem, but could not apply the definitions to solve problems. Level 3, transitional reasoning, was characterized by the ability to identify some of the dimensions of the sampling process along with an inability to fully integrate them. Students at level 4, procedural reasoning, could identify all of the dimensions of the sampling process, but lacked some confidence in integrating them to match a given sampling distribution with its corresponding parameters. The highest level of reasoning observed, level 5, was that at which students could identify the dimensions of the sampling process, integrate them, and demonstrate confidence by stating the dynamics of the process in their own words.

Konold, Well, Lohmeier, and Pollatsek (1993) also examined undergraduates' understanding of the foundations of formal inference within the context of instruction aided by a computer simulation. The purpose of their study was to describe students' understanding before and after instruction guided by a simulation tool. It was conjectured that the tool would help students understand the law of large numbers, that "statistics of larger samples (e.g., means) more closely approximate the corresponding parameters of populations and are thus less variable than those of smaller samples" (Konold, Well, Lohmeier, & Pollatsek, 1993, p. 299). Twelve undergraduates recruited from psychology courses took part in interviews. The students were asked to solve tasks requiring an understanding of the law of large numbers both before and after 75 minutes of instruction aided by a computer simulation illustrating the law. The researchers found that after instruction, students did have some understanding of the law of large numbers, but were

not sure about when it was and was not applicable. Most of the students were able to understand why means of large samples exhibit less variability than means of small samples. However, misunderstandings about percentages and the need for random sampling blocked many students from attaining a complete understanding. Their findings echoed those of Well, Pollatsek, and Boyce (1990), who experienced moderate success in using computer simulations to help students come to understand how sample size affects sampling distributions.

Saldanha and Thompson (2001) conducted a teaching experiment in which they attempted to help tenth- and eleventh-graders begin to conceive of collections of samples as distributions. In a 20-session teaching experiment, computer simulation activities were used to investigate whether or not the outcomes of stochastic experiments were “unusual.” In a reported episode from the experiment, the researchers had students draw a sample of candies from a sack containing half red and half white candies. Students were not told the proportion of each color in the sack. After drawing a number of samples and seeing that there were more white than red candies, they concluded that the sack contained more whites than reds. They were surprised when the instructors told them that the proportion of red candies in the sack was the same as the proportion of white candies. The students came to the conclusion that it would be best to draw numerous samples from the bag in order to reach more appropriate conclusions. At that point, the researchers engaged the students in conducting probability simulations for drawing samples to form a sampling distribution. In using the software program ProbSim to simulate drawing numerous samples, students realized that the samples they had initially

drawn were not that unusual to observe within the overall sampling distribution. In coming to this realization, the students had successfully developed sound intuitive notions about ideas underpinning formal inference.

Shaugnessy, Watson, Moritz, and Reading (1999) have also investigated students' conceptions about some of the ideas underpinning formal inference. In their study, 324 students from grades 4, 6, 9, and 12 gave written answers to tasks asking them to estimate the spread of a distribution within the context of a sampling situation. Only 30% gave a reasonable range of spread for the tasks administered. Few students used words related to variability or spread in responding to the questions. There was no clear pattern of growth through grade levels regarding the measurement of variability in the situations presented in the tasks. Many students overestimated the amount of variability that would be present in the given situation. When Torok and Watson (2000) conducted a follow up study making use of clinical interviews with tasks similar to those of Shaugnessy et al. (1999), their findings were somewhat contradictory in the sense that they did see a clear pattern of improvement of appreciation of variation in sampling situations as students attained higher grade levels. They also described levels of sophistication in response to the tasks administered: students progressed from a weak appreciation of variation to an isolated appreciation of concepts of variation and clustering. They then developed a more complete, yet inconsistent appreciation of the two concepts. Students responding at the highest level demonstrated a complete and consistent grasp of both concepts.

Summary

Several themes emerge from the preceding discussion of students' abilities to analyze data. While instruction can be effective in helping students draw viable conclusions from data sets, students do sometimes draw unsophisticated and poorly supported conclusions in the process of analyzing data sets. Research suggests that teachers need to move students towards seeing data sets as entities rather than as individual points. Teachers then must help students reason multiplicatively rather than additively when comparing sets of data. Teachers charged with the duty of helping students understand formal inference must be aware of the incomplete and incorrect conceptions students tend to bring to the topic, and how some researchers have worked to find ways to overcome them. Technology and social interaction are two tools that appear to have helped students develop more sophisticated strategies for analyzing data.

Collecting Data

Students' thinking in regard to data collection is beginning to be investigated by the research community. Three strands of research within this statistical thinking process are identifiable. The first strand of research findings describes students' knowledge of data collection methods. Specifically, researchers have described students' ability to differentiate between random and nonrandom situations, and also how students' knowledge of data collection methods impacts their analysis of data. A second strand of research deals with individuals' abilities to critique samples collected from populations. Within this strand, students' understandings of how randomization, sample size, and sources of bias affect the representativeness of samples are treated. A third strand of

research describes how students' notions about collecting data develop during instruction. Studies within this strand have been conducted at the college level, and they provide insight about the cognitive hurdles students encounter as they learn about collecting data.

Students' Knowledge of Data Collection Methods

Data gathered by means of random selection are highly valued in several different types of data collection designs. Konold et al. (1991) described students' perceptions of what constitutes a random situation. The study compared "expert" perceptions of randomness to those of "novices." Four statistics instructors and one professional statistician were given the "expert" label, and 20 psychology undergraduates were given the label "novice." Each of the participants was given 18 different cards, upon which the description of a situation was written. Participants were to classify each situation as "random" or "non-random," and then justify their classification during an interview. The researchers identified five different categories of justifications. One category of response identified was that some students would classify a situation as random only if each of its outcomes had the same probability. In another category, some would categorize a situation as random only if it had more than one possible outcome. A third category of response was to classify a situation as random only if there was no prior knowledge about the outcome, so that any predictions would be impossible. The fourth category of response was characterized by an unwillingness to call a situation "random" if any control of outcomes was possible. A fifth category of response, exhibited only by experts, was to determine if a situation was random by comparing it to some standard model of randomness.

Knowledge of the methods used to collect any given set of data can influence how students engage in data analysis. McClain and Cobb (2001) gave middle school students several different sets of contextualized data to analyze during a teaching experiment. They found that students had several questions about how the data was collected when given the sets. Listening to students' concerns, the researchers answered all questions students had about how the data they were to analyze was collected. The students were able to analyze the data meaningfully once their questions about the data collection process had been answered. Batanero, Estepa, Godino and Green (1996) also documented how students' perception of contextual factors surrounding data collection played a role in students' analysis of data. The researchers did not document any cases of students asking questions about the methods of collection of the data presented in contingency tables that they were to analyze. However, some students involved in their study seemed to form their own "personal theories" about the contextual factors surrounding the data. Some of the "personal theories" students generated contributed to a low level of sophistication in data analysis. These studies show that when students are given sets of contextualized data, they often are not satisfied to analyze the data without considering the context from which the data arose and the methods used to collect the data.

Students' Abilities to Critique Samples

Watson and Moritz (2000a) detailed some of the notions students bring to the process of critiquing sampling methods used in studies. They interviewed 62 students from grades 3, 6, and 9 on the basis of responses students had given to a written questionnaire about sampling. The interview questions investigated the students' personal meanings for

the word “sample,” their thoughts on selecting samples, and how they thought increasing sample size affects the representativeness of the sample. The researchers identified six levels of sophistication in response to the set of sampling tasks. At the lowest level, students had only a simple notion of the concept of “sample,” and did not demonstrate understanding of the roles of methods of selection or sample size in choosing a statistical sample. At subsequent levels of sophistication, students began to realize that some type of random or stratified selection was often desirable in obtaining a representative sample. They subsequently also began to realize that large random samples were more representative of the population from which they were drawn than small random samples. Students at the highest level of thinking recognized bias in sampling in multiple contexts and understood the roles of randomization and sample size in producing a representative sample. The results suggest that students come to understand the roles of sample size and sample selection methods in sampling before they are able to recognize bias in multiple contexts.

Watson and Moritz’ (2000b) analysis of a written survey given to students in grades 3-11 sheds further light upon the development of students’ notions of sampling. They detailed students’ development of appropriate meanings for the word “sample.” They also analyze how well students can apply sampling concepts within a context and question claims made on the basis of samples. The levels of thinking revealed in understanding of the word “sample” ranged from a low level of thinking in which the word held no clear meaning, to the highest level of thinking in which students recognized that a sample is a small part of the overall population and spoke of representative

samples. In response to tasks asking students to critique claims made upon the basis of samples, students' responses ranged in sophistication from those who offered no criticism of situations in which bias would naturally occur to those who recognized the need for samples to be representative and unbiased. When the questionnaire was administered to some of the same students 2 to 4 years later, it was found that students tended to improve by one or two levels of sophistication in thinking, but rarely improved by more than two levels. The study suggests that a need exists to help students identify bias in multiple contexts as they develop meanings for sampling terminology and sample selection methods.

Data from NAEP provide information about a large representative sample of U.S. students' responses to tasks involving sampling. Zawojewski and Shaughnessy (2000) note that about $\frac{2}{3}$ of eighth-grade students taking the NAEP could correctly choose the sampling method that would provide the least biased results when given several choices of sampling methods in a multiple choice question. Students also did well on a secure test item designed to assess students' ability to recognize the potential for sample bias, with over half of grade 8 students and $\frac{3}{4}$ of grade 12 students producing correct answers. Over half of eighth-grade students were able to respond correctly to an item testing knowledge of the importance of sample size in producing a representative sample. The results seem to indicate that students tend to have some fairly solid ideas about sampling from a population despite the relative neglect of statistics in current school curricula.

Developing Understanding of Data Collection through Instruction

Derry, Levin, Osana, Jones, and Peterson (2000) described the development of undergraduates' statistical thinking ability in regard to data collection within the context of a semester-long statistics course with 22 undergraduate education majors. Task-based interviews were used to gather data from students before and after the course, and work produced by students over the course of the semester was also gathered. During the course itself, students engaged in studying techniques of experimentation, correlation, the structure of an effective statistical argument, and sampling techniques. A comparison of interview scores from before and after the course indicated that the course was successful in developing some of students' conceptions in regard to the collection of data. Students showed significant gains in being able to recognize the need for correlational evidence in some situations, the design of convincing experiments, and the grasp of the concept of random sampling. At the same time, after the course many students still tended to confuse the concepts of random sampling and random assignment. The researchers noted that it was difficult for students to break from a traditional instructional environment in favor of an inquiry-based one. The difficulties associated with the transition may have limited the quality of learning displayed.

Heaton and Mickelson (2002) also describe education majors' developing conceptions in regard to data collection during a semester-long statistics course. The researchers, who were also the instructors for the course, had the primary goal of helping students develop mature ideas about the process of statistical investigation. In the first major assignment for the course, students were to pose quantifiable questions of interest

about aspects of mathematics teaching and learning at their practicum sights. In the second major assignment, students were to design projects for elementary school children that would require the children to engage in data collection and analysis. Students had difficulty in collecting data and critiquing their own studies during the first assignment. Some formulated insignificant, relatively uninteresting questions. They seemed to struggle with how to ask quantifiable questions and then design data collection methods to match them. The collection of data was also made difficult by the often fast-paced nature of events in classrooms. The second assignment presented its own unique difficulties. In designing assignments, the education majors generally did not carefully take into account the fact that the choice of question to be investigated would have a substantial impact on all other phases of the investigation. In one instance, however, the assignment proved to be highly successful. An education major that engaged a sixth-grade class in a study of questions surrounding racial discrimination got children to informally discuss notions such as construct validity and external validity. The study shows that even though it may be a difficult task, it is possible to help teachers engage children in meaningful investigations involving the collection of data.

Summary

Research shows the complexity involved in understanding data collection. Students' views of data sets are largely shaped by perceptions about how the data in the set was gathered. The concept of randomization and data collection designs incorporating it take considerable cognitive effort for students to understand. The cognitive demand associated with formulating quantifiable questions of interest and gathering data to answer them is

also high. Tasks involving detecting bias and determining appropriate sample size are likewise non-trivial in nature. Despite the high cognitive demands associated with collecting data, research does indicate that many students are highly capable of advancing to sophisticated levels of thinking within this statistical process.

Relationship to this Study

The existing empirical research helped to shape the current study in numerous ways. Several of the studies discussed identify statistical concepts students seem to have difficulty in understanding. Statistical thinking tasks, described in the next chapter, were designed to investigate the nature and prevalence of some of those difficulties and misconceptions among high school age students. Existing literature also provides several instances in which researchers have classified statistical thinking by levels of sophistication of response. These documented levels of response helped shape the lens through which interview data were coded by levels of response. Finally, gaps in the research base helped to shape the study. For instance, the current study sought to fill the gaps that exist in our understanding of the statistical thinking of high school age students and how well some of them understand the tools of formal inference. Since it was guided by, and situated within, the existing body of empirical research, this study added to the overall picture of students' statistical thinking abilities which existing research had already begun to paint.

CHAPTER III

RESEARCH DESIGN AND METHODOLOGY

In chapter 1, I stated that the purpose of this study was to construct a framework to describe the statistical thinking of high school students. In order to accomplish that goal, I constructed a set of statistical thinking tasks and administered them to high school students and recent high school graduates during individual clinical interview sessions. I analyzed the data from the clinical interview sessions in order to identify and describe patterns of response. The different patterns of response were then examined in light of the Biggs and Collis (1982, 1991) cognitive framework. This chapter documents my study design and methodology in detail.

Study Design

A qualitative design was chosen for the study for two main reasons. First, the purpose of the study was to describe statistical thinking processes. A qualitative design is well-suited to this purpose, since qualitative designs allow one to investigate intricate processes rather than just end products (Merriam, 1988; Strauss & Corbin, 1990; Bogdan & Biklen, 1992). Second, as evidenced by the literature review, little research exists that focuses specifically upon the statistical thinking of high school students. A need exists to explore and describe statistical thinking at the high school level and to develop theory about it. Qualitative designs are helpful for studies that involve exploring phenomena,

describing them, and developing theories (Morse, 1991). Given the purpose and the nature of the study, a qualitative design was deemed necessary.

Within the qualitative design, structured, task-based clinical interviews were used as a primary means of data collection. Structured, task-based interviews are those in which an interviewer presents one or more tasks to a problem-solver in a preplanned way (Goldin, 2000). In this study, interviews were also structured in the sense that all respondents received the same set of questions in the same order (Fontana & Frey, 2000). In such interviews, the problem-solver interacts with both the interviewer and a task environment (Goldin, 2000). Hence, interviews can be viewed as “active interactions between two (or more) people leading to contextually based results” (Fontana & Frey, 2000, p. 646).

Task-based interviews helped to focus the study upon describing students’ statistical thinking processes. Goldin (2000) noted that task-based interviews allow researchers to “focus research attention more directly on the subjects’ processes of addressing mathematical tasks, rather than just on the patterns of correct and incorrect answers in the results they produce” (p. 520). This means that task-based interviews were an appropriate choice of data collection within the overall qualitative design, since one of the reasons for choosing a qualitative design was to focus on thinking processes rather than simply on finished products. Goldin (2000) also pointed out that the emphasis on students’ thinking processes rather than on their ability to use algorithms to produce right or wrong answers fits well with the overall spirit of the current mathematics education reform movement. Therefore, descriptions of statistical thinking processes that can be provided by task-

based interviews are of interest to both teachers and researchers concerned with current issues in statistics education at the high school level.

The needs of teachers were taken into account in the decision to structure the task-based interviews, since it is valuable to consider a study's impact on practice (Sowder, 2000; Lester & William, 2002). One advantage to having a structured interview script as part of a study is that the script can later be used by teachers as an assessment instrument (Goldin, 2000). Teachers can gain insight about their students' statistical thinking by administering some or all of the script to students. They can then compare their students' levels of response to those documented in the study, and then use that information to design instruction to help students advance to higher levels. An unstructured interview can not be replicated as easily.

The fact that structured interview scripts are easier to replicate than unstructured interviews is also of interest to mathematics education researchers. According to Goldin (2000), the mathematics education research community is vulnerable to attack from opponents of reform, partially because the research base in mathematics education is perceived as lacking replicable results. Structured interview scripts can be replicated quite easily in subsequent studies. They allow for the investigation of detailed thought processes while they simultaneously help fill the need for replicable mathematics education research. Subsequent researchers may use the interview script to both extend and revise the framework constructed as a result of this study.

Participants

Purposeful sampling (Patton, 1990) was used in the selection of study participants. The goal of purposeful sampling is to “select information-rich cases whose study will illuminate the questions under study” (Patton, 1990, p. 169). The central questions of this study concerned the identification and description of different levels of statistical thinking. Therefore, a maximum variation sampling strategy (Patton, 1990) was used to select the purposeful sample. A maximum variation sample includes people who have had significantly different experiences in some area, and it allows the researcher to describe patterns of variation within the group studied (Patton, 1990). In this study, students were chosen on the basis of the different types of mathematics courses they had taken while in high school. The goal was then to describe the variation within the responses to interview tasks given by these students. Maximum variation samples are not drawn in order to make statistical generalizations to a larger population, but rather to describe variation and significant patterns within the group (Patton, 1990). Accordingly, in this study, I did not seek to statistically generalize my findings to all high school students, but instead sought to describe the patterns in thinking observed among the diverse sample chosen.

There were three different categories of participants interviewed for the study. The participants in the first category were college freshmen who had recently completed a semester-long high school statistics course. Participants in the second category were college freshmen who had recently completed a year-long high school statistics course. The participants in the third category were still in high school at the time of the clinical

interviews. Each of the participants in the third category was enrolled in Algebra, Geometry, AP Calculus, or AP Statistics at the time of the interview sessions. In total, fifteen students participated in clinical interviews. Three of the participants fell into the first category, four into the second, and eight into the third.

Students were recruited from each of the three categories by contacting several university instructors and high school teachers. Several instructors at a Midwestern university helped in the recruitment of students by mentioning my study in their classes and asking for volunteers. The first two categories of students interviewed included students from three of those instructors' classes. The very first category also included a student from a different Midwestern university who was recruited with the help of her former high school statistics instructor. Teachers at two different Midwestern high schools helped in the recruitment of students in the last category by discussing my study with their classes and asking for volunteers. Four students in the last category came from one of the high schools, and four came from the other high school. Figures 2, 3, and 4 summarize some of the important information about the students in categories 1, 2, and 3, respectively. Each of the students' backgrounds is further discussed in Appendix A along with some anecdotal information for each student.

Student	Year in school at time of study	Technology with which the student indicated familiarity
Lisa	College freshman	Simple pocket calculator
Kristen	College freshman	TI-83 and <u>Excel</u>
Laura	College freshman	TI-83

Figure 2. Students who had recently completed a semester-long high school statistics course

Student	Year in school at time of study	Technology with which the student indicated familiarity
Jeff	College freshman	TI-83, TI-86, <u>Fathom</u> , and <u>Minitab</u>
Hillary	College freshman	TI-83, <u>Excel</u> , and <u>Minitab</u>
Paul	College freshman	TI-83
Julie	College freshman	TI-83, <u>Excel</u> , and <u>Fathom</u>

Figure 3. Students who had recently completed a year-long high school statistics course

Student	Year in school at time of study	Course enrolled in at time of study	Technology with which the student indicated familiarity
Crystal	High school senior	AP Statistics	TI-83, <u>Excel</u> , <u>Fathom</u> , <u>Data Desk</u>
Bill	High school senior	AP Calculus	TI-89, <u>Excel</u>
Luke	High school senior	AP Calculus	TI-89, <u>Excel</u>
Daniel	High school freshman	Honors Geometry	Scientific, non-graphing calculator
Jessica	High school sophomore	Geometry	Scientific, non-graphing calculator, <u>Excel</u>
Nancy	High school sophomore	Geometry	Scientific, non-graphing calculator, <u>Excel</u>
Brooke	High school sophomore	Algebra	TI-82, TI-83, <u>Excel</u>
Rick	High school sophomore	Algebra	TI-82, TI-83, TI-85, TI-86, TI-89, <u>Excel</u>

Figure 4. Students still enrolled in high school at the time of the study

Interview Protocol

A set of tasks designed to assess students' statistical thinking across the processes of describing, organizing and reducing, representing, analyzing, and collecting data was administered during the individual interview sessions with students. In order to enhance the study's potential to advise high school statistics instruction, the specific content for each of the tasks came from current curricular recommendations for high school statistics courses. While there is no universal consensus about the statistical content which should be studied by high school students, recommendations have recently been set forth by NCTM (1989, 2000), the College Entrance Examination Board (2001b), mathematics education researchers (e.g., Biehler, 1997), and practicing statisticians concerned about improving the quality of statistics education (e.g., Cobb & Moore, 1997). In this section, I will begin by briefly discussing each of these recommendations. I will conclude the section by explaining how the literature discussed and pilot-testing of tasks helped guide the construction of the statistical thinking tasks administered during the study.

In its Curriculum and Evaluation Standards for School Mathematics, NCTM (1989) set forth recommendations for the statistical content to be studied by high school students. The recommendations were groundbreaking, since the study of statistics in high school was not prevalent in 1989. The statistics standard for grades 9-12 stated that all high school students should be able to read, interpret, and construct charts, tables, and graphical displays for data. All high school students were also to be able to flexibly use measures of central tendency, variability, and correlation to describe and analyze data. In

addition, all high school students were to understand ideas behind sampling, design experiments to test hypotheses, communicate results to others, and evaluate statistical claims made by others. NCTM recommended that college-intending high school students study distributions such as the normal, Student's t , Poisson, and chi square, and understand when it is appropriate to use each distribution in formal hypothesis testing and construction of confidence intervals. The intuitive foundations for formal inference were to be laid by having students construct and analyze sampling distributions using simulations. Technology was to be integrated appropriately in the teaching of all statistical content.

In the Principles and Standards for School Mathematics, NCTM (2000) updated its recommendations for the statistical content to be studied by all high school students. The data analysis and probability standard of the Principles and Standards is similar in spirit and content to the statistics standard of the Curriculum and Evaluation Standards, with a few notable exceptions. In the Principles and Standards, no mention is made of the study of formal inference by high school students, even though all students are still to gain an intuitive understanding of its foundations by constructing sampling distributions using simulations. The Principles and Standards also gave more specific recommendations for the types of graphical displays that are to be understood and used by high school students. Histograms, parallel box plots, and scatter plots are all listed as important types of graphical displays. Finally, it is notable that probability is merged into the data analysis standard, and is not given its own standard as in the 1989 document. Probabilistic content to be studied by all high school students included understanding the concepts of sample

space and distribution, computing and interpreting expected values for random variables, and understanding conditional probability, independent events, and compound events.

In both of its Standards documents, NCTM recommended that high school students study statistics throughout grades 9-12, and did not explicitly recommend that they take the subject as a separate course. Recommended content for a one-year course in statistics for high school students is given by the College Entrance Examination Board (2001b) in its course description for Advanced Placement (AP) Statistics. The course description for AP Statistics is quite ambitious. Accordingly, the College Board at no point recommends that all high school students take the AP course. The goals for the AP course exceed the goals set forth for all students by NCTM (2000). As recommended by NCTM, students are to engage in exploratory analysis of data in which several graphical displays and statistical measures are used. They are to become familiar with different methods of data collection and become proficient in planning, conducting, and critiquing surveys, experiments, and observational studies. AP students also are to study probabilistic content such as computations requiring formal probability theory, independent random variables, the normal distribution, and sampling distributions. In the AP course, students are to study formal statistical inference extensively. They are to understand when it is appropriate to use chi square tests and different types of t tests in formal inference. By the end of the course, students are to understand and use confidence intervals and hypothesis tests for both means and proportions. The inclusion of such an extensive study of techniques of formal inference sets the AP course recommendations apart from other recommendations for high school statistics.

Just as recommendations for the amount of formal inference to be studied by high school statistics students vary by curriculum document, recommendations for the amount of probability to be studied during a statistics course also vary. Cobb and Moore (1997) provide perspective upon how extensively probability needs to be studied in introductory statistics courses which include the study of formal inference. While maintaining that “probability is an essential part of any mathematical education” (p. 820), they also hold the position that a “first course in statistics should contain essentially no formal probability theory” (p. 820). Cobb and Moore (1997) support their position by explaining that “informal probability is sufficient for a conceptual grasp of inference” (p. 821). Students who understand that “the sampling distribution of a statistic...records the pattern of variation of the outcomes of, for example, many random samples from the same population” (p. 821) have enough understanding of probability to have a conceptual grasp of statistical inference. Therefore, content such as the theoretical derivation of probability distributions and formal rules for computing probabilities are best left for courses other than introductory statistics. This position stands in opposition to the AP Statistics course’s inclusion of the study of formal probability rules. It also suggests that Data Analysis and Probability might be better conceptualized as two separate standards by NCTM (2000) rather than being merged into one, since the probability part of the standard suggests that students learn formal probabilistic rules.

While recommendations surrounding the study of formal inference and probability seem to vary, the recommendation that all students learn to use technology appropriately in the process of data analysis seems to be a constant in all recent curricular

recommendations. Biehler (1997) discussed features that should be present in statistical software packages for data analysis at the secondary level. He stated that computer software packages need to strike a balance between ease-of-use and ease-of-learning. He recommended that software allow students to become familiar with methods and concepts of statistics by allowing for exploratory data analysis and not just applications of formal statistical methods. Software should also allow students to conduct Monte Carlo simulations to simulate simple and multistage experiments. Software also needs to be flexible, allowing teachers the ability to redesign it to meet specific curricular goals. Biehler considered these types of data analysis tools as indispensable to the teaching of statistics at the secondary level.

All of the literature discussed in this section was taken into account in the design of statistical thinking tasks. Interview tasks allowing students to demonstrate knowledge of the content included in the “Data Analysis” portion of the “Data Analysis and Probability” standard (NCTM, 2000) were included. In addition, the interview protocol included data analysis tasks whose context allowed for the application of formal inferential methods. In accord with Cobb and Moore’s (1997) views about the role of formal probability within an introductory statistics course, no questions designed to assess students’ knowledge of formal probability were included in the interview protocol. Students’ informal notions about probability, however, were examined in questions dealing with sampling distributions, and in questions that can be answered via simulation. Finally, the interview protocol allowed students to make use of statistical software packages and graphing calculators in responding to tasks. While not all students who

were interviewed were proficient in using one or more software programs for data analysis, it was seen as important to allow students who were proficient in their use to have access to them so they would solve interview tasks in a manner which was natural for them.

The wording of the tasks, the order in which they were presented, and the context in which they were set were advised by the collection of empirical data (Groth, 2002a). In the first phase of pilot-testing, a set of tasks was devised and administered to four different students representing a range of mathematical ability. As a result of this phase, unwanted ambiguities in wording, contexts confusing for students, and tasks not serving to reveal a wide range of statistical thinking were eliminated. Tasks that did serve to reveal a wide range of statistical thinking were retained. In areas where the interview script did not give adequate information about a statistical thinking process, tasks were added. In the second phase of pilot-testing, the entire interview script generated as a result of the first phase was completed by an AP Statistics class in the form of written responses. As a result of the second phase, slight revisions in terms of wording, contexts, content of tasks, and order of presentation of tasks were made. The end result of the pilot-testing process was the interview script in Appendix B. Tables showing the process each task was designed to assess and how its content was aligned with current curricular recommendations for high school statistics are presented in appendices C, D, and E.

Procedure

Each of the interview participants was informed that it would take a total of approximately 2-3 hours to answer all of the interview questions. Six of the participants

in the first two categories decided to split the total interview time between two interview sessions. Julie was the only student in the first two categories who decided to do the entire interview in one session. Most of the participants in the last category were interviewed during study hall periods, so the total interview time for these students was generally split among three different 50-minute study hall periods. Only one of the students in the last category, Daniel, decided to do the entire interview in one sitting by scheduling an interview session after school hours.

Before each interview session, each student was asked to complete a brief questionnaire (Appendix F). The first item on the questionnaire was administered in order to attempt to determine which types of technology the students were able to use when solving statistical problems. At least one piece of technology the students indicated familiarity with was made available during the clinical interview sessions, and students were allowed to use the technology to whatever extent they wished during the interview sessions. The software programs Excel and Fathom were made available during every clinical interview session, as was a TI-83 calculator. The purpose of the remaining items on the questionnaire was to have students share some anecdotal information about their previous experiences with statistics. The information they shared on these items is reported in appendix A.

During the clinical interview sessions, the interview protocol (Appendix B) was administered in the same order to each of the students. Each student started by completing task 1 and then proceeded, in the order shown, through the entire interview protocol until completing all parts of task 5. As students were interviewed, I collected

data by taking field notes, audio recording responses, and keeping any written work they completed. The audio recordings were later transcribed for analysis.

Data Analysis

Analysis of interview transcripts was advised by the constant comparative method described by Maykut and Morehouse (1994). Analysis of the interview responses took place concurrently with data collection. I grouped responses to the interview sub-tasks into five categories according to the statistical thinking processes (describing, organizing and reducing, representing, analyzing, and collecting data) to which they pertained. Some sub-tasks failed to give useful information about thinking within any of the five processes, and the reasons why they failed to do so are reported in detail in the next chapter.

After the responses had been placed into five groups according to statistical thinking processes, sub-tasks that elicited responses about similar statistical content were clustered together. Each of the clusters provided information about students' thinking for a statistical thinking subprocess. For example, within the broad process of organizing and reducing data, four different clusters of interview items provided information about students thinking within the four subprocesses of using measures of center, using measures of spread, recognizing the effect of data transformation upon center and spread, and organizing raw sets of data.

Within each statistical thinking subprocess identified, I expected to see evidence of different modes of thinking as well as levels within each of the modes (Biggs & Collis, 1982, 1991). Hence, I examined interview responses within each of the subprocesses and

grouped them according to their relative complexity. I then wrote descriptors to capture the essence of each different pattern of response identified. The descriptors that were written are reported in the next chapter.

The end result of the constant comparative method of data analysis was the generation of sets of descriptors for different patterns of thinking within each of the five processes of describing, organizing and reducing, representing, analyzing, and collecting data. I examined each set of descriptors for agreement with the levels of thinking described in the Biggs and Collis (1982) cognitive framework. I then matched each descriptor with both a mode of development (concrete-symbolic or formal) and a level within the mode (unistructural, multistructural, or relational).

Some of the aspects of a check-coding procedure described by Miles and Huberman (1994) were used to finish the process of data analysis. First, a random sample of six students was drawn. Then, two other researchers familiar with the Biggs and Collis cognitive framework analyzed responses given by each of the six students drawn. Each of the researchers categorized the student responses according to the descriptors I had formulated during my own data analysis. I then met with the two researchers to discuss the conclusions they had come to during data analysis. In cases where disagreement occurred about categorizations of students by descriptors or the evidence supporting the formulation of sets of descriptors, the disagreements were discussed until consensus was reached. In cases where disagreements occurred, the originally written descriptors were revised or student responses were re-categorized. The next chapter reports the end result of the data analysis process.

CHAPTER IV

RESULTS

The interview protocol that served as the primary data gathering instrument for the study (appendix B) was designed to elicit various degrees of sophistication in patterns of response to questions involving describing, organizing and reducing, representing, analyzing, and collecting data. In the first main section of this chapter, I report upon the patterns that were identified and discuss the relative sophistication of each. In this section, the interview tasks that led to the identification of the patterns of response are discussed along with illustrative student responses. In the second main section, I discuss the connections between the patterns identified and the Biggs and Collis (1982, 1991) cognitive model.

Patterns of Response to Statistical Thinking Tasks

Patterns of response of various degrees of sophistication were evident within each of the five statistical thinking processes investigated. The analysis of interview data led to the identification of three such patterns within the process of describing data and four such patterns within representing data. Within each of the processes of organizing and reducing, analyzing, and collecting data, there were several subprocesses. Each subprocess has its own set of patterns of response. For example, the subprocess of “using measures of center” lies within the broad process of organizing and reducing data, and four distinct patterns of response were identified for the subprocess. The processes of

analyzing and collecting data are similarly split into subprocesses. Each of the patterns identified within each statistical thinking process and subprocess are discussed in detail in this section.

I have adopted several writing conventions in relating my findings. First, I have summarized the patterns of response from least to most sophisticated in each of the following sections by using figures. The description of the least sophisticated pattern of response appears at the bottom of each figure, the intermediate patterns above it, and the most sophisticated patterns on the very top. Second, a detailed discussion of each of the patterns summarized in the figures appears immediately after each figure. In each case, the detailed discussion progresses from the least sophisticated pattern of response that was evident to the most sophisticated pattern evident. Third, I have used the phrases “patterns of response” and “patterns of thinking” interchangeably, recognizing that some may argue that students’ responses are not always completely indicative of their thought processes. Finally, in discussing interview protocol tasks, I have used a decimal system that places the task in front of the decimal point and the subtask immediately after (e.g., task 3, part 2 is abbreviated as “3.2”).

Describing Data

As defined in previous chapters, describing data involves reading data displayed in both graphs and tables. Several interview items (Appendix B) were aimed at assessing students’ thinking in both areas. However, patterns of response of various degrees of statistical sophistication were observed only in regard to some of the questions pertaining

to graph reading (1.10, 2.5, 4.1). Hence, the patterns of response to be discussed in this section pertain only to students' abilities to read statistical graphs.

Three different graph reading tasks led to the identification of patterns of thinking within the process of describing data. In one such task (1.10), students were asked to explain the meanings of points in several different statistical graphs that were produced from a set of data an individual had collected over the course of a shopping mall survey consisting of several different questions. Both one variable (dotplot and boxplot) and two variable (scatterplot) graphs had been produced from the data. In task 2.5, students were asked to find the highest value in a one variable display of the distribution of test scores for a population of students. Task 4.1 involved describing the meaning of a point displayed in a scatterplot showing the relationship between ice cream consumption and temperature.

Some of the interview questions designed to identify various degrees of sophistication in patterns of response in regard to describing data failed to do so. One question pertaining to graph reading (1.11) was considered unhelpful in discerning patterns of response because students who had more familiarity with the technology used to produce the graphs in the question had an advantage over those who did not. Those who were familiar with the conventions of TI graphing calculators were able to interpret the "max" and "min" labels included at various points in the question, while the others could at best make reasonable conjectures about them. The interview questions asking students to read tables (2.3, 5.2) did not elicit different patterns of response. All students interviewed exhibited no difficulty in reading values from tables. Each student simply

correctly read the required information from each table and gave no further explanation.

The patterns of response discernible in connection with describing data derived from tasks that actually were successful in eliciting thinking about graph reading are summarized in figure 5.

Pattern descriptor	Students whose responses reflected the pattern
Relates information about the appropriate number of variables when discussing meanings of points on both one and two variable statistical graphs. Correctly relates information about display conventions of a variety of graphs.	Crystal, Jeff, Julie, Brooke
Relates information about the appropriate number of variables when discussing meanings of points on both one and two variable statistical graphs.	Bill, Luke, Kristen, Nancy, Jessica, Paul, Daniel, Laura, Rick
Relates information about only one variable when discussing meanings of points on both one and two variable statistical graphs.	Lisa, Hillary

Figure 5. Describing statistical graphs

In the least sophisticated pattern of response identified, students related information about only one variable when discussing meanings of points on both one and two variable statistical graphs. Lisa and Hillary both realized that the dotplot and the boxplot displaying data about responses to mall survey questions (task 1.10) contained information about one of the variables in the data set. They also both demonstrated the ability to read the one variable graph displaying test scores presented in task 2.5. Neither of them, however, completely described the information contained within the points of the scatterplots presented. For example, when Hillary was asked to explain what the point

farthest to the left in the scatterplot in task 1.10 (comparing hours of sleep per week to age) represented, she replied, “That would be like the minimum, like the minimum hours of sleep or the minimum age. Oh, wait, on this one it would just be the minimum hours of sleep, not age.” When asked to explain the meaning of the point farthest to the left in the scatterplot presented in task 4.1 (comparing ice cream consumption to temperature), she replied, “That’s the lowest temperature, the day with the lowest temperature.” In each case, her descriptions mentioned only one of the relevant variables.

In the next pattern of response identified, students demonstrated the ability to describe the meanings of data points contained in both one and two variable statistical graphs. Kristen’s responses to task 1.10a (scatterplot showing age vs. hours of sleep) and 1.10d (dotplot showing ages of people in the sample) are illustrative. She successfully described the meaning of the point farthest to the left on the scatterplot, stating, “Well, it’s kind of hard to get specific with the age, but you would say that someone who is under 25, probably 16, because that’s the lowest age on here, gets 20 hours of sleep a week.” She also successfully described the meaning of the point farthest to the left in the dotplot, stating, “That is just showing one person who is 16.”

The most sophisticated pattern of response identified within the subprocess was that at which students successfully read both one and two variable statistical graphs and also fully described the meanings of display conventions within each type of graph presented. Crystal, for example, in response to task 1.10 stated that the leftmost point on the scatterplot (age vs. hours of sleep) showed “a 16 year old who gets 20 hours of sleep.” She showed the ability to read one-variable displays and their conventions as well,

correctly identifying, for example, the bar near the center of the boxplot as “the median amount of sleep.” In this pattern of response, complete descriptions of meanings of points on both one and two variable displays were coordinated with complete descriptions of display conventions.

In summary, three different patterns of response were identified in regard to reading statistical graphs. Students giving the least sophisticated responses left important components out of their descriptions of statistical graphs. Two variable displays were described as if they were one variable displays, and some display conventions were misidentified. Responses in the next pattern differentiated between one and two variable displays, but some display conventions were still misidentified. The most sophisticated pattern of response was that at which the differentiation between one and two variable displays was coordinated with the correct reading of the display conventions of the graphs.

Organizing and Reducing Data

The statistical thinking process of organizing and reducing data involves arranging, categorizing, or consolidating a given set of data into summary form. In order to investigate students’ thinking about this process, questions were included in the interview script to attempt to uncover students’ thinking in regard to four subprocesses: using measures of center (1.4; 2.6, 5.8b), using measures of spread (2.2; 3.4; 5.8a), recognizing the effect of data transformation upon center and spread (2.10), and organizing raw sets of data (1.3). Descriptors characterizing patterns of thinking for each of the four subprocesses are discussed in this section.

Subprocess 1: Using Measures of Center

As discussed in chapter 2, students' use of measures of center has been investigated extensively by researchers. Some researchers have chosen to focus only upon students' understanding of the arithmetic mean (e.g, Pollatsek, Lima, & Well, 1981; Strauss & Bichler, 1988; Cai, 2000). Others have focused upon students' use of measures of center in general (e.g., Mokros & Russell, 1995; Zawojewski & Shaughnessy, 2000; Watson & Moritz, 1999a). My study of students' understanding of measures of center had the latter focus, since I was interested in investigating students' choices of measures of center in various situations.

Three different interview tasks elicited students' thinking about measures of center. In one task (1.4a), students were asked to find the typical income in a skewed data set. In another (2.6), students were to find the typical test score within a slightly skewed data set presented graphically. Finally, in a third task (5.8b), students were asked to find the true weight of a fish given a skewed set of measurements. The strongly skewed nature of the data in some of the data sets made the choice of a measure of center a non-trivial matter. The patterns of thinking identified for the subprocess and the students whose responses reflect them are summarized in figure 6.

One task (1.4b) that was designed to elicit different patterns of thinking about using measures of center failed to do so. In the task, I asked students to find typical political party affiliation for people in a data set. Students all answered the task in a similar manner, finding the typical political party affiliation by simply comparing the numbers of

people belonging to each different party. Different patterns of thinking relating to the use of measures of center were not evident in responses to the question.

Pattern descriptor	Students whose responses reflect the pattern
Uses reasonable formal measures to locate centers of data sets.	Hillary, Paul
Uses a combination of reasonable formal and visual measures to locate centers of data sets.	Bill, Nancy, Crystal
Uses a combination of formal and visual measures to find centers of data sets, only some of which are reasonable for the given set of data.	Lisa, Kristen, Jeff, Julie, Jessica, Rick, Daniel, Brooke
Uses only visual approaches to find centers of data sets, only some of which are reasonable for the given set of data.	Laura, Luke

Figure 6. Using measures of center

As noted in figure 6, one of the important criteria distinguishing among patterns of sophistication in thinking for this subprocess was the use of formal or visual measures reasonable for the given sets of data. The data set of salaries for people in a sample (task 1.4a) contained a value of \$2,000,000. This value was an outlier, since it was more than 1.5 times the interquartile range away from the third quartile of the data set. Likewise, the data set of measurements for the weight of a fish (task 5.8b) contained an outlier. The observation of 10.1 pounds for the data set was an outlier on the low side. Hence, I did not consider the mean to be a reasonable measure of center for either task (1.4a or 5.8b), since it would be unduly influenced by the outliers. I did consider the median to be a reasonable measure of center in each case. I also considered the use of the mean to be

reasonable if the extreme points in the data set were discarded. Visual measures that produced results close to those produced by those two formal approaches were considered reasonable. In the test scores task (2.6), although the data set was slightly skewed right, there were no outliers. Hence, the use of the mean, median, or mode was considered reasonable for describing the typical value in the data set.

Students giving responses within the least complex pattern used only visual approaches to find the centers of data sets. In some cases, the visual approach used was not reasonable because it did not account well for skewness. Luke's responses are illustrative. When asked to find the typical income of the data set in task 1.4a, he replied, "I would say about \$50,000 to \$55,000. Because the majority of the data is around 50 to 60, and there's some a little bit higher, and some a little bit lower. Then there's like some outliers of 2 million and 16,000, and 14,000. But the majority of the data is between \$50,000 and \$55,000." In response to task 2.6, he stated the typical test score would be "between 6 and 9" because "from about 6 to 10, there's a high concentration of kids." Finally, when asked to find the true weight of the fish in task 5.8b, he stated, "About 22 [pounds]. It looks like about the average. I ignored 10.1 [pounds], because like I said earlier, some outliers you just have to ignore, and realize there was something wrong with the data or something. So, I took the average of the other 6 [without computing it], and it's about 21.5, or 22 [pounds], somewhere around there." Hence, Luke used only a visual strategy in each context. No use is made of the formal measure of median to describe the centers of the skewed sets. In the first case, his visual estimate of \$50,000 to \$55,000 is fairly far away from the actual median income of \$42,500.

In the next pattern of thinking within the subprocess, students were able to use two different relevant strategies to locate the centers of data sets. The use of both formal and visual approaches was evident in responses reflecting this pattern. In using the two different approaches, however, students sometimes still produced unreasonable answers by not accounting well for skewness. For example, when asked to find the typical income in task 1.4a, Kristen stated, “Oh, in my opinion? Oh, I have no idea. I never know, like...I would say 24,000 [dollars], because there are lots of 20 thousands, but then there are some really high ones, and then there’s lower ones...somewhere around 24 [thousand dollars].” Her visual approach, in this case, had produced an estimate fairly far from the median income. In response to the fish weighing task (5.8b), Kristen stated, “if you just added all those weights together and then divided by 7, you got 19.7 [pounds] as the average. If you just looked in the middle, you got 21.5. So, then I just add 21.5 to 19.7 and divided the two, and got 20.6. So, I just did the average of both averages.” This formal strategy, although somewhat unconventional, did take the skewness of the data set into account. In general, responses reflecting this pattern used some formal measures in describing centers of data sets, but the formal and visual approaches did not always produce results reasonable for the given data sets.

Student responses reflecting the next pattern incorporated both formal and visual methods in order to produce reasonable values for the centers of the given data sets. Bill, a student in this category, used a reasonable visual strategy in order to answer the typical income task (1.4a), and he used a reasonable formal strategy to answer the fish weighing task (5.8b). In response to the former task, he stated, “About 45 [thousand dollars],

maybe? No, wait a minute. Maybe about 40. First I see that I have more incomes from the younger generation. And even though none of theirs reach 40...they're all about 26, 28, I used those to outbalance the fact that you've got one over here at 50, and 86, an 110,000, stuff like that. I tried to balance them out. And then your median category, the 30's, they're all making around 40 or 45." In response to the fish weighing task, he used a reasonable formal strategy: "21.316 [pounds]. I took the average of the greatest 6 (using calculator), and just threw out the 10.1, because I don't think that's right, anyway." When presented with the test score task (2.6), he opted for a visual approach in which he equated mode with typical value. In response to the task, he chose 6 as the typical value, stating "it had the highest mark on the graph, leading me to assume that the most people received a 6." Responses reflecting this pattern in general were characterized by a fit between the strategy used for finding center and the context of the data at hand.

Some students interviewed used reasonable formal approaches to find the centers of each of the data sets. Hillary's responses help to illustrate. Her response to the typical income task (1.4a) was, "If you just average them, you get \$140,000 a year for the average, but you might have the one that earned 2 million a year and the one that earned 5,000 a year might skew it...might mess up the data and make it look larger than it really is. You might want to throw those two out, and then figure out what the average is for the rest of those. If you did that, (using calculator) then the average would be 44,000 a year." She used a similar strategy for the fish weighing task (5.8b): "21.3 [pounds]. I just discarded the 10 because it was so extreme, and I took the average (using calculator) of the 6 remaining measurements." When asked to find the typical test score for students in

Arkansas (task 2.6), she correctly identified the mode as 6. She further mentioned that the typical score would likely lean toward the median because the data set was somewhat skewed. Hence, responses in the most sophisticated pattern within the subprocess were characterized by the inclusion of reasonable formal strategies for finding the centers of data sets in each context.

Overall, the responses to tasks within the subprocess of using measures of center revealed an array of patterns of thinking. A progression from the use of only visual measures to the use of only appropriate formal measures was apparent. Intermediate patterns of response involved the gradual incorporation of formal measures. The intermediate patterns also revealed differing degrees of fit between the measures of center chosen and the contexts at hand. The most sophisticated pattern of response was characterized by the use of reasonable formal measures to describe the centers of each of the data sets.

Subprocess 2: Using Measures of Spread

Three different questions in the interview protocol were aimed at uncovering students' understanding of measures of spread (2.2, 3.4, 5.8a). Each of the three tasks helped in the construction of descriptors for the subprocess. In task 2.2, valuable information about the measures of spread students had at their disposal was obtained as they compared the spreads of test scores for three different classes. In task 3.4, similar information was obtained as students were asked to compare the spreads of salaries for department store workers listed in two different columns in a chart. Responses to task 5.8a helped to round out the picture of students' thinking within the subprocess, as

students were asked how spread out they considered a set of weight measurements to be. The skewness of the set of measurements in this subtask made the choice of measure of spread more difficult, since simple use of the range would not capture the tightly clustered nature of most of the data in the set. Three different patterns of response were discerned as interview responses to the three tasks were analyzed, and they are summarized in figure 7.

Pattern descriptor	Students whose responses reflected the pattern
Gives quantifications and subjective verbal descriptions of spread that are suitable for given sets of data.	Luke, Crystal, Paul, Laura, Nancy
Gives quantifications and subjective verbal descriptions of spread. Some descriptions or quantifications are not suitable for given sets of data.	Kristen, Bill, Jeff, Lisa, Julie, Jessica, Hillary, Rick
Gives subjective verbal descriptions of spread rather than quantifications.	Brooke, Daniel

Figure 7. Using measures of spread

As noted in the pattern descriptors in figure 7, one of the criteria distinguishing among patterns of sophistication in thinking for this subprocess is the ability to incorporate descriptions and quantifications of spread suitable to the data set at hand. In the test scores task (2.2) and the department store workers salary task (3.4), I considered the range to be a suitable measure of spread, and any verbal description that captured the idea of the range to be suitable. In each of the two tasks, there were no outliers (i.e., no points farther than 1.5 times the interquartile range away from the first or third quartiles)

present in the data set that would cause the range to be an unreasonable measure of center. In the fish weighing task (5.8a), the data point showing 10.1 pounds was an outlier. Hence, I did not consider the range or descriptions capturing the idea of the range to be suitable, because they neglect to describe the tightly clustered nature of most of the data set. Standard deviation also was not considered a reasonable measure of spread for the data set, since it would be unduly influenced by the outlier in the data set. Formal measures of spread such as interquartile range or else descriptions that noticed the cluster in the data set were considered suitable.

The least sophisticated approach to the tasks within the subprocess was demonstrated by students who gave verbal descriptions of the spreads of data sets rather than quantitative measures of the spreads of the data sets. Brooke's responses reflected this pattern. When asked to describe the spread of the test scores in task 2.2, she said, "I would say for class B it's probably the most spread out. Because it's not missing, but there's a bigger jump between 15 and 20." Her response does not quantify the spread of class B, but rather discusses "jumps" between discrete data points. When asked to describe the spread of the first set of salaries in task 3.4, Brooke replied, "They're really not that spread out. If you would get rid of the extremes like 5.50, I think you would pretty much see that it's not incredibly spread out." Again, she focused upon describing parts of the set rather than quantifying the spread of the set. Finally, when asked about the spread of the measurements of the weight of a fish in task 5.8a, she said, "Well, they're not that spread out...most of them don't...I mean, none of them are the same. But I know that happens when you get on the scale sometimes. But, they're pretty similar except for

there is one, and that is the 10.1. Maybe he only weighed half the fish.” While the verbal descriptions reflecting this pattern capture some of the important characteristics of the spreads of the data sets, no evidence was present that students were able to assign a numerical value to spread.

Responses in the next pattern identified indicated that students were able to give both verbal descriptions of spreads of data sets and quantifications of spread. However, their descriptions and quantifications did not always take important characteristics of the data sets into account. Kristen’s responses are illustrative. She verbally described spread and also quantified spread in the test score task (2.2). She stated, “Class C. Well, it’s kind of between B and C. Like Class A is all about in the same range, it only differed by like 6 points or whatever. But, for C I just have...they go from 7 all the way to 20...there’s just a few skips in between...it’s pretty spread out.” Her response to the department store worker income task (3.4) showed her ability to quantify the overall spread by looking at the range. She said, “Pretty spread out. There’s a 13 dollar difference there, no, 14, no even more than that. From 5.50 to 20 is pretty spread out.” Her response to the fish weighing task (5.8a) gives an example of a situation in which a quantification for spread was given that did not take the skewness of the data set into account. She seemingly did not notice the tightly clustered nature of the points at the high end of the data set, stating simply, “About 13 pounds...this is the same fish? That’s weird. It’s pretty spread out. How one fish can be 10 pounds and the other end 23...that’s something my dad would do.” The verbal descriptions and quantifications of spread reflecting this pattern of response were not always well-suited to the data at hand.

In the most sophisticated pattern of response identified, students were able to adjust their quantifications and verbal descriptions of spread to capture the important characteristics of the spreads of the data sets they were asked to examine. Luke's responses help to illustrate this pattern of thinking. He quantified the spread in the test scores task (2.2). The following interaction took place in response to the task:

Luke: C. It has the largest range.

I: Largest range meaning...

Luke: Lowest to highest test score.

Luke similarly used the range in order to quantify the spread of department store worker salaries (task 3.4). In answering the fish weighing task (5.8a), Luke took the salient characteristics of the data set into account in describing the spread. He stated, "Well, they're not very spread out, except for the 10.1 [pounds]. And the other ones look fairly close, they're about 21 or 22. And the 10.1, I don't know what happened, if there's something wrong with the scale or something. But, except for that point, the one outlier, they're not very spread out." Since students giving responses reflecting this pattern demonstrated the ability to adapt verbal descriptions and quantifications to include descriptions of clusters within a data set when applicable, their responses were considered to be more sophisticated than those reflecting the previous pattern.

The three patterns of response identified within the subprocess of using measures of spread bear some resemblance to the first three patterns identified within the subprocess of using measures of center. The least sophisticated pattern of response evident within each subprocess was that at which students made use of subjective visual strategies for

reducing data sets. At a more sophisticated pattern of response within each subprocess students made use of some formal objective strategies along with subjective strategies for reducing data, but those strategies were not always well-suited to the data at hand. At the third pattern of response within each subprocess, the subjective and formal objective strategies that were incorporated for data reduction were each well-suited to the data. This similarity in structure suggests that patterns related to thinking about measures of spread somewhat resemble those related to measures of center. Even though this is the case, it is possible for students to give more sophisticated responses to tasks within one of the two subprocesses.

Subprocess 3: Recognizing the Effects of Data Transformation upon Center and Spread

NCTM (2000) recommended that high school students should “recognize how linear transformations of univariate data affect shape, center, and spread” (p. 324). Because of this recommendation, one task (2.10) on the interview protocol was designed to investigate students’ patterns of thinking in a situation where each value of a univariate data set was increased by a constant of 15. The situation in the task was that a teacher increased the test scores for each student in a class by 15 points. The students interviewed were asked to describe the impact that this type of linear transformation would have upon the center and spread of the data set. Three different patterns were evident as responses to this item were analyzed. They are summarized in figure 8.

Pattern descriptor	Students whose responses reflected the pattern
Recognizes the effect of data transformation on both measures with a complete quantitative description of the impact on one measure of center.	Laura, Luke, Bill, Jeff
Recognizes the effect of data transformation on both measures without completely quantifying the effect upon center.	Kristen, Hillary, Lisa, Crystal, Daniel, Julie, Jessica, Rick
Recognition of the effect of data transformation on only one of two measures: center or spread.	Paul, Nancy, Brooke

Figure 8. Recognizing the effects of data transformation upon center and spread

The least sophisticated pattern of response within the subprocess was that at which students recognized the effect of data transformation on only one of the two measures: either the center or the spread. Brooke, for example, recognized that adding 15 points to each test score would cause the typical score to increase, saying, “Well, on the graph it would show that it would increase. But you still would have the same shape, in terms of it would be going down, however the scores would be higher.” She seemed to have no recognition of the non-impact on spread, stating, “I think the spread would probably be more evident if you had more numbers in between. Because the way you have it now, there’s just one number difference in all of the scores. But if you add 15 on, then there would be a noticeable difference.” Her response indicates that she felt adding 15 on to each score would actually increase the distances between the points in the data set. In general, responses in this category incorporated only one relevant aspect of the task, recognizing either that the typical score would increase or that the spread would remain unchanged.

In the next pattern of response identified for this subprocess, students were able to grasp two relevant aspects of the task, recognizing the impact of data transformation upon both center and spread. However, they were not able to fully quantify the change regarding the center. A typical response reflecting this pattern was given by Crystal. In response to the question of how the proposed data transformation would impact the typical score, she replied, “Adds 15 points to every person’s score? That would drastically raise the average score. And it wouldn’t make sense, because it’s only a 20 point test, and then people would get over 20, so it wouldn’t really make sense. It would drastically raise the average score.” She indicated recognition of the increase of the typical score, but did not quantify the increase. She simply stated that it would increase “drastically.” In response to the question about spread, Crystal replied, “The spread would actually stay the same, because every person got 15, so the spread would stay the same if you raised the numbers...just because...it would all stay the same.” While responses reflecting this pattern correctly described the impact of the data transformation upon both center and spread, they did not fully quantify the impact upon center.

At the most sophisticated pattern of response identified, the impact upon center and spread was correctly described and quantified. Jeff’s response to the task reflected this pattern. Jeff recognized that in the task situation, “The mean would be raised by 15 points.” In terms of spread, he recognized that, “The lowest point would be raised by 15, as well as the highest point. The range would still be the same.” Hence, Jeff and the other students whose responses reflected this pattern fully quantified the impact on both center and spread in terms that were suitable for the given context.

The three different patterns discernible for the subprocess show a gradually increasing awareness of the impact of data transformation upon center and spread. In the least sophisticated pattern, students seem to grasp the impact upon only one of the two measures. In the next pattern, an understanding of the impact upon both measures is demonstrated, but is not fully quantified. Students whose responses reflected the third and most sophisticated pattern identified gave quantitative descriptions not included in responses reflecting the previous two patterns.

Subprocess 4: Organizing Raw Sets of Data

The research community is just beginning to gather information describing how students organize raw sets of data (e.g., Mooney, Langrall, Hofbauer, & Johnson, 2001). In order to gather this type of information about high school students, in task 1.3 I asked students to organize a data set containing answers people had given to questions administered during a shopping mall survey. The data set contained both measurement variables (age, annual income, and hours of sleep per week) and categorical variables (favorite color, political party affiliation, and a “yes” or “no” response to a question about computer literacy) pertaining to the sample of people that had been interviewed. Both types of variables were included in the data set because of the NCTM (2000) recommendation that high school students be able to “understand the meaning of measurement data and categorical data” (p. 324). As students organized the given set of data, three different patterns of response were apparent. They are summarized in figure 9.

Pattern descriptor	Students whose responses reflected the pattern
Uses both relevant strategies for grouping demonstrated in the previous pattern: (i) Forming groups within an ordering of values of a measurement variable; (ii) Forming groups by using a pre-existing categorical variable within the data set.	Daniel, Hillary, Paul, Luke, Jeff, Laura
Uses the strategy of ordering the values of a measurement variable to organize data. One of the following two grouping strategies is also incorporated: (i) Forming groups of values within the ordering; (ii) Forming groups by using a pre-existing categorical variable within the data set.	Bill, Rick, Nancy, Crystal, Kristen, Julie, Jessica
Uses the strategy of ordering the values of a measurement variable to organize data.	Brooke, Lisa

Figure 9. Organizing raw sets of data

The least sophisticated pattern of response apparent was that at which students used only the strategy of ordering the values of a measurement variable in order to organize the data set. Within this pattern, no grouping strategies were evident. Lisa, one of the students whose response reflected this pattern, chose to organize the data set from oldest to youngest. Brooke, the other student whose response reflected the pattern, organized the data set by putting the data set in order from greatest to least according to the measurement variable of annual income. No clusters or groups were formed by either student in organizing the data.

At the next discernible pattern of response, students still used the strategy of organizing the data by ordering the values of measurement variables from least to greatest. However, this was not the only organization strategy used. Responses reflecting

this pattern also incorporated one grouping strategy in organizing the data. Some used the strategy of creating clusters and groups within the values of a measurement variable, while others used pre-existing categorical variables within the data set to form groups. Bill, for example, organized the data from least to greatest according to age. Then, within that organization, he formed groups according to decades (i.e., 20's, 30's, 40's, etc.). Julie also organized from least to greatest according to age. However, she did not form groups within that ordering, but used the strategy of separating the data into groups by political parties. Responses reflecting this pattern were more complex than those the previous pattern because one strategy for grouping was incorporated with a strategy for ordering according to values of a measurement variable.

The most sophisticated pattern of response for the subprocess was that at which students organized data by using both of the grouping strategies used by students whose responses reflected the previous two patterns. Each student whose response reflected this pattern used the strategy of forming groups within values of a measurement variable as well as the strategy of forming groups by using a pre-existing categorical variable. Daniel, for example, sorted the cards by the pre-existing categorical variable of favorite color. He did the same with the pre-existing categorical variable of political party affiliation. He incorporated another type of grouping strategy as well, mentioning the possibility of forming groups of \$10,000 each within the measurement variable of income. Hillary, another student whose response reflected this pattern, formed clusters within an ordering of the measurement variable of age. She also discussed grouping by using the pre-existing categorical variable of political party affiliation. Each of the

responses reflecting this pattern at some point incorporated two relevant strategies for forming groups within the set of data.

Hence, students' responses within the subprocess of organizing raw sets of data had a total of three discernible patterns. The least sophisticated pattern was that at which students used a simple ordering of values of measurement variables to organize data. At the next pattern discernible, a grouping strategy was used in addition to an ordering strategy in order to organize the data set. The most complex pattern of response for the subprocess was that at which students incorporated two different relevant grouping strategies in organizing the data.

Representing Data

Representing data is the process of displaying a given set of data by using graphs. Some researchers investigating students' thinking in this area have chosen to focus on subskills related to producing graphs, such as scaling and labeling axes (e.g., Padilla, McKenzie, & Shaw, 1986). Others have focused on students' abilities to produce certain types of graphs, such as line graphs (e.g., Berg & Phillips, 1994; Zawojewski & Shaugnessy, 2000). Still others have taken a different approach, investigating the types of graphs students produce when asked to effectively represent given sets of data (e.g., Mevarech & Kramarsky, 1997; Mooney, 2002). My study of students' thinking while representing data falls into this last category, since I presented students with situations and data sets and asked them what types of graphs would be effective in each case.

Responses to several different tasks were incorporated in forming a set of descriptors for the patterns of response in regard to representing data. Tasks 1.5, 1.6, and 1.7 each

required students to produce graphical representations for bivariate situations. Since NCTM (2000) recommended that high school students “display and discuss bivariate data where at least one variable is categorical” (p. 324), tasks 1.5 and 1.6 each involved at least one categorical variable. Task 1.5 dealt with one categorical variable (favorite color) and one measurement variable (income). Task 1.6 involved two categorical variables (computer literacy and political party affiliation). Task 1.7 was also a bivariate situation, but it involved two measurement variables (age and income). In tasks 4.8 and 4.9, students were asked to conceptualize and discuss displays for a set of univariate measurement data. In each of the task situations, students were prompted to give multiple representations. Four different patterns were evident in students’ responses to these data representation tasks. The four patterns are summarized in figure 10.

Pattern descriptor	Students whose responses reflected the pattern
Conceptualizes multiple valid representations in both univariate and bivariate situations.	Bill, Rick, Crystal
Conceptualizes graphs in both univariate and bivariate situations, incorporating multiple representations in some bivariate situations.	Lisa, Hillary, Daniel, Paul
Conceptualizes multiple graphs for bivariate situations <u>or</u> conceptualizes one valid graph for each univariate and bivariate situation.	Jeff, Luke, Jessica, Julie, Kristen, Laura, Brooke
Conceptualizes one graph for bivariate situations.	Nancy

Figure 10. Representing data

Some of the tasks designed to elicit students’ thinking about data representation were not used in the construction of the pattern descriptors in figure 10. Task 3.2 was designed

to determine patterns of response for representing data in a context where values for men's and women's salaries were displayed. Some students responded to the task as if it were a bivariate situation, while others conceptualized the situation as the comparison of two univariate measurement variables. While this difference in perception itself is interesting, it did not help in the formation of descriptors for different patterns of sophistication in response, since both conceptualizations of the situation are equally valid. Also, students' different abilities to read some of the graphs in tasks 1.8 and 1.9 made it virtually impossible to compare the components of their responses which dealt with data representation. A further complication is that students who had greater familiarity with the graphical conventions of TI calculators were able to interpret the boxplots and histograms in 1.8 and 1.9 more effectively than those who were not familiar with them (this is the same difficulty that arose in task 1.11). The patterns of thinking summarized in figure 10 and described in detail below were not discerned in any manner in the responses to these problematic tasks.

The least sophisticated pattern of response was displayed by Nancy. She was able to conceptualize one graph for each of the bivariate situations presented. She used a line graph to display information about favorite color and income (1.5), pie graphs to break down computer literacy by political party affiliation (1.6), and a scatterplot to compare age and hours of sleep per week (1.7). She was not able to conceptualize a graphical representation to display information about the variable of ice cream consumption (task 4.8). When asked if she could do so, she replied, "I don't think so, because people are going to wonder where you got these numbers [in the chart showing the data] from." Her

response indicated that she did not see how it would be possible to construct a graphical representation to display information about just one variable.

Several responses were more structurally complex than Nancy's. At the next pattern of response identified, students showed the ability to do more than simply conceptualize one graphical representation for each bivariate situation. Luke, Jessica, and Julie gave responses similar to Nancy's in that they were not able to conceptualize a graphical representation for the univariate measurement data situation. However, the responses of Luke, Jessica, and Julie were more structurally complex in the sense that each of these students demonstrated the ability to conceptualize multiple representations for some of the bivariate situations (e.g., Luke conceptualized two graphs for favorite color vs. income and also for computer literacy vs. political party affiliation). Jeff, Kristen, Laura, and Brooke gave responses similar to Nancy's in the sense that they conceptualized only one graphical representation for each bivariate situation. However, the responses of this group of students were more structurally complex in the sense that each of them included the conceptualization of one representation for the univariate measurement data situation (e.g., Jeff discussed using a histogram to represent the ice cream consumption data). Therefore, responses classified within this pattern incorporated one more relevant aspect in their responses than those in the previous pattern. In summary, one group of responses reflecting this pattern incorporated the additional aspect of the production of multiple representations for some bivariate data, while another group incorporated the additional aspect of producing a graphical representation for a univariate measurement data set.

Each of the responses reflecting the next pattern incorporated both the idea of multiple representations for bivariate data and a representation for the univariate measurement data situation. Lisa, for example, conceptualized multiple graphical representations for tasks 1.5 and 1.6, and one graphical representation for task 1.7. She decided upon the use of a dotplot in the univariate measurement data situation, saying that such a graph would make it easier to see concentrations of values and the overall average value for the variable. In general, responses reflecting this pattern all incorporated multiple representations for some bivariate situations and one graphical representation for the univariate situation.

One final pattern of response was identified in regard to representing data. In this pattern, students incorporated the idea of multiple representations for data in both univariate and bivariate data situations. Bill, for example, conceptualized multiple representations for tasks 1.5 and 1.7, and one representation for task 1.6. When presented the univariate measurement data situation in task 4.8, he discussed using both a line graph and a bar graph to display the data. Responses reflecting this pattern differed from those at the previous pattern in the sense that multiple representations were discussed for both univariate and bivariate situations.

Responses to data representation tasks differed in degree of complexity. At the least sophisticated pattern of response evident, only one graphical representation was conceptualized for each of the bivariate data situations, and none could be conceptualized for univariate measurement data. Responses reflecting the next pattern incorporated either the idea of multiple representations or the representation of both univariate and

bivariate data. It was not until the next pattern of response that students incorporated multiple representations as well as representations for both bivariate and univariate situations. The most sophisticated responses incorporated the idea of multiple representations across both univariate and bivariate situations.

Analyzing Data

The process of analyzing data involves identifying trends and making inferences or predictions from a data display or set. Responses to tasks in the interview protocol led to the identification of patterns of response for eight different subprocesses within the process of analyzing data. The first of the subprocesses includes patterns of response involving the comparison of univariate data sets (2.1, 2.7). The second subprocess involves the analysis of the relationship between a sample mean and the population from which it was drawn (2.8). The third of the subprocesses requires the identification of atypical points within a tabular data set (3.3). The fourth describes patterns of response for making a multiplicative comparison (5.7).

The last four subprocesses all involve students' analysis of bivariate data sets containing two measurement variables. The patterns of response evident when students identify atypical points in such sets of data were investigated and described (task 4.5). Interpolation (task 4.3) and extrapolation (task 4.4) were the next two subprocesses investigated and described. Finally, responses given to tasks involving the description of the relationship between two variables (4.2 and 4.10) led to the identification of patterns of response for the subprocess of describing bivariate relationships when both variables involved are measurement variables.

Some of the tasks administered during the interview sessions that were designed to elicit students' thinking about data analysis did not prove to be helpful in discerning patterns of response. When task 2.9 did not elicit different patterns of response among the first few students interviewed, I changed one of the numbers in the problem. This led to the construction of a better interview item for future studies, but it made comparison among students' responses for the purpose of this study impossible. Tasks 3.5, 3.6, 4.6, and 4.7 were designed to elicit thinking about formal inference. None of the students interviewed demonstrated such thinking, and there was no discernible difference in patterns of sophistication among the responses given. Finally, tasks 5.3, 5.4, and 5.5 were designed to elicit students' thinking about using simulations or formulas to find the likelihood of events. None of the students interviewed used these strategies to answer the questions, and the responses that were given did not indicate a variety of patterns of thinking. The patterns of response reported below for each of the subprocesses come from the tasks in the interview protocol that did serve to indicate a variety of patterns of thinking. In subprocesses that included more than one task, I categorized students by the most sophisticated response they gave to the tasks.

Subprocess 1: Comparing Univariate Data Sets

The comparison of univariate data sets is an important data analysis skill that has been investigated and described in several studies of middle school students (e.g., Watson & Moritz, 1999a; McClain & Cobb, 2001). Two of the tasks included in the interview protocol served to elicit a variety of patterns of response from students in regard to the comparison of univariate data sets. Task 2.1 asked students to decide which two classes

in a set of data were the most similar in performance based upon the variable of test score. Task 2.7 required students to examine two graphs showing distributions of test scores for students in two different states and compare the two states using the graphs. The patterns of response summarized in figure 11 and described in detail below come from the analysis of students' responses to the two tasks. In cases where a student used a more sophisticated approach in one situation than in the other, the student is classified within the chart in figure 11 according to the more sophisticated pattern of response given.

Pattern descriptor	Students whose responses reflected the pattern
Compares sets of data by using multiple relevant attributes of the aggregates.	Kristen, Jeff
Compares sets of data by using one relevant attribute of the aggregate sets such as shape, center, or spread. The comparisons made are often supplemented by point-by-point comparisons or irrelevant features of the data sets.	Laura, Julie, Luke, Crystal, Nancy, Paul, Hillary, Lisa, Bill, Daniel, Jessica
Compares sets of data by using a point-by-point strategy.	Rick, Brooke

Figure 11. Comparing univariate data sets

The least sophisticated responses within the subprocess incorporated only one strategy for comparing the univariate data sets in each of the tasks. They included only point-by-point comparisons in each case. Rick's comparison of test scores for two classes (task 2.1) relies upon point-by-point comparisons:

I think it would be B and C. Even though B has two less students than C did, I think they were very similar. If you look at their lowest kid's low score at 7, and if once you go and keep on going, they both had a kid that got in the 20's that got all of them right. They both had kids, actually two of them, that only got one wrong, that was 19. Both had a kid that only scored 15. And then they also both had the same kid that scored 14. And 13 and 12 is very similar. And the biggest difference would have to be that two kids are missing, and in class B the kid scored 9 and the...in class C the kid scored 11. And that's the biggest difference between scores, is by 2 points. That, and that there's two kids missing. But out of those two classes, class B and class C are very similar.

His comparison of test scores in two different states (task 2.7) also uses only point-by-point comparisons:

Um, then I would say that the scores in Wisconsin seem to be higher than the state of Arkansas. Seems like there are fewer 7's, but there are a lot more 11's and 13's. Even the 20's and 19's seem to be higher. And everything seems to be higher, except for the 1's and the 3's and the 5's, that's all lower. But it keeps on going up, once you keep on going up, the test scores eventually are better.

Students' explanations within this pattern of response gave little evidence that they viewed data sets as aggregates, since relevant features associated with aggregates such as shape, center, and spread were not incorporated.

Responses reflecting the next pattern gave more evidence that students viewed data sets as aggregates, since they used one relevant characteristic such as shape, center, or

spread in order to make comparisons among sets. In some explanations reflecting this pattern, however, point-by-point comparisons and irrelevant features of data sets were incorporated in comparisons. Julie, for example, gave the following response when asked to compare the test score data sets in task 2.7: “I would say that overall the students in Wisconsin had a higher average on the test, but there were still students in both states that scored either a 1 or a 20. So, it was even that way, but overall the students had a higher average test score in Wisconsin.” The data sets are compared on the basis of the relevant characteristic of the average. A point by point comparison is also incorporated, as she compared the numbers of students getting high and low test scores in each data set. In general, responses reflecting this pattern compared data sets based upon one of the relevant characteristics of each set of data, and were at times supplemented by point-by-point comparisons or irrelevant features of the data sets.

The responses reflecting the next pattern clearly demonstrated the view that the data sets being compared were aggregates with several different important characteristics. Responses reflecting this pattern incorporated two different relevant features in drawing comparisons between data sets. One student, Jeff, gave two responses reflecting the pattern. He examined the range of scores and the centers of the test score data sets (task 2.1), stating, “The range of values in both of them are from 7 to 20. The middle looks to be right around 15 for both.” In drawing comparisons between the test score data sets presented in task 2.7, he used the characteristics of shape and center, stating, “I’d say it’s slightly higher. Because I would say the middle is about 10 or 11. This one’s pretty symmetrical, peaking at 11.” Both responses reflected the most sophisticated pattern of

thinking, since in each one he used two different relevant characteristics to draw comparisons between data sets. The first response incorporated spread and center, and the second incorporated center and shape. Responses reflecting this pattern indicate a perception that data sets are aggregates with several different important characteristics that can be used in order to draw comparisons.

In summary, three patterns of response were discernible within responses to interview tasks involving comparisons of univariate data sets. The first pattern described incorporated only point by point comparisons in response to the tasks. The next pattern saw students begin to use one relevant characteristic of the data sets as aggregates in order to draw comparisons, with these comparisons often supplemented by irrelevant features of the data sets or else point-by-point comparisons. Responses reflecting the next pattern incorporated multiple relevant characteristics in drawing comparisons. It was clear in responses reflecting the most sophisticated pattern that data sets were treated as aggregates with several multiple characteristics that can be used for the purpose of drawing comparisons.

Subprocess 2: Analyzing Sample Means

Understanding the relationship between a statistic and a parameter is an important part of the foundation for statistical thinking (NCTM, 2000). Several researchers have investigated students' intuitive and formal understandings of the relationship between statistics of samples and the corresponding parameters of the populations from which they were drawn (e.g., Confrey & Makar, 2001; Chance, Garfield, & DelMas, 2001; Konold, Well, Lohmeier, & Pollatsek, 1993; Saldanha & Thompson, 2001; Shaugnessy,

Watson, Moritz, & Reading, 1999). My study included an interview question to examine students' thinking about the relationship between the mean of a sample and the mean of the population from which it was drawn. Task 2.8, similar in structure to items used by Konold, Well, Lohmeier, & Pollatsek (1993), elicited different patterns of response in regard to the effect of sample size upon that relationship. The patterns are summarized in figure 12.

Pattern descriptor	Students whose responses reflected the pattern
Recognizes that the larger the random sample, the more likely its mean will be close to the population mean; and that smaller samples are more likely to produce extreme results.	Julie, Hillary, Laura, Luke, Lisa, Bill, Crystal, Jessica
Recognizes that the larger the random sample, the more likely its mean will be close to the population mean.	Jeff, Kristen, Nancy, Brooke
Recognizes that different size samples can be drawn from a population, but does not recognize the impact of changing sample size upon the sample mean.	Paul, Daniel, Rick

Figure 12. Analyzing sample means

Responses reflecting the least sophisticated pattern of response identified within the subprocess indicated that students recognized that different size samples can be drawn from a population, but demonstrated little understanding of how sample size influences that variability of the sample mean. For example, when Paul was asked if he thought there would be any difference between the means of two samples of different sizes, he replied, “No. Um, because it is a simple random sample, so everyone has an equal

chance of getting picked. So by trying to do that, the ones that were picked will be spread out and averaging out to be basically the same, and that would be the same for both groups.” When asked if one of the samples was more likely than the other to yield an average of 17, he stated, “No. Because not...too few of students scored 17 or above. It’s not going to average out to be 17, most likely below, because most students did score less than 17.” Responses reflecting this pattern revealed little more than the recognition that samples of different sizes can be drawn from a population.

In the next pattern of response, students understood what it meant to draw a sample from a population, and also understood that the larger the random sample drawn, the closer its mean is likely to be to the population mean. They did not, however, have sophisticated enough understandings of sample mean variability to realize that a small sample is more likely than a large sample to produce an extreme result. Brooke’s responses to each part of the task help to illustrate. In response to the question of whether or not there would be a difference between the averages of a large and a small sample, she replied, “I think that there would be a difference. Because when you only survey or check 5 students’ scores, you really don’t have enough data to see where the average was, but when you take 15, you have a little bit better of an idea. Chances are that all students won’t have a separate score, most of them will have some type of repeating number that would be the same thing.” Her response seems to indicate an intuitive understanding of the fact that the mean of a large sample is usually nearer the population mean than the mean of a small sample. However, when asked if one of the two sample means was more likely than the other to produce a result greater than 17, she replied, “Probably the second

teacher, because she has a lot more students to choose from. Adding them all together would probably give a bigger number.” Her response indicated that she had the correct relationship reversed. She thought that the larger sample was more likely to produce extreme results. Her responses were more sophisticated than those reflecting the previous pattern, but lacked the connectedness expected of a more mature response.

Students whose responses reflected the most sophisticated pattern identified for the subprocess demonstrated a more complete intuitive understanding of the relationship between the means of samples and the means of populations. Hillary’s responses to both parts of the given task exemplify this type of connected intuitive understanding. In response to part a, she stated, “The one that takes the 15 will probably get closer to the real average of all the students, just because she’s using more scores. And, if the other one only used 5, then she could get some extremes in there. Even if it is at random, she could pick one extreme, and it would ruin it.” For part b, she responded, “Um, the one that picked 5 will be more likely to get something above 17. Because there’s only about...there’s only, like a small percentage that got above 17. So, if they pick 15, they’re going to have to start picking from the ones that are lower than 17, and that will bring the average down. But if they are only picking five, then they could pick the five that are, all five could be above 17.” The responses reflecting this pattern demonstrate a more complete intuitive understanding of the nature of the subprocess that was lacking at the previous patterns of response identified.

Three different patterns of response, then, were identified within the subprocess of analyzing sample means. The least sophisticated pattern was marked by students grasping

only the relevant concept of sampling, and the most sophisticated was marked by students demonstrating a more complete intuitive understanding of the relationship between sample size, sample mean, and population mean. Notably, none of the students interviewed discussed the formal notion that the distributions of small sample means do, across all contexts, exhibit greater variability than the distributions of large sample means of samples drawn from the same population. Instead, all students reasoned intuitively from the given context and did not incorporate general principles used in formal methods of data analysis.

Subprocess 3: Identifying Atypical Points in a Tabular Data Set

Part of the process of analyzing data is making comparisons within data sets. This competency falls within Curcio's (1987) category of "reading between the data." Two different questions in the interview protocol gave students opportunities to make comparisons among the points in tabular data sets and identify atypical values. Task 2.4 asked students to identify any unusually high or low values in a set of test scores for a class, and task 3.3 asked students to identify any unusually high or low values among salaries listed for department store workers. Each of the two questions gave students the opportunity to identify outliers within the data sets and give reasons why they considered the points identified to be outliers. The two tasks differed from those included in Friel's (1998) study in that the formal term "outlier" was not included in either question, and that the data were presented in the form of tables rather than boxplots. The purpose of putting the data in tables rather than boxplots was to allow students to generate their own strategies for identifying atypical points rather than suggesting a strategy through a given

representation. The patterns of response that were obtained within the subprocess are summarized in figure 13. In cases where students gave responses reflecting different patterns of sophistication for each task, the students' names appear in the table next to the more sophisticated pattern of response given.

Pattern descriptor	Students whose responses reflected the pattern
Formal procedure used to identify atypical points	Paul
Distance from main cluster of points or distance from the center of the data set is discussed in the identification of atypical points.	Luke, Laura, Jeff, Brooke, Julie, Daniel
Distances between individual data points are discussed in the identification of atypical points.	Bill, Hillary, Kristen, Jessica, Rick
Qualitative criteria alone are used to identify atypical points. No quantitative comparisons are made between points in the data set.	Lisa, Crystal, Nancy

Figure 13. Identifying atypical points in a tabular data set

The least sophisticated pattern of response within the subprocess was characterized by responses that took only contextual factors into account in determining atypical points. Crystal gave the following characteristic response when asked if she saw any unusual values in a chart showing department store worker salaries (task 3.3):

Yes, I do. I think 8.50 is pretty low. That's interesting that the highest paid person only gets 8.50 an hour. So, that's pretty low. But then, you go over to the men's, and there's one for 5.50. That's crazy. So, I don't know what kind of store that would be. (I: asks Crystal to explain why 8.50 is low). I would consider it to be low because they took the highest paid woman at that job, and usually that would

probably be a manager type person who gets the highest paid. And, seeing that's 8.50, I know that managers don't usually get paid 8.50. So, maybe that would say that all of the people in that business, all of the managers or higher people are men, and the women are just kind of, like, associates or something like that.

(In response to the second part of the task) I would say that 5.50 an hour is pretty low, again, just because it's supposed to be the highest paid person, the highest paid man at that company. And, that's like, minimum wage, or maybe below, so that's kind of different. And, maybe, again, um, that might be um, a place where a lot of women are higher up, because some jobs are dominated by women. So, it might be a job like that.

Identification of atypical points within the data set rested on contextual factors alone in this pattern of response. No quantitative comparisons among points within the data set were made. Qualitative criteria alone were used.

In the next pattern of response identified, students began to incorporate quantitative comparisons among points within the set in their explanations. However, these quantitative comparisons were generally quite unsophisticated, in that they simply involved discussion of jumps between individual data points in the set. Hillary gave a response characteristic of this pattern when asked if there were any unusually high or low values within a table showing test scores for a class (task 2.4). She stated, "Um, in class c, the 7 is pretty low compared to the other ones. Because the other ones are within 2 or 3 of each...like between each other there's 3, but between 7 and the next score, there's

4.” In order to identify atypical points, she examined the distances between individual points in the set. While responses reflecting this pattern incorporated quantitative comparisons, the quantitative comparisons that were made were simply those among distances between discrete points.

Responses reflecting the next pattern indicated that students saw data sets as aggregates. They used the quantitative criterion of distance from a central cluster of points in order to identify atypical data values in some cases. In other cases, they used the criterion of distance from the center of the set of data. Both of these strategies mirror the structure of common formal methods that are used to identify outliers (see Yates, Moore, & McCabe, 1998 for a detailed discussion of common formal methods). The strategy of examining distance from a central cluster of points comes out well in Luke’s response to task 3.3 (identifying unusually high or low salaries in a table). He stated, “Yeah, like the low salaries, the 5.50, the 6, and the 7, they look real low, because most of them are around 18 or 19 or 17 or 20. And so, those look like, the 5.50 especially, really low. But the higher values, the 20’s and the 19’s, those don’t look too high, it just seems about average.” The three data points identified as unusually low were labeled as such because of their relatively large distance away from the main cluster of data points in the set. Hence, salient characteristics of the data set, and not just quantitative differences among individual points, came into play in the identification of unusual data values in responses reflecting this pattern.

In the most sophisticated pattern of thinking identified, one student used a unique strategy to identify atypical points within each column of data presented in task 3.3

(department store worker salaries). Paul produced a boxplot for each column of data on a TI-83 calculator, and then examined each column for atypical values. This strategy is similar to responses reflecting the previous pattern, in that he examined the data sets for unusual values by judging their distance from a central cluster of points. However, his strategy is different from those reflecting the previous pattern in that he introduced the formal tool of the boxplot in order to help him make the comparisons. Such a strategy was not suggested by anything present in the context of the task or in the data set. Paul reasoned from the general principle that boxplots can be used as guides in identifying atypical points in univariate data sets.

An interesting array of patterns of response emerged in the identification of atypical points in a tabular data set. The least sophisticated pattern was marked by the use of only contextual criteria to identify atypical points, the second by the use of primitive quantitative criteria, and the third by the use of quantitative criteria that made use of the distances of points from a central cluster of points in the data set. One more sophisticated pattern of response was also evident, in that one student used a formal strategy to assist in the identification of atypical points. The formal method that was used mirrored the structure of response at the previous pattern, but it formalized the structure by using a boxplot to examine distance from the center of the data set to the highest and lowest values in the set.

Subprocess 4: Making Multiplicative Comparisons

Task 5.7 was included in the interview protocol in order to elicit different patterns of response from high school students in regard to analyzing two-way tables. In the task,

students were asked to examine a two-way table showing the number of trout and non-trout caught by two different fishermen fishing in the same lake. Students were told that one of the fishermen had used new bait designed to help catch trout, and the other did not. They were then asked whether or not they felt the new bait helped one catch more trout. Multiplicative reasoning, in some form, was necessary to give a complete answer to the task, since one fisherman had caught slightly more fish than the other. The requirement to use multiplicative reasoning was built into the task as a result of Cobb's (1999) observation that this type of reasoning is an important part of many data analysis tasks. The patterns of thinking that emerged in response to the fishing bait task (5.7) are summarized in figure 14.

Pattern descriptor	Students whose responses reflected the pattern
Makes only multiplicative comparisons when only multiplicative comparisons are warranted.	Paul, Crystal, Jeff, Luke
Makes additive and multiplicative comparisons when only multiplicative comparisons are warranted.	Hillary, Laura
Makes additive comparisons when multiplicative ones are warranted.	Kristen, Lisa, Bill, Julie, Daniel, Nancy, Jessica, Brooke, Rick

Figure 14. Making multiplicative comparisons

At the least sophisticated pattern of response, several of the students interviewed made only additive comparisons between fish populations and no multiplicative ones. Daniel gave the brief, yet highly illustrative response reflecting this pattern, “Not really. Because they only caught three less non-trout, and five more trout.” He was able to draw

comparisons that might have some relevance to answering the question. However, nowhere in the responses reflecting this pattern is the recognition of the fact that differing numbers of fish caught by the fishermen make additive comparisons very limited in value.

Student responses reflecting the next pattern incorporated both additive and multiplicative comparisons. However, no choice was made between the two types of comparisons. The responses indicated no preference for either type of comparison. Hillary's response reflected this pattern, since she stated, "I would think that it does, because for boat B, he caught less non-trout fish than boat A did, and he also caught more trout. And he also has a higher percentage of trout." In her response, both additive and multiplicative comparisons are made, and one is not preferred over the other. While responses reflecting this pattern incorporate one more relevant strategy than those reflecting the previous pattern, they do not indicate that one strategy is more suitable for drawing comparisons than the other.

Responses reflecting the next pattern did indicate a preference for multiplicative comparisons in the given situation. Jeff, for example, came to a decision by using percentages to draw a multiplicative comparison between the catches of the two fishermen. He stated, "I think it [type of bait] would [make a difference], because the guy who wasn't using the bait, he only caught a third of his catch was trout. The other guy, with the new bait, he caught, um, 47% of his catch was trout." His response indicated recognition of the fact that percentages are an efficient means of comparison in the given situation, since the numbers of fish caught by each fisherman were unequal. The use of

the most relevant strategy in order to come to a decision in the given situation indicated thinking characteristic of this pattern.

Three different patterns emerged among the responses given to the task within the subprocess of making multiplicative comparisons. The least sophisticated pattern was marked by students' use of only one weak strategy for drawing comparisons between the two fishermen. The second pattern was characterized by the use of both additive and multiplicative strategies with no preference expressed for either one. The most sophisticated pattern was marked by recognition of the need to make multiplicative comparisons in the given situation.

Subprocess 5: Identifying Atypical Points in a Graphical Bivariate Data Set

One of the tasks included in the interview protocol elicited patterns of thinking in identifying atypical points in data presented in a graph. In task 4.5, students were asked to identify atypical points within a bivariate data set that was accompanied by a scatterplot display. The two variables involved were both measurement variables, one being ice cream consumption and the other outside temperature. The responses to the task were used in order to formulate the set of descriptors for the subprocess of identifying the atypical points within a graphical bivariate data set. This subprocess was not merged with the subprocess of identifying atypical points in a tabular set because of important qualitative differences in the patterns of response that were apparent. Four different patterns of response emerged within the subprocess. Their qualitative characteristics are summarized in figure 15.

Pattern descriptor	Students whose responses reflected the pattern
Distance from a visual line of best fit is discussed in the identification of atypical points.	Lisa, Luke
Distance from the central cluster of points is discussed in the identification of atypical points.	Crystal, Hillary, Kristen, Julie, Jessica, Brooke
Subjective and ill defined quantitative comparisons are used in the identification of atypical points.	Laura, Jeff, Paul, Bill, Rick
Contextual criteria alone are used to identify atypical points.	Daniel, Nancy

Figure 15. Identifying atypical points in a graphical bivariate data set

The least sophisticated pattern of response within the subprocess was characterized by the use of contextual criteria alone to identify atypical data points. No evidence of the use of quantitative comparisons among values in the data set was given. Daniel’s response is typical of the least sophisticated pattern. When asked if any unusual points existed within the data set, he stated, “Yeah, it looks like they are selling quite a bit of ice cream when it’s only 23 degrees outside.” His response incorporated only a contextual factor from the setting of the task, namely that one would not expect to sell much ice cream when it is 23 degrees outside. The point with an x-coordinate of 23 degrees itself did not deviate in any noticeable manner from the overall trend of the data (see scatterplot accompanying task 4.5). The apparent lack of attention to the overall trend of the data characterized responses reflecting this pattern.

In the next pattern, quantitative comparisons were incorporated within the responses given. However, they were not incorporated in a coherent manner. Rick’s response to the task helps to illustrate this pattern of response. He stated (referring to the scatterplot

accompanying task 4.5), “The unusual point that I saw was around 47 degrees. That seemed very, very low. That received .3. I mean, if you just take 40 degrees, then you’re degrees colder...some people were eating .4. There was a huge difference. Also, when it was 78 degrees, it was below .4. And when it was 40 degrees, some people were eating ice cream to .4, so...4 seemed unusually high, and .70 seemed unusually low. Those are the extremes.” His explanation seems to jump from point to point. While quantitative comparisons are incorporated, no systematic comparison strategy is evident in the responses reflecting this pattern.

In the third pattern identified, a systematic quantitative comparison strategy became evident. It also became evident that students began to see the data set as an aggregate rather than a set of discrete points. When asked to identify atypical points, responses reflecting this pattern discussed the distances points were from the central cluster of points within the data set. Brooke’s response exemplifies this pattern. She stated (see the scatterplot accompanying task 4.5), “Not really unusual, there’s somewhat of a peak at about 72 degrees, where there’s more people eating ice cream. And at that same point, there’s less people eating ice cream. But other than that it all basically falls into a cluster.” This pattern of response was characterized by a clear indication that the data set was viewed as an aggregate, and that atypical points were those that fell outside the central cluster of the data.

One final pattern of response was identified in regard to the subprocess. In the most sophisticated pattern observed, students identified atypical points in the data set by discussing their distances from a hypothetical line of best fit. In response to the relevant

task, Luke stated, “Not really, maybe on the one day when it was 72 degrees, the highest point on the graph up here, that’s a little bit unusual, because it doesn’t exactly fit. But maybe people just wanted to eat a lot of ice cream that day...that looks like the only unusual one. The others stay about the same distance from the line of best fit.” Responses reflecting the most sophisticated pattern observed for the subprocess incorporated the formal statistical tool of the line of best fit in order to determine whether or not data points could be considered atypical.

The patterns of thinking for the subprocess of identifying atypical points in a graphical display and the subprocess for identifying atypical points in tabular displays have similarities and differences. At the least sophisticated pattern of thinking within each subprocess, contextual criteria are used without quantitative criteria in order to identify atypical points. In the next pattern within each subprocess, weak quantitative criteria come into play, although the nature of the weak quantitative criteria differs somewhat in each of the subprocesses. For the third pattern of each subprocess, it becomes apparent that students view data sets as aggregates, since distance from the central cluster of points is used in order to determine whether or not points are atypical. At the fourth pattern within each, a formal statistical tool is introduced in order to help determine whether or not points are atypical. The nature of the formal tool introduced and precisely how it is used differs for each of the subprocesses. By keeping the two subprocesses related to identifying atypical points separate, some of the detail about the nature of the formal tools being used is preserved. The slight qualitative differences of other patterns are also preserved.

Subprocess 6: Interpolating within Bivariate Data

An important part of analyzing data is making inferences and predictions from data sets. Because of the importance of this aspect of analyzing data, in task 4.3 students were asked to do so with a set of bivariate data. For the task, they needed to predict a reasonable value for the ice cream consumption given a value for temperature that was not listed in the given data set. The patterns of response for interpolation are summarized in figure 16.

Pattern descriptor	Students whose responses reflected the pattern
Visually determined best fit line used to interpolate.	Bill, Lisa, Luke
Trend of the data is used to interpolate.	Laura, Paul, Jeff, Jessica
Data in the immediate vicinity is used to interpolate.	Daniel, Julie, Kristen, Hillary, Crystal, Nancy, Brooke, Rick

Figure 16. Interpolating within bivariate data

In the least sophisticated pattern of response identified, students used quantitative criteria to make the necessary interpolation, but they used only data in the immediate area. Hillary's response to the task illustrates this pattern well. Looking at the graph and the data in the area of interest, she stated, "I'd say [the total consumption would be] somewhere around .3 or .5 [pints of ice cream per 100 people]. Because if you average [the y-coordinates] between when it's 44 and 55 degrees, it's .34 [pints of ice cream per 100 people], but if you add in the one [data point] where it's 52 degrees, then it [the total

consumption] is .375 [pints of ice cream per 100 people]. Because there is no measurement for 50 [degrees], so I would just use the ones [the observations] that are closest to 50 [degrees] to kind of guess.” Responses reflecting this pattern gave no indication that students saw the data as an aggregate, since only data in the immediate vicinity were used in order to make an interpolation. The trend of the data set as a whole was not mentioned.

In the next pattern, the trend of the data was mentioned in the explanation of the interpolation process. Paul’s response is indicative of this pattern. He predicted, “Around .4. Um, it’s in the middle, it seems like it kind of separates into two groups. Below 50 degrees, um, the average consumption is less than .4, but above 50 it is above. So for the average, it would be around the .4 area.” His explanation indicated that he took the trend of the entire data set into account in the process of interpolation by taking into account the data both above and below 50, even though this data was not all in the immediate vicinity of interest. The response illustrates the use of quantitative criteria for interpolation used in concert with the view of the data set as an aggregate.

One last pattern of thinking was identified within the subprocess. In the next pattern, students used a hypothetical visually determined best-fit line in order to help in the interpolation process. Luke was one of the students who used such a strategy. In reply to the question, he stated, “I would say about 3.8 pints per 100 people...or .38, I mean. Well, like I said, it’s going to be a straight line through here, so I just kind of went up to 50 and saw where the point would be, like the middle of the points that are around it. And it looks like it’s going to be close to .4 but not all the way to exactly .4.” The introduction

of a hypothetical object to help reason out the answer was a unique characteristic of this pattern of response.

In summary, three patterns of response were evident within the interpolation subprocess. In the first pattern, quantitative criteria were used in order to interpolate, but only the data in the immediate vicinity of the point of interest was taken into account. Responses reflecting the next pattern took the trend of the entire data set into account in making an interpolation. The most sophisticated pattern identified was that at which students introduced the idea of a hypothetical line of best fit.

Subprocess 7: Extrapolating from Bivariate Data

In task 4.4 (the Ice Cream task), students were asked to make a prediction for the amount of ice cream consumption at a temperature lower than any of those given in the data set. While the process of extrapolating from bivariate data would seem on the surface to elicit the same patterns of response as those observed in interpolating within bivariate data, many more patterns of response were identified for extrapolation. Six different patterns of thinking were identified, as shown in figure 17.

Pattern descriptor	Students whose responses reflected the pattern
Formally determined best fit line is used in order to extrapolate.	Paul
Visual/algebraic best fit line is used in order to extrapolate.	Bill
Visual line of best fit is used in order to extrapolate.	Hillary, Lisa, Jeff, Luke, Jessica
Trend/direction of the data is used in order to extrapolate.	Laura, Crystal, Nancy, Brooke
Data that would occur in the immediate vicinity are used in order to extrapolate.	Kristen, Julie, Rick
Context and personal theories are used in order to extrapolate.	Daniel

Figure 17. Extrapolating from bivariate data

The least sophisticated pattern of response to the task was given by Daniel. He made use of non-quantitative contextual factors alone in order to make an extrapolation. When asked to predict how much ice cream would be consumed if the temperature hit 10 degrees Fahrenheit, he replied, “Probably none at all. Because they could just go outside and grab some snow and make their own snow cones.” This response was easily the least sophisticated of all those obtained. No quantitative criteria for the decision were articulated whatsoever.

In the next pattern of response to the task, students generated quantitative data based on personal theories in order to predict data points that may occur in the area of interest to the prediction. Once data points had been predicted for the area, the prediction was thought to be close to the conjectured points. Julie, for example, looked at the data that were closest to the area of interest and made a prediction based mainly upon those data,

and not the overall trend. She predicted (see scatterplot accompanying task 4), “Well, based on the graph and the lists, the lowest point on here is only 27 [degrees], and that’s not even represented, so there would be none. According to this information, it would be zero pints.” In this pattern of response, only an imagined part of the data set was used in order to make a prediction. The perception that data are perceived as an aggregate is not apparent.

The next pattern of response to this task did provide evidence that students viewed the data set as an aggregate. In responses reflecting this pattern, the trend or direction of the data was used in order to make a prediction. Crystal’s response captures the nature of this pattern well. She stated, “OK, well, it doesn’t show under 20 degrees Fahrenheit, so they didn’t do a survey of that, but seeing that it’s a positive association, you could just take from that relationship, and I would just assume that it would be under .25, maybe around .2 pints would be consumed. Just looking at the association that’s already going, and taking in backward, making it negative and seeing that like at 25, it’s going to be about .25, so it’s going to decrease from that, and so I’m saying .2.” This pattern of response was marked by students taking the entire data set into account in making an extrapolation.

At a more sophisticated pattern of response, students introduced a hypothetical visual best fit line in order to help make extrapolations. One student whose response was characteristic of this pattern was Hillary. In response to the task, she stated, “I’d say .24, because if you draw a line of best fit in there, then you get to about 10 degrees would be .24 or .23.” Students giving responses reflecting this pattern not only paid attention to the

trend of the overall data set, but also introduced a hypothetical best fit line in order to capture the trend and make reasonable predictions from it.

Bill's response to the extrapolation task was a bit more sophisticated than those at the last pattern detailed. He used a visual line of best fit in order to extrapolate, just like the students giving responses reflecting the previous pattern. However, Bill also assigned an algebraic equation to the line he had constructed in order to help make his extrapolation more precise. He explained, "I would actually do an equation. So, basically an equation $y = mx + b$. And, you said if $x = 10$. My slope is about, we agreed on, $.375$ over 30 . Plus about $.25$. I mean, this is an approximation. I would say about $.2625$." His response differed from those reflecting the previous pattern in that he introduced an algebraic equation for the visual line of best fit he had constructed.

One slightly more sophisticated pattern appeared in response to the extrapolation task. Paul gave a response that can be characterized as more sophisticated than that of the previous patterns. Like the students who gave responses reflecting the previous two patterns, Paul used a best fit line in order to make an extrapolation. However, unlike the students who used visually determined best fit lines, Paul relied upon a formal procedure to determine the line of best fit. Recognizing that a least squares regression would produce an objective line of best fit rather than a subjective one, he entered the data into two lists in the TI-83 calculator. He then used the calculator to generate a line of best fit, and he determined the value of the line at 10 degrees. His response was considered more sophisticated than those reflecting the previous patterns because he tied together all of the relevant aspects of the problem in order to make an objective extrapolation.

The extrapolation task elicited an exceptionally wide array of patterns of response. Specifically, it elicited a much wider array of patterns of response than the interpolation task. Although the descriptors for the two subprocesses could be merged into one set, they were kept separate in order to emphasize the differences in patterns of thought the two different tasks elicit. Responses to the extrapolation task involved elicited more responses reflecting the less sophisticated patterns of thinking. At the same time, they also elicited more sophisticated use of the formal statistical tool of the line of best fit.

Subprocess 8: Analyzing Bivariate Relationships

The last of the subprocesses examined under the overarching process of analyzing data was the subprocess of analyzing bivariate relationships. NCTM (2000) considered the analysis and description of bivariate measurement data important, stating that for such data, high school students should “describe its shape, and determine regression coefficients, regression equations, and correlation coefficients using technological tools” (p. 324). As students analyze such bivariate relationships, they also need to understand that if two variables are correlated, “the correlation may be due to an underlying cause” (NCTM, 2000, p. 328).

Two interview tasks were designed in order to prompt students to analyze a given bivariate relationship. In the first task (4.2), students were asked to explain as fully as possible the bivariate relationship between the two measurement variables of temperature and ice cream consumption presented. In the second task (4.10), students were asked to comment on the possibility of a causal relationship between the two variables involved. Their responses to the two tasks revealed four different patterns of response within the

subprocess. The patterns seem to be differentiated by the manner in which students answered questions about causality between the two variables involved. Those patterns are summarized in figure 18.

Pattern descriptor	Students whose responses reflected the pattern
Goes beyond discussing the direction of the quantitative relationship between the two variables displayed. Identifies intervening variable beyond the two variables shown in the data set.	Julie
Describes direction of the relationship between the two variables involved and uses that relationship and quantitative anomalies to argue for or against a causal relationship between the two variables.	Lisa, Bill, Rick
Describes direction of the relationship between the two variables involved and uses that relationship to help argue for a causal relationship between the two variables.	Luke, Nancy, Crystal, Hillary, Paul, Brooke, Kristen
Describes direction of the relationship between the two variables involved.	Daniel, Jeff, Jessica, Laura

Figure 18. Analyzing bivariate relationships

Students giving responses reflecting the least sophisticated pattern within the subprocess demonstrated the ability only to describe the general direction of the quantitative relationship between the two variables involved. Daniel’s description, for example, simply was, “As it gets warmer, people eat more ice cream.” While this captures part of the relationship between the two variables involved, it only captures one very small part of the overall relationship. In this pattern of response, the quantitative relationship was not referred to in arguing for or against a causal relationship between the two variables.

In the next pattern of response identified, students began to use the data given in order to form arguments about cause and effect relationships between the two variables. Crystal, for example, commented, “It’s a positive association, it pretty much goes upward as the temperature gets warmer, more ice cream is consumed, which makes sense, because it’s supposed to cool you off it’s kind of a summer-type food.” When asked if she thought that there was a causal relationship between the variables of ice cream consumption and temperature, she used the data set to back up a claim that temperature caused ice cream consumption to increase. She stated, “I would agree with that statement [that higher temperatures cause ice cream consumption to increase], because this study shows that there’s a positive association. When temperature increases, more ice cream is consumed. So, I would agree with that.” In responses reflecting this pattern, an explicit connection was made between the data display and any proposed causal relationship between the two variables.

In the next pattern of response within the subprocess, students went further in forming arguments about the quantitative rationale for causal relationships between the two variables presented in the problem. In addition to arguing that the overall trend of the quantitative data and the context seemed to indicate a causal relationship, responses reflecting this pattern identified anomalies within the data set presented that seemed to provide evidence against a direct causal relationship. When asked to comment on a newspaper headline saying that one variable in the data set caused the other, Bill said, “I agree. Basically the data set shows exactly what he said. You could even go on to say in some instances the higher temperature may not have directly caused the higher

consumption, such as the unusual point [he had previously identified the point (71, .549) on the scatterplot in task 4 as the “unusual point”]. But overall, his statement is agreed with.” Therefore, responses reflecting this pattern were characterized by the lack of complete acceptance of a simple causal relationship between the two variables presented in the data set.

A more sophisticated pattern was identified in Julie’s responses to the tasks within the subprocess. The following exchange took place when Julie was asked to comment on the possibility of a causal relationship between the two variables presented in the data set:

Julie: Well, since there is a correlation, you can say the temperature affects the ice cream consumption, but you can’t say that it causes, because there are other factors that may influence ice cream consumption. Temperature doesn’t cause people to eat ice cream, so...

I: What other factors might come into play?

Julie: When it’s a high temperature, people may go outside more, so ice cream might be a way that people relieve warmness, but that would be the most obvious factor. And when it’s warmer, people would rather go out because of the weather to eat ice cream...so...yeah.

Julie’s response indicated that she recognized that quantitative evidence presented within a data set does not necessarily mean that one variable causes another. She went on to identify a variable not directly shown in the data set that could have an impact upon ice cream consumption. Her response was more sophisticated than the others in that she hypothesized about the impact of a relevant variable not included in the given data set.

The last subprocess identified within the process of analyzing data included four different patterns of response. The first three patterns became progressively more sophisticated until students questioned simple cause and effect relationships between two variables by referring to quantitative evidence presented within the given data set. One student exhibited a more sophisticated pattern of response by going beyond the quantitative data presented and basing an argument against simple causation on a single relevant variable not explicitly presented in the data set.

Collecting Data

The process of collecting data includes planning, conducting, and critiquing surveys, experiments, and observational studies. Three different subprocesses within the process of collecting data were addressed: designing a non-experimental study, designing an experimental study, and critiquing a study. Five tasks (1.1a, 1.1b, 1.1c, 1.1e, 3.7) elicited students' thinking within the subprocess of designing non-experimental studies. One task (1.1d) elicited thinking within the subprocess of designing an experimental study, and one (1.2) led to the identification of patterns of response within the subprocess of critiquing a study.

One set of tasks (3.1, 5.1) was designed to elicit students' thinking about randomness, which is an important part of both experimental and non-experimental design. However, because of students' different understandings of the contexts in which the tasks were set, it was not possible to discern patterns of response about randomness from the responses given to the tasks. Students' success in responding to these tasks seemed more dependent upon their knowledge of fishing and department store

organization than their knowledge of randomness. Hence, the patterns of response described in this section were identified as a result of the tasks that involved designing a non-experimental study, designing an experimental study, and critiquing a study.

Subprocess 1: Designing a Non-Experimental Study

Important statistical thinking skills for high school students include being able to “understand the differences among various kinds of studies and which types of inferences can legitimately be drawn from each” (NCTM, 2000, p. 324) and to “know the characteristics of well-designed studies, including the role of randomization in surveys” (NCTM, 2000, p. 324). In tasks 1.1a, 1.1b, 1.1c, and 1.1e, students were asked to design studies in order to answer questions of interest about people who live in the state of Florida. No study design was imposed upon them, and they were free to approach the questions in any manner they deemed reasonable. In task 3.7, students were asked to expand a survey that had taken place in one state in order to include the entire United States. Students’ responses to the tasks led to the formulation of the descriptors for patterns for the subprocess summarized in figure 19.

Pattern descriptor	Students whose responses reflected the pattern
Concern for the representativeness of study participants is articulated, and randomization used in concert with stratification in some cases to obtain it.	Paul, Crystal, Hillary, Kristen
Concern for the representativeness of study participants is articulated, and stratified sampling or random sampling is used in some cases to obtain it.	Rick, Bill, Luke, Daniel
Concern for representativeness of study participants is articulated, but no strategy is described to attempt to achieve it.	Jeff, Lisa
Discusses data gathering instruments or procedures without articulating concern for the representativeness of those to be included in the study.	Laura, Julie, Brooke, Jessica
Discusses data gathering instruments or procedures, but primarily relies upon pre-existing studies for information.	Nancy

Figure 19. Designing a non-experimental study

The least sophisticated pattern evident in response to the subprocess of designing a non-experimental study was that exhibited by Nancy. She relied primarily on pre-existing studies in order to gather the information needed to answer each question posed. She briefly mentioned interviewing people to determine whether or not the governor would be re-elected in the next election, but the idea was not developed. She made use of pre-existing information and studies in books, periodicals, and the internet in order to answer each of the other questions in the tasks. For example, when asked to determine the success of a law that raised the minimum driving age from 16 to 18, she stated, “Well, it seems like I’m relying on the internet a lot, but that’s basically how I would. I guess you’d have to look up the accident claims from insurance companies and see if the claims

were higher or lower or whatever after the law was passed.” While she recognized that empirical data would be useful for answering the questions of interest, she relied on others to gather the data for her rather than developing her own data gathering techniques.

In the next pattern identified, students recognized the need for empirical data and also began to develop their own ideas for gathering the data. However, they discussed their data gathering techniques without expressing concern that the data they would gather would be representative of the population from which it was drawn. Brooke, for example, proposed the following plan for predicting whether or not the governor would be re-elected in the election in fall:

In response to the election question, I think I would do some kind of poll that people could answer on the internet or, like, by responding to a phone number in the newspaper. And you also would send door to door scouts out, that ask people if they vote, are they planning on voting for the governor. Then I would probably just display that by a percentage of people that we polled.

Although she discussed a number of data gathering techniques, none of them attempt to ensure that the sample drawn would be representative of the overall population. In fact, the data gathering techniques proposed by students whose responses reflected this pattern were quite likely to produce non-representative samples.

In the next pattern, students proposed data gathering techniques and recognized the importance of obtaining representative samples in sampling situations. Jeff, for example, recognized that it would be important to obtain data for the governor’s re-election from

“a wide range of the population from various backgrounds.” In response to the same question, Lisa felt it was important to “make sure that there’s all types of people, and that there is no economic bias” within a surveyed sample. However, she did not offer a strategy for obtaining such a representative sample. Therefore, the students responding within this pattern not only proposed data gathering techniques, but also at some point articulated that it was important for the data gathered to be representative of the larger population.

Students giving responses reflecting a more sophisticated pattern proposed formal methods to attempt to ensure that the sample drawn in sampling situations would be representative of the population of interest. In this pattern of response, students incorporated either simple random sampling or stratified sampling in order to attempt to gain a representative sample. Bill, for example, suggested using stratified sampling when asked to expand the department store survey to the entire United States (task 3.7), saying, “You could take a direct approach, call each department store, that’s just crazy. Or like the census bureau does, you don’t have to go to every single house. You could go to, I mean, you could research department store wages in a couple of states, north, east, south, west, all around, and then get an average of that.” Daniel suggested a random sampling strategy in response to the same question, saying, “He could pick some cities at random, and pick one man and one woman from a department store from that city chosen at random.” Responses reflecting this pattern incorporated one formal strategy to attempt to obtain a representative sample.

In the most sophisticated pattern of response identified within the subprocess, students incorporated both random and stratified sampling in the design of some studies. Paul's response to the question of determining the typical income of adults in the state of Florida (task 1.1a) illustrates:

OK, um, there's a couple of ways, I guess, you could do this. One would be to do a census, you know, and take information from everyone. You'd have to find ways to get to everyone, so they'd have to mail it back, and make sure people...send people out to go find them...that's one way. Another way would be to do a simple random sample, I might stratify the districts, or the counties, I guess. So, you'd have the same amount if there's urban, rural, suburban areas. Then take the same amount...or even it out that way. Then I guess you'd just find the mean salary for each of the adult households.

While students giving responses reflecting this pattern showed evidence of more complex thinking than students giving responses reflecting the previous pattern, responses reflecting this pattern did lack components that one would expect in more sophisticated responses. Stratification and randomization strategies were both used in some cases, but some of the important mechanics and logistics of the strategies were not discussed.

The five patterns of thinking identified for designing a non-experimental study become increasingly more sophisticated along the dimension of obtaining representative data. Students' thinking about representativeness has emerged as an important component of statistical thinking in other studies as well (e.g., Watson & Moritz, 2000a; Zawojewski & Shaugnessy, 2000). At the climax of the first three patterns, students expressed concern

about the representativeness of at least some of the samples drawn. In the fourth pattern, they incorporated one formal method, either stratified or random sampling, in order to obtain a representative sample in some cases. At the fifth, there were able to use both strategies together in one study design. A more sophisticated pattern of thinking would presumably involve discussion of the mechanics of random and stratified sampling.

Subprocess 2: Designing an Experimental Study

NCTM (2000) recommended that high school students understand experimental studies and be able to conduct them in order to answer quantifiable questions of interest. Researchers have begun to investigate how to design instruction in order to help students learn the principles of experimental design (e.g., Derry, Levin, Osana, Jones, & Peterson, 2000). In order to supplement the current research efforts underway in this area, a task was included on the interview protocol (task 1.1d) that elicited students' thinking about experimental design. In the task, students were asked to evaluate the effectiveness of a hypothetical drug that had just been developed. Five different patterns were identified in the responses to the task. They are summarized in figure 20.

Pattern descriptor	Students whose responses reflected the pattern
Recognizes the possibility of conducting an experiment when applicable and discusses two or more formal controls to be put on experiment in order to obtain valid results.	Julie, Paul
Recognizes the possibility of conducting an experiment when applicable and discusses a control to be put on the experiment in order to obtain valid results.	Crystal, Lisa
Recognizes the possibility for an experiment when applicable.	Bill, Luke, Daniel, Kristen
Uses data gathering method other than an experiment when an experiment would be applicable. No mention of the possibility of using an experimental design.	Hillary, Laura, Jeff, Jessica, Rick, Brooke
Relies solely upon pre-existing studies.	Nancy

Figure 20. Designing an experimental study

Nancy’s response to the task was indicative of the least sophisticated pattern of response evident. When asked how she would evaluate the effectiveness of the new drug that had been developed, she stated, “Well, I guess you’d have to go to like a direct source from...I mean if you couldn’t get the answers in a book or a periodical or online or something...If you could actually find someone who had actually encountered the virus and was working on a drug for it...if you had that kind of access.” Hence, she relied solely upon pre-existing empirical studies in order to answer the given question.

In responses reflecting the next pattern, students proposed their own methods for gathering data in order to answer the question. However, they did not mention conducting an experiment as a possible study design. Jessica, for example, proposed gathering information about success and failure from people who had used the drug, stating, “I

would speak to doctors, see what their opinion is. But, probably more importantly, talk to people who have contracted the disease and have taken the drug to see how they felt, have their symptoms got better, and how it worked out for them.” While several data gathering techniques were discussed in responses reflecting this pattern, none of them were experimental in nature.

Students whose responses reflected the next pattern recognized that it would be possible to conduct an experiment in order to obtain the needed information. Bill suggested testing the drug on animals. Daniel suggested finding some people who had the virus and testing the drug on them. Kristen and Bill both suggested analyzing the results of clinical tests without naming specific experimental subjects. Each response reflecting this pattern indicated the recognition of the possibility of carrying out an experiment in order to answer the question of interest.

In the next pattern, students not only recognized the possibility of conducting an experiment to answer the question of interest, but they also proposed a control to be put on the experiment in order to help ensure that the experiment would produce valid results. Crystal proposed, “For this one, I would take a group of people who actually have the West Nile virus, and I would make sure that they are all in the same stage, so that some aren’t worse off than others. Then I would give part of the people a fake-type drug, and one the real drug, and see how the differences in their improvement turn out.” Lisa recommended using animals as test subjects rather than humans, but she also incorporated the use of a placebo in her experimental design. The transition to this

pattern, therefore, was marked by the students' abilities to begin to discuss components of the design of an experiment they would use in order to answer the question of interest.

The responses of two students formed an even more sophisticated pattern, because they discussed experimental designs with more than one control that could be used in order to determine the effectiveness of the drug. Julie, for example, stated, "OK, well you would have to do some sort of experiment where you have the actual drug, and then, like a placebo. Two groups are chosen at random and put into random groups. Um, and it would help if it was double blind. And, just have one group taking the new drug and one group taking the placebo." Her design incorporated double blinding and a placebo. Paul's response was similar, and it mentioned the use of the same two controls. Presumably, responses reflecting an even more sophisticated pattern would include full discussions of the benefits and drawbacks of controls such as placebos and double blinding within experimental design.

In summary, five patterns of response were evident within the subprocess of designing an experiment. The first three patterns climaxed with students recognizing the applicability of experimental methods in the given situation. Responses reflecting the fourth pattern were marked by the ability to discuss the design of a hypothetical experiment for the situation in detail. The last two patterns identified reflected increasingly more complex experimental designs. Responses reflecting the fifth pattern identified included more detailed descriptions of quality controls that would be put on the experiment than those reflecting the previous pattern.

Subprocess 3: Critiquing a Study

NCTM (2000) emphasized that it is important for high school students to “know the characteristics of well-designed studies” (p. 324). Watson and Moritz (2000a, 2000b) have described some of students’ notions that impact their abilities to critique samples that are gathered for studies. I sought to extend this line of investigation in this study by investigating high school students’ abilities to critique a study given some information about the sampling design used for it. In task 1.2, students were asked to critique a study conducted by someone else. They were given several bits of information about the study, including the time at which it was conducted, its location, and the questions asked. There were three patterns of response evident. Descriptors were formed for each of the three patterns in order to characterize thinking within the subprocess of critiquing a study. This information is summarized in figure 21.

Pattern descriptor	Students whose responses reflected the pattern
Critiques study on the basis of some, but not all, of the following: sample size, data gathering instruments, or data gathering procedure.	Laura, Kristen, Paul, Brooke, Rick
Critiques study on the basis of participants included or instruments used to gather data.	Bill, Jeff, Daniel, Julie, Crystal, Lisa, Hillary, Jessica
Critiques study on the basis of perceived importance of the questions under investigation.	Luke, Nancy

Figure 21. Critiquing a study

In the least sophisticated pattern of response evident within the subprocess of critiquing a study, students simply critiqued the study on the basis of the perceived

importance of the questions under investigation. Luke, for example, went through each individual question that was asked in the study and discussed whether or not it would yield pertinent information about people who lived in Florida. Nancy made a similar critique. Each of the students giving responses reflecting this pattern made critiques similar in nature about each of the questions under investigation. Students giving responses reflecting this pattern made this type of critique consistently across all of the questions.

At the next evident pattern of thinking, students critiqued the study on the basis of one aspect of the study design. Bill, for example, questioned whether or not the person conducting the poll would be able to obtain accurate results from simply asking interview questions of people. Jeff pointed out that the people interviewed during the early morning at the mall would not be the same as those who would shop later on during the day. This pattern of thinking was considered more advanced than the previous one because it actually addressed questions related to study design rather than just perceived importance of the topics of the study. Each of the responses reflecting this pattern included one valid concern about study design.

Students whose responses reflected the most sophisticated pattern identified within the subprocess critiqued the study on the basis of more than one relevant study design aspect. Each of the students giving responses reflecting this pattern critiqued the study on the basis of some, but not all, of the following: participants included, sample size, data gathering instruments, or data gathering procedure. Paul's response is illustrative of this pattern:

His method of collecting is biased to a voluntary response. Those people willing to take the poll, an opinion kind of thing...um...they're gonna take, uh do the interview...kind of...run away from those kind of people at the mall. The data wouldn't be that useful because it is only certain people that are willing to do the survey. Plus, I would think that at a mall, you're going to have more younger people doing this than older people, so it seems a little biased toward age.

His response discussed some, but not nearly all, of the relevant points upon which the study presented in the task could be criticized. Presumably, a student giving a response reflecting a more sophisticated pattern would comment upon most all of the relevant aspects upon which the study could be criticized.

The three patterns of response identified within the subprocess of critiquing a study begin with students paying attention only to the perceived interest of the questions under investigation. While this is certainly a valid dimension along which one can critique any given study, it is far from being all that can be said. At more complex patterns of response, students began to mention aspects of study design that could have been done differently. The second pattern identified was marked by attention to one of the aspects of study design that could be critiqued, while the third pattern identified was marked by critiques of several aspects of study design.

Summary

Several different patterns of response were identified in the analysis of interview transcripts. The patterns range from the very simple to the very sophisticated. Since the patterns of response identified vary in this manner, they provide a portrait of a spectrum

of performance. This spectrum provides a means for identifying various levels of response in the area of high school statistics. In the next main section of this chapter, I will discuss the relationship between the spectrum of patterns of response identified and the cognitive levels described by Biggs and Collis (1982, 1991).

Relationship to the Theoretical Perspective

Overview

The patterns of response to the statistical thinking interview items can be characterized in terms of modes and levels within the Biggs and Collis (1982, 1991) cognitive model. Responses given by students during the clinical interview sessions seemed to be characteristic of two different modes of development: concrete symbolic and formal. Levels of response within each of these two modes were apparent in the patterns identified. The levels can generally be identified with unistructural-multistructural-relational (UMR) cycles. A UMR cycle begins with the unistructural level, at which students focus on one relevant aspect needed to complete a task. At the next level, multistructural, students focus upon several relevant aspects, but they do not integrate them to form a coherent whole. The climax of a UMR cycle is the relational level, in which the relevant aspects are integrated to produce a coherent structure and meaning. Some researchers maintain that there can exist several UMR cycles within any given mode (e.g., Pegg & Davey, 1998; Watson, Collis, Callingham, & Moritz, 1995). For example, one complete UMR cycle in the concrete symbolic mode could be followed by another cycle that involves the consolidation and application of a concept developed in the first cycle.

In this section, I will begin by discussing the characteristics of both the concrete symbolic and formal modes and how I separated concrete symbolic patterns of response from formal ones. I recognize that there is a great deal of room for disagreement about what qualifies a response as characteristic of the formal mode of development as opposed to the concrete symbolic mode. Experienced researchers have been hesitant to place the formal mode label on observed patterns of statistical thinking. For example, Watson and Moritz (2000c) identified two levels of thinking in regard to students' understanding of average outside of a UMR cycle they had identified within the concrete symbolic mode. Rather than making the claim that the levels reflected the formal mode, they simply labeled the two levels "application of average in one complex task" (A1) and "application of average in two complex tasks" (A2). Hence, the following discussion is not meant to be the final word on the characteristics of formal mode statistical thinking. Instead, it is intended to be a catalyst for further discussion of characteristics of statistical thinking within the formal mode of development.

Similarly, my matching of patterns of thinking with levels within modes is not meant to provide the final word on the matter. Instead, I have provided plausible matches between the patterns of thinking observed and the levels within the Biggs and Collis (1982, 1991) model. Data from a larger sample of students could modify some of the descriptors I have formed for the various patterns of thinking observed during this study. This may change the matching of levels to patterns of response. There are also some cases where one might argue that alternative level assignments could be made to patterns of thinking. I have pointed out some of these cases in what follows, but the reader is

encouraged to attempt to identify more such cases in the following discussion. The following discussion should be viewed as the beginning of further discussion in regard to connections between the levels in the theoretical perspective incorporated and observed patterns of response in statistical thinking.

Concrete Symbolic and Formal Modes

The Concrete Symbolic Mode and Statistical Thinking

Biggs and Collis (1991) discussed the nature of the concrete symbolic mode. During this mode, students master the application of second order symbol systems to experiential reality. Tools such as written language, mathematical symbols, and maps can be applied to describe and make sense of concrete referents. In essence, “there is logic and order both between the symbols themselves, and between the symbol system and the world” (Biggs & Collis, 1991, p. 63) for students who function at the highest level within the concrete symbolic mode. While coherence within symbol systems and coherent applications of symbol systems to reality both occur during the concrete symbolic mode, students do not develop the ability to think more abstractly and theoretically until the formal mode of development.

Several of the patterns of thinking displayed by students during the clinical interviews were concrete symbolic in nature. In some of the subprocesses identified, patterns of response gradually became more sophisticated until students viewed data sets as aggregates. Viewing data sets as aggregates is an essential part of mastering the application of statistical symbol systems to reality. It is also an essential part of perceiving statistical data sets themselves as coherent wholes (McClain & Cobb, 2001).

Patterns within other subprocesses climaxed within the concrete symbolic mode with students exhibiting a high degree of competency dealing with statistical graphs and their application to the representation of data sets. In still other subprocesses, patterns of response became more sophisticated until students used and integrated all relevant cues embedded in the tasks in the production of answers. While a high degree of fit was obtained between the response and the task setting in these cases, such responses were still concrete symbolic because they did not transcend the given task setting to incorporate theoretical aspects.

The Formal Mode and Statistical Thinking

The formal mode of development was also described by Biggs and Collis (1991). Thinking in the formal mode incorporates theories and general principles not used within the concrete symbolic mode. Formal mode thinkers are able to use the underlying principles of a discipline in order to solve problems. The ability to reason about abstract underlying principles represents a significant jump from the concrete symbolic mode in the capacity for abstract thought. Pegg and Davey (1998) summarized the Biggs and Collis (1991) conception of the formal mode, saying, “The individual can consider more abstract concepts and work in terms of ‘principles’ and ‘theories.’ The individual is no longer restricted to a concrete referent” (p. 117).

Responses that could be considered characteristic of the formal mode of thinking were given by some students in the present study. Responses characteristic of formal mode thinking were evident in two main ways. The first way was in the application of formal theoretical statistical tools to problem situations. The second way was in the

application of formal statistical principles of data collection and data analysis to problem situations. Of course, there is not always a clear separation between statistical tools and statistical principles. In the following discussion, I have adopted the convention of calling theoretical objects such as best fit lines and measures of center statistical tools. Methods of experimental and non-experimental design and formal theories have been called statistical principles. Responses characteristic of formal mode thinking incorporated the application of either formal statistical tools or statistical principles at some point.

Levels and Modes within Patterns of Response

Describing Data

The three patterns of response observed within describing data can be characterized as a UMR cycle within the concrete symbolic mode that climaxes with a high level of competence in reading statistical graphs (figure 22). The least sophisticated pattern is characteristic of the unistructural level, since the one relevant aspect students incorporate in responses is the conveying of information about one of the variables that is represented by any given data point. The next pattern in the subprocess is multistructural in nature, since the relevant abilities to describe information about both one and two variable points both come into play. At the most sophisticated pattern of response within the subprocess, the ability to describe information contained by points representing both one and two variables is coordinated with a correct description of the display conventions of graphs. Since the various relevant abilities are coordinated in this last response pattern, it is relational in nature. None of the responses given to the describing data tasks were considered examples of formal mode thinking since none of the students went outside the

contexts of the given problems to discuss statistical principles of graph reading in general.

Level and Mode	Describing Data: Describing Statistical Graphs
Relational/Concrete Symbolic (R-CS)	Relates information about the appropriate number of variables when discussing meanings of points on both one and two variable statistical graphs. Correctly relates information about display conventions of a variety of graphs.
Multistructural/Concrete Symbolic (M-CS)	Relates information about the appropriate number of variables when discussing meanings of points on both one and two variable statistical graphs.
Unistructural/Concrete Symbolic (U-CS)	Relates information about only one variable when discussing meanings of points on both one and two variable statistical graphs.

Figure 22. Relationship between the theoretical model and patterns of response for describing data

Organizing and Reducing Data

The patterns of response evident within the process of organizing and reducing data were characteristic of thinking within both the concrete symbolic and formal modes of development. One pattern of response characteristic of a level in the formal mode (FO) was identified in regard to the subprocess of using measures of center. Each of the other patterns of thinking identified within organizing and reducing data were characteristic of levels within the concrete symbolic mode. The relationship between levels and modes and the patterns of thinking identified is summarized in figure 23 and discussed in detail below.

Level and Mode	Subprocess 1: Using Measures of Center	Subprocess 2: Using Measures of Spread	Subprocess 3: Recognizing the Effects of Data Transformation upon Center and Spread	Subprocess 4: Organizing Raw Sets of Data
U-FO	Uses reasonable formal measures to locate centers of data sets.	Not observed.	Not observed.	Not observed.
R-CS	Uses a combination of reasonable formal and visual measures to locate centers of data sets.	Gives verbal descriptions and quantifications of spread that are suitable for given sets of data.	Recognizes the effect of data transformation on both measures with a complete quantitative description of the impact on one measure of center.	Uses both relevant strategies for grouping demonstrated in the previous pattern: (i) Forming groups within an ordering of values of a measurement variable; (ii) Forming groups by using a pre-existing categorical variable within the data set.
M-CS	Uses a combination of formal and visual measures to find centers of data sets, only some of which are reasonable for the given set of data.	Gives verbal descriptions and quantifications of spread. Some descriptions or quantifications are not suitable for given sets of data.	Recognizes the effect of data transformation on both measures without completely quantifying the effect upon center.	Uses the strategy of ordering the values of a measurement variable to organize data. One of the following two grouping strategies is also incorporated: (i) Forming groups of values within the ordering; (ii) Forming groups by using a pre-existing categorical variable within the data set.
U-CS	Uses only visual approaches to find centers of data sets, only some of which are reasonable for the given set of data.	Gives verbal descriptions of spread rather than quantifications.	Recognition of the effect of data transformation on only one of two measures: center or spread.	Uses the strategy of ordering the values of a measurement variable to organize data.

Figure 23. Relationship between the theoretical model and patterns of response for organizing and reducing data

Subprocess 1: Using measures of center. Increasing attention to important task factors marks the differences between the patterns identified for subprocess 1 of organizing and reducing data (using measures of center). The lowest three patterns within the subprocess can be conceptualized as one complete UMR cycle within the concrete symbolic mode. The least sophisticated pattern is unistructural in nature because only one strategy, a primitive visual one that does not necessarily fit the context, is used in order to describe the centers of data sets. The next pattern identified is multistructural in nature because both formal and visual approaches are incorporated. Still, only some of the approaches produce results reasonable for describing the centers of the data sets given. Relational level responses are exemplified by the most sophisticated pattern of thinking, which is characterized by a good fit between the data set and the approaches, either formal or visual, used to describe the centers of the sets of data.

The most sophisticated pattern of response within subprocess 1 of organizing and reducing data (using measures of center) seems to qualify as a level of thinking within the formal mode. Responses reflecting this pattern break completely from the use of visual measures of center in favor of exclusive formal methods that are well-suited to the data at hand. The switch to the exclusive use of formal methods is significant, since the centers of the data sets can then be objectively communicated to others who are familiar with the same statistical tools for finding measures of center. Just one level of thinking was apparent within the formal mode for the subprocess, since just one formal statistical tool for finding center was used in each of the problem situations that were presented. The use of only one formal statistical tool for finding center could indicate that the pattern of

thinking is unistructural in nature. Presumably, higher levels of thinking within the formal mode would include the discussion of several different formal measures of center that could be used in each given situation.

One might also argue that the most sophisticated pattern of thinking identified within the subprocess of using measures of center marks the beginning of another cycle within the concrete symbolic mode. Following this line of argument, one might say that the use of formal measures of center in different contexts does not qualify as a big enough leap in abstraction to place such thinking within the formal mode. It could be argued that the pattern is unistructural in nature, since just one formal measure of center is incorporated for each data set. Hence, others may characterize the most sophisticated pattern of thinking within the subprocess of using measures of center simply as the unistructural level of a new concrete symbolic UMR cycle.

Subprocess 2: Using measures of spread. The three patterns of response for subprocess 2 of organizing and reducing data (using measures of spread) seem to comprise a UMR cycle within the concrete symbolic mode that climaxes with a good fit between the data set and the approaches used to describe its spread. The first pattern of thinking appears to be unistructural, since students use just one primitive strategy to describe spread. Only verbal descriptions are used, and quantifications are notably absent. In the next pattern of response, verbal descriptions and quantifications are incorporated. However, not all of them are well-suited to the data, since some of them do not take clusters within the data set into account. Responses within the pattern therefore could be characterized as multistructural. The most sophisticated pattern of response within the

subprocess is relational, since the quantifications and descriptions for spread are well suited to the data set, and clusters and skewness are taken into account. Taking clusters and skewness into account is seen as indicative of more sophisticated and integrated thinking since such issues must be taken into account when one begins to work exclusively with formal measures of spread. Standard deviation, for example, does not work as well as interquartile range to describe spread for skewed data sets.

Subprocess 3: Data transformation effect on center and spread. Subprocess 3 within organizing and reducing data (data transformation impact on center and spread) contains three patterns of varying degrees of sophistication. The least sophisticated pattern was that at which students recognized the impact of data transformation upon just one of the relevant measures: either the center or the spread. The mastery of just one of the relevant aspects is indicative of unistructural thinking. At the next pattern in the subprocess, students recognize the direction of the impact upon both a measure of center and a measure of spread. However, while the impact is described qualitatively, it is not quantified. Since one more relevant aspect is incorporated, but the important aspect of quantification of impact is not, the second pattern of response could be considered multistructural. Responses reflecting the most sophisticated pattern within the subprocess did bring the important aspect of quantification of impact into play, so they could be considered relational in nature. Hence, the three levels within the subprocess seem to comprise one full UMR cycle in the concrete symbolic mode.

Subprocess 4: Organizing raw sets of data. The three patterns of thinking identified for subprocess 4 of organizing and reducing data (organizing raw sets of data) also seem

to comprise one complete UMR cycle in the concrete symbolic mode. The least sophisticated pattern of response was that at which students only incorporated linear orderings in their organization of data sets. The incorporation of one relevant organization strategy makes this pattern resemble a unistructural level of thought. The next pattern of response was that at which the idea of grouping was incorporated along with linear orderings. However, only one of two grouping strategies suggested by the context of the data set was used. Since the incorporation of one more relevant aspect took place, but an important strategy was left out, the responses reflecting this pattern are characteristic of multistructural thinking. At the most sophisticated pattern of response within the subprocess, students used linear orderings along with both types of relevant grouping strategies suggested by the context of the data set. Since the integration of all of the relevant aspects occurred in this pattern, it is relational in nature.

Representing Data

While the levels of thinking associated with describing data reflect increasing competence in extracting information from graphs, the levels associated with representing data reflect increasing sophistication in constructing graphs to display information. The four levels of response associated with the subprocess do not seem to comprise one UMR cycle, as was the case with describing data. Rather, the lowest level seems to comprise the end of one UMR cycle, and the last three levels seem to comprise a full UMR cycle (figure 24). None of the levels seem to go beyond the concrete symbolic mode, since, as with describing data, students did not go outside of the context given for the tasks to discuss principles of statistical graph construction in general.

Level and Mode	Representing Data
Relational – Concrete symbolic, cycle 2	Conceptualizes multiple valid representations in both univariate and bivariate situations.
Multistructural – Concrete symbolic, cycle 2	Conceptualizes graphs in both univariate and bivariate situations, incorporating multiple representations in some bivariate situations.
Unistructural – Concrete symbolic, cycle 2	Conceptualizes multiple graphs for bivariate situations <u>only</u> or conceptualizes one valid graph for each univariate and bivariate situation.
Relational – Concrete symbolic, cycle 1	Conceptualizes one graph for bivariate situations.

Figure 24. Relationship between the theoretical model and patterns of response for representing data

The least sophisticated pattern of response for representing data was that at which students were able to produce one valid graphical representation for each of the bivariate data sets given. These responses seem to be relational in nature, since presumably levels of response in the cycle leading up to it would involve students being unable to produce graphs for some of the given bivariate situations. The next evident pattern of response is more structurally complex because of the incorporation of one more relevant aspect than what was included in responses at the previous level. Students begin to conceptualize multiple graphs for bivariate situations, or else conceptualize a valid graph in both univariate and bivariate situations. Since this pattern encapsulates the previous one and adds just one more of these relevant aspects, it seems characteristic of unistructural thinking. The third pattern within the subprocess incorporates yet another relevant aspect. Multiple representations for bivariate graphs are incorporated along with the conceptualization of graphs for both univariate and bivariate data sets. Since an

additional relevant aspect is included in this pattern of response, it can be considered multistructural. Finally, at the most sophisticated pattern of response evident, students incorporate multiple representations across both univariate and bivariate situations. Since responses reflecting this pattern integrate all of the relevant aspects introduced in the previous two, the responses are characteristic of the relational level.

Alternatively, one could argue that the two least sophisticated patterns of response could be merged into one. This line of thinking would assert that a student who can construct multiple graphs for a bivariate situation is not exhibiting more structurally complex thinking than one who can construct just one graph for the same situation. Following this line of thinking to its logical conclusion would result in the collapsing of the two lowest levels within the subprocess. This would mean that there would be just one UMR cycle in the concrete symbolic mode evident in the responses. In making the choice to consider the construction of multiple representations indicative of a higher level of thinking, however, I have kept the two least sophisticated patterns of thinking distinct.

Analyzing Data

Some of the patterns of response for the analyzing data tasks appear to be concrete symbolic in nature, while others are characteristic of the formal mode. The level and mode that matches each pattern of response for analyzing data is displayed in figure 25. In figure 25, the subprocesses relating to identifying atypical points (subprocesses 3 and 5) have been merged together. The subprocesses relating to interpolation and extrapolation (subprocesses 6 and 7) have also been merged. The rationale for combining these subprocesses is explained below along with a detailed description of the

connections between all of the patterns of response and the Biggs and Collis (1982, 1991) cognitive model for levels of thinking.

Level and Mode	Subprocess 1: Comparing Univariate Data Sets	Subprocess 2: Analyzing Sample Means	Subprocesses 3 and 5: Identifying Atypical Points
R-FO	Not observed.	Not observed.	Not observed.
M-FO	Not observed.	Not observed.	Not observed.
U-FO	Not observed.	Not observed.	Formal tool used to help identify atypical points.
R-CS	Compares sets of data by using multiple relevant attributes of the aggregates.	Recognizes that the larger the random sample, the more likely its mean will be close to the population mean; and that smaller samples are more likely to produce extreme results.	Distance from main cluster of points or distance from the center of the data set is discussed in the identification of atypical points.
M-CS	Compares sets of data by using one relevant attribute of the aggregate sets such as shape, center, or spread. The comparisons made are often supplemented by point-by-point comparisons or irrelevant features of the data sets.	Recognizes that the larger the random sample, the more likely its mean will be close to the population mean.	Distances between individual data points are discussed in the identification of atypical points.
U-CS	Compares sets of data by using a point-by-point strategy.	Recognizes that different size samples can be drawn from a population, but does not recognize the impact of changing sample size upon the sample mean.	Contextual criteria alone are used to identify atypical points. No quantitative comparisons are made between points in the data set.

Figure 25. Relationship between the theoretical model and patterns of response for analyzing data

Level and Mode	Subprocess 4: Making Multiplicative Comparisons	Subprocesses 6 and 7: Making predictions from Bivariate Measurement Data	Subprocess 8: Analyzing Bivariate Relationships
R-FO	Not observed.	Formally determined best fit line is used in order to make predictions.	Not observed.
M-FO	Not observed.	Visual/algebraic best fit line is used in order to make predictions.	Not observed.
U-FO	Not observed.	Visual line of best fit is used in order to make predictions.	Goes beyond discussing the direction of the quantitative relationship between the two variables displayed. Identifies intervening variable beyond the two variables shown in the data set.
R-CS	Makes only multiplicative comparisons when only multiplicative comparisons are warranted.	Trend/direction of the data is used in order to make predictions.	Describes direction of the relationship between the two variables involved and uses that relationship and quantitative anomalies to argue for or against a causal relationship between the two variables.
M-CS	Makes additive and multiplicative comparisons when only multiplicative comparisons are warranted.	Data in the immediate vicinity are used in order to make predictions.	Describes direction of the relationship between the two variables involved and uses that relationship to help argue for a causal relationship between the two variables.
U-CS	Makes additive comparisons when multiplicative ones are warranted.	Context/personal theories are used in order to make predictions.	Describes direction of the relationship between the two variables involved.

Subprocess 1: Comparing univariate data sets. The patterns of response for subprocess 1 of analyzing data (comparing univariate data sets) can be characterized as one UMR cycle within the concrete symbolic mode. The cycle progressively builds to the point that it becomes evident that students perceive each of the data sets being compared as aggregates. The least sophisticated pattern of response can be considered unistructural because only one primitive strategy is used for comparing the data sets. Point-by-point strategies are the only ones incorporated. In the second pattern, responses incorporate the use of one of the characteristics of data sets such as shape, center, or spread. Hence, these responses show a progression toward the view of the data sets as aggregates. However, the second pattern of response can at best be considered multistructural, since the responses frequently incorporate the use of point-by-point comparisons or irrelevant features of data sets. The third pattern of response for the subprocess is characteristic of the relational level, since only multiple relevant characteristics of the aggregate data sets are taken into account in drawing comparisons.

Subprocess 2: Analyzing sample means. The three patterns of response associated with subprocess 2 of analyzing data (analyzing sample means) can be characterized as a complete UMR cycle within the concrete symbolic mode that is marked by an increasing awareness and integration of contextual factors. In the least sophisticated pattern of response evident, students seem to be aware only of what it means to draw a sample from a population. Since the one relevant idea of sampling is incorporated in responses, the first pattern is unistructural in nature. Another relevant element is incorporated in responses reflecting the next pattern. Students grasp the fact that larger samples are more

likely to produce sample means close to the population mean. However, they do not demonstrate an understanding of the fact that means of smaller samples are more likely to produce extreme results than those of larger samples. The more complex, yet incomplete, nature of responses reflecting this pattern suggests that the pattern is indicative of the multistructural level. An integrated understanding of both facts is demonstrated in responses in the highest pattern of response, so it can be considered indicative of the relational level. Since none of the students interviewed invoked statistical theory to explain the variability of sample means, none of the patterns of response exhibited were characteristic of the formal mode of development.

Subprocesses 3 and 5: Identifying atypical points. Subprocesses 3 and 5 under analyzing data (identifying atypical points) have patterns of response that mirror each other. The first three patterns of response within each subprocess make up a UMR cycle within the concrete symbolic mode that climaxes with the acquisition of the view of a data set as an aggregate. The least sophisticated pattern within each subprocess was that at which students used qualitative criteria alone to identify atypical points, and did not incorporate any quantitative comparisons among data points. The incorporation of the one relevant strategy of incorporating contextual factors makes the response unistructural in nature. In the next pattern within each subprocess, quantitative criteria begin to be incorporated, but the quantitative criteria are primitive in that they generally involve the consideration of jumps between individual data values. Responses reflecting this pattern are multistructural since they are more structurally complex, yet also lack the connected nature relational responses. In the third pattern within each subprocess, a relational level

of response, the distances from central clusters within the aggregate data set are used in order to identify atypical points. While formal methods are not used at the relational level to identify atypical points, the strategy of identifying such points by considering their distances from the center of the data set is precisely the strategy encapsulated in formal procedures for finding outliers.

The most sophisticated patterns of response for subprocesses 3 and 5 (identifying atypical points) appear to fall within the formal mode. In responses reflecting each of the patterns, a formal statistical tool was used in order to search for atypical points within a data set. The student giving the most sophisticated response within subprocess 3 (the tabular situation) examined a calculator-produced boxplot in order to see if there were any outliers present in the given set of data. The students giving the most sophisticated responses within subprocess 5 (the graphical situation) used the criterion of distance from a best fit line to identify atypical points. The use of these formal tools marks a significant leap in abstraction from students at the previous levels, who used the distance from subjectively perceived clusters in order to determine whether or not points were atypically high or low. Since just one formal statistical tool was incorporated in the responses reflecting the most sophisticated pattern of thinking for each subprocess, the pattern of response could be characterized as unistructural in nature within the formal mode. Higher levels of response within the mode might involve either a more detailed discussion of the formal statistical tool used or perhaps the incorporation of several different statistical tools for finding outliers.

Of course, one could argue that the incorporation of a best fit line or a boxplot does not mark a transition to the formal mode. It seems to be clear that the examination or description of a best fit line or a boxplot given along with a problem would not qualify as formal mode thinking. However, in each situation in which these formal statistical tools were used by students, they were suggested by the students themselves. They did not appear within the context of the problem. According to Watson, Collis, Callingham, and Moritz (1995), a defining characteristic of the formal mode is “deductive reasoning with abstract concepts and rules” (p. 249). Best fit lines and boxplots seem to qualify as abstract concepts, and the students introduced these abstract concepts as reasoning tools. Hence, in this subprocess and in some others that follow, the introduction of a best fit line or a boxplot to a problem situation to aid reasoning was considered to mark a transition to the formal mode.

Subprocess 4: Making multiplicative comparisons. The three patterns of response identified for subprocess 4 in analyzing data (making multiplicative comparisons) can be characterized as one complete UMR cycle within the concrete symbolic mode. The least sophisticated pattern of response could be considered unistructural, since additive comparison is the one relevant (although weak) strategy that is incorporated. The second pattern is multistructural in nature. Responses reflecting the pattern incorporate the use of both additive and multiplicative strategies, but no choice is made between the two strategies. It is not acknowledged that multiplicative comparisons are more powerful for the given data set because the number of fish caught by the second fisherman is slightly greater than the number of fish caught by the first. The third pattern within the subprocess

can be considered relational because the responses express a clear preference for multiplicative reasoning in the given situation. Perceiving the data sets as aggregates is essential in choosing to use multiplicative reasoning to compare the numbers of fish caught by the two fishermen.

Subprocesses 6 and 7: Interpolation and extrapolation. The three least sophisticated patterns of response for subprocess 7 can be characterized as one UMR cycle in the concrete symbolic mode that climaxes with a view of the data set as an aggregate. The least sophisticated pattern within the subprocess is unistructural in nature, since the use of non-quantitative contextual criteria is the only relevant aspect of the task that comes into play in the response. The second pattern can be described as multistructural, since quantitative criteria are incorporated, but still only parts of the data set are attended to. The third pattern within the subprocess can be considered relational, since the trend of the aggregate data set is taken into account in making an extrapolation. In this pattern, indicative of the relational level, the relevant quantitative aspects are integrated in order to produce the response.

The two least sophisticated patterns of response within subprocess 6 (interpolation) parallel the multistructural and relational levels within subprocess 7 (extrapolation). The least sophisticated pattern of response within subprocess 6 parallels the multistructural level in subprocess 7, since only part of the data set is taken into consideration in making an interpolation. The second pattern of response within subprocess 6 parallels the relational level within subprocess 7, since the trend of the aggregate data set is taken into account in making a prediction. The two least sophisticated patterns associated with

interpolating from bivariate data, therefore, can be characterized as comprising the last two levels of a UMR cycle within the concrete symbolic mode.

In subprocess 7 of analyzing data (extrapolation), formal mode responses were evident because of the incorporation of a theoretical best fit line. A theoretical best fit line formally encapsulates using the trend of bivariate measurement data in order to make predictions. The last three patterns for subprocess 7 seem to comprise a UMR cycle within the formal mode in which students' use of best fit lines becomes more and more sophisticated. In the first pattern of response, the idea of a best fit line is incorporated to aid extrapolation, but none of the algebraic or other theoretical characteristics of such a line are incorporated. Hence, the response is characteristic of unistructural thinking. (The most sophisticated pattern of response for subprocess 6 also falls here, since it has the same characteristics). In the next pattern of response an algebraic equation was written for a visually placed best fit line. The fact that the best fit line was visually placed makes predictions more subjective than they need to be. Hence, this pattern can be considered multistructural in nature. A more relational response is achieved in the most sophisticated pattern, where a procedure is executed that results in a best fit line that can be objectively replicated by anyone following the same procedure.

Subprocess 8: Analyzing bivariate relationships. The first three patterns of response in subprocess 8 (analyzing bivariate relationships) seem to comprise a UMR cycle within the concrete symbolic mode marked by increasing attention to relevant quantitative factors. The first pattern of response within the subprocess is that at which a simple directional description of the relationship between two variables is given. Since the

simple directional description comprises the entire substance of the responses reflecting this pattern, they are unistructural in nature. Multistructural responses, displayed in the second pattern identified, take some of the relevant features of the data set into account when asked to determine whether or not one variable causes another. They do not, however, take all of the relevant quantitative features of the data set into account. This does not happen until the third pattern of response, which can be characterized as relational because of the attention to all relevant quantitative features that are given in the data set in determining causation.

The most sophisticated pattern of response identified for subprocess 8 (analyzing bivariate relationships) appears to fall within the formal mode of development. The student whose response reflected this pattern applied the statistical principle that intervening, unseen variables can impact the bivariate relationship that is displayed. Since just one possible intervening variable was identified, the response could be characterized as unistructural. Presumably, higher levels of response would involve the identification and synthesis of a number of possible intervening variables.

Collecting Data

Responses characteristic of both formal and concrete symbolic mode thinking were obtained from the items dealing with data collection. Examples of formal mode responses were obtained for each of the subprocesses within collecting data. The alignment between the patterns of thinking observed in each subprocess and the theoretical model for this study are summarized in figure 26.

Level and Mode	Subprocess 1: Designing a non-experimental study	Subprocess 2: Designing an Experimental Study	Subprocess 3: Critiquing a Study
M-FO	Concern for the representativeness of study participants is articulated, and randomization used in concert with stratification in some cases to obtain it.	Recognizes the possibility of conducting an experiment when applicable and discusses two or more formal controls to be put on experiment in order to obtain valid results.	Not observed.
U-FO	Concern for the representativeness of study participants is articulated, and stratified sampling or random sampling is used in some cases to obtain it.	Recognizes the possibility of conducting an experiment when applicable and discusses a control to be put on the experiment in order to obtain valid results.	Not observed.
R-CS	Concern for representativeness of study participants is articulated, but no strategy is described to attempt to achieve it.	Recognizes the possibility for an experiment when applicable.	Not observed.
M-CS	Discusses data gathering instruments or procedures without articulating concern for the representativeness of those to be included in the study.	Uses data gathering method other than an experiment when an experiment would be applicable. No mention of the possibility of using an experimental design.	Critiques study on the basis of some, but not all, of the following: sample size, data gathering instruments, or data gathering procedure.
U-CS	Discusses data gathering instruments or procedures, but primarily relies upon pre-existing studies for information.	Relies solely upon pre-existing studies.	Critiques study on the basis of perceived importance of the questions under investigation.

Figure 26. Relationship between the theoretical model and patterns of response for collecting data

Subprocess 1: Designing a non-experimental study. The first three patterns of response within subprocess 1 of collecting data (designing a non-experimental study) can be characterized as one complete UMR cycle in the concrete symbolic mode. At the least sophisticated pattern of response, students predominantly use just one mode of inquiry in order to answer the questions. Since pre-existing studies alone form the essence of response reflecting this pattern, it has a unistructural nature. In the second pattern of response, more modes of inquiry are mentioned. However, the need for the important component of sample representativeness is not recognized. Since the second pattern is more complex, yet does not integrate an important component, it is multistructural in nature. Finally, in the third pattern within the subprocess, the need for representativeness is articulated. Since this important aspect is finally integrated, the third level can be considered relational in nature.

The last two patterns identified in subprocess 1 for collecting data appear to fall within the formal mode. Students whose responses reflected the last two patterns applied statistical principles that would help in obtaining a representative sample from the population. In the first of the two patterns, students showed the ability to apply just one statistical principle that would help in this way. At some point, they made use of either random sampling or stratified sampling. Since the application of just one of these formal strategies was evident, the responses could be considered unistructural within the formal mode. A more sophisticated pattern of response was also evident, at which students made use of both stratified and random sampling in the attempt to gain representative samples from populations. Since another relevant strategy was incorporated in responses

reflecting this pattern, they could be considered multistructural. Although no higher levels of response were evident, relational level responses might involve more detailed discussions of both stratified and random sampling strategies, perhaps including discussions of the mechanical aspects necessary to carry out each type of sampling strategy.

Subprocess 2: Designing an experimental study. The first three patterns of response within subprocess 2 of collecting data (designing an experimental study) are similar to the first three described for subprocess 1 (designing a non-experimental study). Just as in subprocess 1, the least sophisticated pattern of thinking within subprocess 2 is characterized by the incorporation of just one mode of inquiry, the use of pre-existing studies. Therefore, the pattern is unistructural in nature. The second pattern of response is that in which a mode of inquiry other than a pre-existing study is introduced. However, experimental methods are not mentioned. Since the second pattern is more complex than the first, yet lacks the important aspect of the recognition of the possibility of experimental design, the pattern can be characterized as multistructural. In the third pattern, the important idea of the possibility of experimental design is finally incorporated. The incorporation of this key idea makes the response relational in nature.

Alternatively, one might argue that the two least sophisticated patterns of response for designing an experimental study are both unistructural in nature. One following this line of argument might assert that the use of non-experimental data gathering methods is no more structurally complex than relying upon pre-existing studies. The logical conclusion of this line of argument would be that both of the two least sophisticated

patterns exemplify unistructural thinking within the concrete symbolic mode. Since I consider the use of data gathering methods a more complex pattern of thinking than simply looking up results from studies that have already been conducted, I opted to keep a separation between the complexity of the two patterns of thinking.

Subprocess 2 of collecting data (designing an experimental study) also elicited two patterns of thinking that seem to have their place within the formal mode of development. At the fourth pattern of sophistication identified, students discussed designing an experiment and putting a formal control on it in order to ensure the validity of the results obtained. Since just one formal control was introduced in responses reflecting this pattern, it could be considered unistructural in nature. In the most sophisticated pattern identified, more than one formal control was imposed upon the experimental design. Hence, the pattern of response could be considered multistructural. Relational level responses, although not evident among those given, might include more detailed discussions about the formal controls used and the benefits and drawbacks of each (e.g., ethical issues may be involved in testing a drug on humans or in giving one group a placebo and the other the actual drug).

Subprocess 3: Critiquing a study. The least sophisticated pattern of response for the subprocess of critiquing a study appears to be unistructural in nature. Students whose responses reflected this pattern simply critiqued the study based upon how important they considered its subject matter to be. This is a relevant dimension along which to critique any given study, but there are certainly many other dimensions along which one can critique studies. Students whose responses reflected the next two patterns of thinking

incorporated some of those dimensions in their critiques. In seeking a fit between the theoretical model and the patterns of thinking identified for critiquing a study, I merged the two most complex patterns of thinking observed into one, since both appear to be multistructural in nature. Students in each of the two most complex patterns of thinking critiqued the study according to some, but not all, possible relevant dimensions. None of the responses in the two most complex patterns incorporated critiques of all of the dimensions along which the study could be evaluated. Hence, none of the patterns of response seemed to reflect the coherence characteristic of relational level thinking.

Summary

Several patterns of statistical thinking characteristic of the concrete symbolic mode have been described in this chapter. Students develop the ability to read both one and two variable statistical graphs within this mode. They develop the ability to choose ways to describe center and spread that are suitable for a given data set, and develop the ability to quantify the effect data transformation has upon the center and spread of a data set. Students develop the ability to organize raw sets of data along several relevant dimensions. In regard to representing data, they develop the ability to construct representations for different types of data, and at a higher level of thought construct multiple representations for different types of data. They develop sound conceptions about sampling, while beginning to view data sets as aggregates rather than discrete sets of points. Data collection concepts begin to develop within the mode as well, as students begin to recognize the need to conduct experiments, begin to think about the importance of representative samples, and begin to critique studies along concrete dimensions.

The transition to the formal mode of thinking is marked by significant leaps in abstract and hypothetical thought. Some glimpses of formal mode statistical thinking appear to have been provided by this study. Within the process of organizing and reducing data, some students were able to apply abstract and valid formal methods for finding centers of skewed sets of data across different contexts. Within the process of analyzing data, some students used abstract formal procedures to identify atypical points within different data sets. Some used a hypothetical line of best fit in order to aid in the process of interpolating and extrapolating from bivariate data. Some also were able to hypothesize about unstated intervening variables that would influence the relationship between two variables displayed in a scatterplot. Within the process of collecting data, some were able to make strides towards incorporating relevant formal procedures in the design of surveys and experiments. While glimpses of formal mode statistical thinking were obtained in this study, much work remains for future researchers to continue to identify and characterize statistical thinking within this mode of development. Some directions for further research in this area, as well as some of the other implications of this study, are discussed in the next chapter.

CHAPTER V

REFLECTIONS ON THE STUDY

The goal of the present study was to answer two research questions: (i) What are the defining characteristics of different patterns of high school students' statistical thinking within the processes of describing, organizing and reducing, representing, analyzing, and collecting data? (ii) What levels of statistical thinking can be associated with each of the patterns? In the last chapter, I provided detailed answers to both of these research questions. In this chapter, I will summarize the study, provide a critique of it, and discuss its significance for researchers and teachers.

Summary of the Study

The current study gives a broad picture of high school students' statistical thinking. Taking into account current recommendations for the content of high school statistics (NCTM, 2000; College Board, 2001b; Schaeffer, Watkins, & Landwehr, 1998), an interview protocol was designed to provide a portrait of students' thinking across the processes of describing, organizing and reducing, representing, analyzing, and collecting data. Through the analysis of responses given by 15 students during clinical interview sessions, several patterns of response were identified and described within each of the five statistical thinking processes of interest. The patterns of response identified were then matched with levels of sophistication within the Biggs and Collis (1982, 1991) cognitive model. Taken together, the levels of sophistication identified form a framework

that describes statistical thinking across a variety of processes. In this section, I will summarize the patterns of thinking that were evident within each process, their connections to the theoretical model, and how the findings within each process relate to other statistical thinking research.

Describing Data

Three different patterns of response were identified within responses to tasks concerning the process of describing data. The least sophisticated pattern of response was that in which students related information about only one of the relevant variables in statistical graphs. In the next discernible pattern of response, students related information about the appropriate number of variables in displays. Responses fitting the most sophisticated pattern related information about the appropriate number of variables and included accurate descriptions of display conventions of graphs. As argued in the previous chapter, the three patterns of response seem to comprise one complete unistructural-multistructural-relational (UMR) cycle within the concrete symbolic mode.

The findings about students' thinking within the process of describing data confirm and supplement those of previous research. The fact that several levels of thinking were discerned within the process helps support the contention of Friel, Curcio, and Bright (2001) that making sense of graphs is a complex endeavor rather than a simple one. The study also shows that high school students may experience some of the same difficulties reading graphs as middle school students. According to Mooney (2002), lower levels of thinking for middle school students within describing data are partially marked by an inability to decipher some of the display features of graphs. The findings of the present

study suggest that the same is true for lower levels of high school students' thinking. Additionally, lower levels of thinking for high school students in the present study were marked by lack of attention to one of the variables displayed in some graphs. The lack of attention to variables being displayed in graphs and the inability to interpret display features and conventions may help explain why many students experience difficulty in reading various types of statistical graphs (Zawojewski & Shaughnessy, 2000).

Organizing and Reducing Data

The present study documented students' thinking in regard to using measures of center and spread. The least sophisticated pattern of thinking within each of the subprocesses of using measures of center and using measures of spread was that in which highly subjective and hence sometimes inaccurate descriptions of center and spread were given. In the next pattern of response for each subprocess, students included the use of objective formal tools along with subjective descriptions in describing center and spread. However, some of their descriptions of center and spread were not well suited to the data. In the third pattern of response for each, objective formal tools and subjective descriptions were both still incorporated in describing center and spread, but those tools and descriptions were always well suited to the data. There was then a fourth pattern of thinking only within the subprocess of using measures of center, in which students used only formal methods to describe center, and those formal tools were always well-suited to the data. In the previous chapter, it was argued that the first three patterns of thinking in each of the subprocesses of using measures of center and using measures of spread each comprise a UMR cycle within the concrete symbolic mode. It was also argued that the

fourth pattern of response in regard to using measures of center was characteristic of a unistructural level of thinking within the formal mode.

The findings of the present study describe students' thinking in regard to applying measures of center and spread in different contexts rather than their understandings of the characteristics of specific tools for finding center and spread. Some early research studies in this area had the latter focus, specifically describing students understanding of the arithmetic mean (Pollatsek, Lima, & Well, 1981; Strauss & Bichler, 1988; Cai, 2000) and variance (Mevarech, 1983). The present study gives evidence that the levels of thinking evident among high school students as they apply measures of center and spread to different data sets are similar to those evident among middle school students. Mooney's (2002) study suggests that lower levels of thinking for middle school students are characterized by a poor fit between the methods used to describe center or spread, while higher levels are characterized by a good fit. The same was true for the students interviewed for the present study. Groth's (2002b) study suggests that there are also higher levels of thinking in regard to using measures of center among high school students than those documented during the present study. Groth (2002b) found that some high school students have the ability to draw upon several formal tools in order to describe centers of data sets.

The present study also documented students' thinking about how data transformation effects measures of center and spread. The least sophisticated pattern of thought identified was that in which students demonstrated an understanding of the impact of a transformation on only one of the two measures. In a more sophisticated pattern of

response, students demonstrated an understanding of the impact of the transformation on both measures, but did not quantify it. Students giving responses reflecting the most sophisticated pattern did quantify the impact. In the previous chapter I argued that the three patterns of thinking identified could be characterized as a UMR cycle in the concrete symbolic mode. Even though recognizing the effect of data transformation is an important part of high school statistics (NCTM, 2000), there is not a rich body of research describing students' thinking about it. Hence, the levels of thinking described by the present study help to fill a void in the research literature.

The final subprocess investigated by the present study concerned students' thinking while organizing raw sets of data. The least sophisticated pattern of response to interview tasks was that at which the only strategy students used for organizing data was to group values of a measurement variable from least to greatest. The next pattern of response was marked by the strategy of using either clusters within values of a measurement variable to organize data or else the use of pre-existing categorical variables to group the data. The most sophisticated responses were those which incorporated both strategies for grouping the data. I argued that the three patterns of response identified can be characterized as a UMR cycle in the concrete symbolic mode.

The findings about students' levels of thinking in regard to organizing data supplement previous findings. First of all, the levels of thinking for organizing data identified by the present study somewhat resemble those described by Mooney, Langrall, Hofbauer, & Johnson (2001) for middle school students. In their study and in the present one, progressively higher levels of sophistication are marked by students' clustering of

values within a given variable. The present study also addresses a direction for further research in the area identified by Mooney, Langrall, Hofbauer, and Johnson (2001). In the discussion of their study, they pointed out that they assessed students' thinking in regard to categorical data only, and stated that tasks involving numerical data might reveal additional patterns of thinking. The patterns of thinking identified during the present study detail students' organization of a data set containing both numerical and categorical variables.

Representing Data

The present study also gives insight about high school students' abilities to represent data. Students giving the least sophisticated patterns of response for the process were able to conceptualize just one graphical representation for each given bivariate situation, while those responding at higher levels were able to conceptualize multiple graphs for both univariate and bivariate situations. In the previous chapter, I argued that the patterns of thinking evident for representing data could be characterized as either a complete UMR cycle in the concrete symbolic mode or else the end of one UMR cycle in the concrete symbolic mode and the beginning of another in the same mode.

The representing data portion of the framework details students' abilities to conceptualize data displays for various situations (as in Mevarech & Kramarsky, 1997) rather than their abilities to construct different parts of a given display (as in Padilla, McKenzie, & Shaw, 1986; Berg & Phillips, 1994). This focus fit well within the overall spirit of the framework, since the other parts of the framework also describe students' application of statistical tools to problem situations rather than their understandings of the

components of specific statistical tools. The study reflects one of Mevarech and Kramarsky's (1997) findings in that many students expressed a preference for line graphs to express data whenever possible. However, in the present study, unlike in Mevarech and Kramarsky (1997), all students were prompted to produce multiple graphical representations for a given set of data. When prompted, some students were able to display the data correctly using other types of displays. Hence, the present study illustrates that even though students may initially express a preference for using a certain type of graph to represent data in a given situation, it does not necessarily mean that they are so tied to that representation that they are incapable of conceptualizing alternative valid representations for the given set of data.

Analyzing Data

Of all the statistical thinking processes investigated, the most information was gained about the process of analyzing data. It was found that students exhibiting more sophisticated thinking pay attention to progressively more relevant features when asked to compare univariate data sets. Responses within the subprocess of comparing univariate data sets ranged from those indicative of students who conceptualized data sets as discrete points to those who viewed data sets as aggregates and made comparisons based upon multiple relevant characteristics of the aggregates. I argued in the previous chapter that the three patterns of thinking evident for the subprocess could be characterized as a complete UMR cycle within the concrete symbolic mode.

The levels of thinking that emerged in regard to comparing two data sets indicate that not all students at the high school level naturally see data sets as aggregates when asked

to make comparisons between them. The same seems to be true for students at the middle school level, since some students interviewed by Mooney, Langrall, Hofbauer, and Johnson (2001) did not take all of the given data into account when asked to make comparisons between two sets. The work of McClain and Cobb (2001) shows that properly designed instruction can help middle school students begin to see data sets as aggregates and make comparisons between data sets based upon multiple relevant attributes of the aggregates. Presumably, some of the high school students interviewed for the present study had not experienced instruction which would help them begin to develop this type of thinking.

Insight was also gained about students' understandings about the foundations of formal inference. Three different levels of thinking were evident in the subprocess of analyzing sample means. At the highest level of thinking, students had an intuitive grasp of how sample size impacts the variability of sample means. Students not responding at the highest level demonstrated incomplete conceptions about the relationship between population means and sample means similar to those documented by Konold, Well, Lohmeier, and Pollatsek (1993). For example, students responding at the pattern of thinking I characterized as multistructural understood that the larger the sample that is drawn, the closer its mean is likely to be to the population mean, but did not fully realize that smaller samples were more likely to exhibit greater variability. These types of incomplete understandings impair a student's ability to gain a complete conceptual understanding of the law of large numbers, which is foundational to statistical inference. No students actually invoked a formal statistical principle such as the law of large

numbers during the interview sessions, so the three patterns of response that were identified for the subprocess seem to comprise one UMR cycle within the concrete symbolic mode.

Students' abilities to identify atypical points within a data set were also illuminated as a result of this study. Responses given by students in the present study ranged from incorporating relatively unsophisticated contextual and/or qualitative criteria for the identification of such points to using a formal tool to identify such points quantitatively. At the most sophisticated pattern of thinking observed within the subprocess of identifying atypical points in a tabular display, one student used a boxplot as a tool for identifying atypical points. At the most sophisticated pattern of thinking observed in regard to identifying atypical points in a scatterplot, students used the criterion of distance from a line of best fit in order to classify points as atypical. In the previous chapter, I argued that the first three patterns for each of the subprocesses concerned with identifying atypical points could each be considered a UMR cycle within the concrete symbolic mode, while the fourth pattern observed for each subprocess seemed to indicate formal mode thinking.

Although the identification of atypical points, or "outliers," is considered an important part of high school statistics (NCTM, 2000; College Board, 2001a), up to this point there has not been a great deal of research about it. Friel (1998) has done some work in this area, discussing middle school students' abilities to identify outliers in boxplot displays. However, the present study can be viewed as the beginning of another direction for research about students' thinking while identifying outliers. Unlike Friel's (1998)

study, in this study I did not suggest that students use a specific tool in order to identify outliers. I presented sets of data and asked them to use whatever means they saw fit to identify atypical points within the sets. Some chose to use rather primitive methods, while others incorporated formal statistical tools to aid their thinking.

Students' thinking while making multiplicative comparisons within data presented in the form of a two-way table was also examined. In the least complex pattern of sophistication, students made only additive comparisons when multiplicative ones would have been appropriate. In the next pattern, responses incorporated both additive and multiplicative comparisons without expressing a clear preference for one or the other. The most sophisticated pattern of response was that in which responses clearly expressed a preference for multiplicative comparisons. In the previous chapter I argued that the three patterns of thinking could be characterized as a complete UMR cycle in the concrete symbolic mode.

The findings regarding students' use of multiplicative comparisons within the context of a two-way table point out several issues. First of all, just as students do not automatically see data sets as aggregates, the present study and others (Watson & Moritz, 1999a; McClain & Cobb, 2001; Mooney, Langrall, Hofbauer, & Johnson, 2001) suggest that students also do not automatically make multiplicative comparisons in situations where they are appropriate. Students tend first to make additive comparisons, and then later on shift to making multiplicative ones. The findings of the present study also suggest that analyzing a two-way table is far from a trivial task. Batenero, Estepa, Godino, and Green (1996) illustrated this fact by describing students' cognitive

difficulties coming to grips with issues of association between two variables in reading these types of tables. The present study illustrates another dimension of complexity in reading such tables, specifically, recognizing when it is necessary to make multiplicative rather than additive comparisons among parts of the data displayed in the tables.

A fair amount was learned about students' analysis of bivariate data. Responses classified as characteristic of thinking within the formal mode incorporated best fit lines as tools to identify atypical points within bivariate data, while those in the concrete symbolic mode used more subjective means. The use of a best fit line also seems to separate formal mode thinkers from concrete symbolic thinkers within the subprocesses of interpolating and extrapolating data. Notably, different levels of response were given by students in response to interpolation and extrapolation tasks, even though the two tasks appeared to be quite similar on the surface. Extrapolation tasks seemed to elicit more formal mode thinking than interpolation tasks. Formal mode thinking was also evident when students were asked to analyze bivariate relationships, with one student identifying variables not in the data set which could influence the ones displayed.

Responses characteristic of formal mode thinking seemed to be most abundant in regard to bivariate data analysis. As noted in chapters 2 and 4, little statistical thinking research exists that purports to describe statistical thinking within the formal mode. In the present study, I claim to have encountered a number of examples of formal mode thinking in the area of bivariate data analysis. Hence, the findings of the present study in regard to bivariate data analysis seem to provide a springboard for future researchers concerned with investigating thinking within the formal mode of development. The

findings in regard to bivariate data analysis could be used to provoke the research community to work on further describing and investigating the characteristics of formal mode statistical thinking.

Collecting Data

Information was also obtained about students' data collection abilities. The subprocesses detailed within collecting data also confirmed previous findings that students at higher levels of thinking tend to recognize the importance of representative samples (Watson & Moritz, 2000a), and that students at higher levels of thinking also recognize bias within study design (Watson & Moritz, 2000b). Finally, it was found that within the process of collecting data, some students do not recognize situations in which experiments would be useful, while students at higher levels of thinking recognize when they would be useful, and can begin to describe experimental design in some detail. In the previous chapter, I argued that the ability to incorporate formal principles and methods in applicable situations in collecting data marked the transition from concrete symbolic to formal mode thinking.

Critique of the Study

I noted at the close of chapter 1 that no statistical thinking framework can ever really be considered "complete." There will always be room for further investigation and refinement of any given framework. While this study has helped provide a broad picture of high school students' statistical thinking that organizes and supplements some past statistical thinking research, and hence can be used to advise teaching and research, it can also be critiqued along a number of dimensions. In this section, I will provide my own

critique of the study. Specifically, I discuss the need for further investigation and description of formal mode statistical thinking and critique the interview protocol.

In this study, no hints of thinking within the formal mode of development were evident in students' responses to tasks within the processes of describing data and representing data. It could be that this can be attributed to a "ceiling effect" inherent in some of the tasks in the interview protocol. That is, it may be that the students interviewed felt that concrete symbolic responses were sufficient for the tasks posed, even though they would have given responses characteristic of formal mode thinking if given the appropriate tasks. Future studies need to explore the issue of formal mode thinking within the processes of describing and representing data in more detail. A resource base of tasks that tend to elicit formal mode thinking within these two processes would be of use to teachers and researchers alike.

Another type of task that needs to be formulated is one that tends to elicit thinking about formal statistical inference. Even though hints of formal mode thinking seemed to be evident in responses given to data analysis tasks posed during the study, none of the students at any time used techniques of formal inference to analyze data. This is puzzling in light of the fact that two tasks (task 3 and task 5) were set in contexts designed to elicit thinking about techniques of formal inference. At least some of the students involved in the study had been exposed to the concept of formal inference, since this is a prominent topic on the AP Statistics course syllabus. None of the tasks posed, however, seemed to connect with their background knowledge of formal inference. It is an open question

whether or not tasks posed in a different manner would have prompted the students to think about the use of techniques of formal inference.

It is also an open question in some cases how the levels of thinking observed in this study were tied to the tasks used to assess a subprocess. For example, some students' inability to conceptualize a graph for the univariate situation (graphing ice cream consumption in task 4.8) in the interview protocol may come from the fact that the variable of interest was originally part of a bivariate data set (ice cream consumption vs. temperature). It is possible that students may have had more success with a variable not originally included in a bivariate set. The reason why all students were able to produce graphs for the bivariate situations may be because the variables for those tasks were not originally part of another set of data.

Similarly, students' thinking while analyzing data may have been tied tightly to some of the task contexts. This is especially problematic for the process of analyzing data, since students' thinking within some of the subprocesses was evaluated by only one interview question. For example, in regard to analyzing bivariate relationships, it is an open question whether or not any students would have discussed causality if not explicitly asked to do so in task 4.10 (students were asked if they thought higher temperatures caused ice cream consumption to increase). Also, in regard to analyzing sample means, it is possible that a task other than the one used (2.8, involving the analysis of two different size randomly drawn samples of test scores) would have been more effective in eliciting students' intuitive notions about sampling distributions. Using

only one task to assess students' thinking within a subprocess limits the information that can be gained about the subprocess.

Finally, it should be noted that some of the responses given by students who participated in the current study do not necessarily reflect their true levels of understanding. It is possible that tasks set in different contexts would have elicited different responses in some instances. Gathering individual students' responses to several different tasks would allow one to more accurately approximate their true levels of understanding. The current study merely provides the basis of a framework for further investigations of students' thinking.

The current study can be critiqued along several dimensions, some of which I have touched upon in this section. Some of the interview tasks may have had a "ceiling" which limited the highest level of response that could be given by students. While it appears as if thinking that could be considered characteristic of the formal mode of development was evident in the study, there is still a great amount of work left to be done in further describing and investigating this mode. More data need to be gathered in order to continue to describe the more sophisticated levels of students' thinking. This data could come from both interviewing a larger sample of students and from adding interview questions to assess some of the subprocesses that were assessed by only one task in the current study.

Implications

The current study has implications for professionals concerned with improving high school statistics education. The study has identified several starting points for further

research. It has also served to identify important issues for teachers. In this section, I will summarize some of the ways in which the study serves to advise the practice of both researchers and teachers.

Implications for Researchers

I contend that the study serves to provide glimpses at levels of formal mode thinking within the context of the application of the Biggs and Collis (1982, 1991) model to the discipline of statistics. Some formal mode thinking was evident within each of the processes of organizing and reducing, analyzing, and collecting data. As discussed in chapter 2, formal mode thinking had previously been largely untouched by empirical research, since most studies using the Biggs and Collis model in the area of statistical thinking had focused upon the concrete symbolic and ikonic modes. These glimpses of formal mode thinking should prove useful as researchers continue to explore thinking within the mode. For researchers who do not completely agree with the manner in which I placed some of the patterns of response with the formal mode, the study can serve as a catalyst for further investigation of thinking within the mode.

Students were asked to deal with a number of different contexts for each of the processes investigated in this study. Cobb and McClain (2001) and Jones et al. (2001) are among those who have noted that context plays a great role in determining students' levels of response to statistical thinking tasks. An interesting issue remaining for investigation, therefore, is how different contexts influence the responses that are given to tasks. The goal is not to somehow eliminate all context confusion issues from statistical thinking tasks, since interpreting a context is an essential statistical thinking skill. No

authentic statistics problems are context free. Instead, I see the goal in this area as obtaining a greater understanding of how different types of contexts elicit different types of response from the same individual. This type of investigation might help researchers tailor tasks to individuals in order to obtain a more complete picture of how the individual they are interviewing thinks within different statistical processes.

Another direction for further investigation in the area of high school statistical thinking would be to add to the framework formed as a result of this study by investigating statistical subprocesses left out of this framework. One subprocess that would warrant investigation under the broad process of analyzing data would be using simulations to answer statistical questions. Even though some of the tasks included in the interview protocol for this study attempted to elicit students' thinking about the use of Monte Carlo simulations to answer statistical questions (5.3, 5.4, 5.5), no students used simulations to complete them. Therefore, there seem to be at least two directions for further research within subprocesses involving simulation usage. One direction would be to devise questions that more effectively elicit students' thinking about simulations and use them to elicit levels of response that could be merged with the framework formulated by the present study. Recent research on high school students' understanding of simulations (Zimmerman, 2002) could help inform such a line of inquiry. A second direction for inquiry would be to take tasks 5.3, 5.4, and 5.5 from this study and administer them to a student population more likely to use simulations to answer the questions. Such a population may be found in classrooms where teachers frequently encourage students to construct and carry out simulations. In any event, a description of

students' levels of thinking in regard to constructing simulations would be a valuable addition to the overall framework, since simulations provide a conceptual bridge between the fields of probability and statistics.

The framework constructed as a result of this study may also be profitably used by researchers in order to inform teaching experiments. Jones et al. (2001) illustrated how a statistical thinking framework grounded in Biggs and Collis' (1982, 1991) cognitive theory could be used for such a purpose. The group of researchers used the elementary school statistical thinking framework (Jones et al., 2000) to form the hypothetical learning trajectory at the beginning of the teaching experiment. The hypothetical learning trajectory was essentially a prediction about the levels of thinking through which students would progress as they developed their understanding of statistics. By the end of the teaching experiment, the researchers had succeeded in helping almost all of the students in the study progress beyond the lowest level of thinking in the framework. This result suggests that a teaching experiment advised by a research-based framework may also prove to be profitable at the high school level.

As another thought for further research directions, studies administering the interview protocol constructed for this study to a larger population could be carried out. This may result in additions to the framework and refinements of descriptors for levels of thinking. If the interview script were administered to a larger population, it may elicit levels of thinking not displayed by the students who participated in this study. If new levels of thinking were displayed in response to some of the tasks, they could be merged into the overall framework constructed as a result of this study. This could be a promising

direction for further research, since it might help complete some of the incomplete cycles of thinking observed during this study. The completion of some of the cycles might also result in the refinement of some of the descriptors formulated during this study, since some descriptors were formed without the benefit of being able to observe a complete cycle of thinking within a given mode.

Finally, it is imperative that the research community concerned with investigating high school students' statistical thinking continue to work on describing the characteristics of formal mode statistical thinking. I have forwarded some preliminary thoughts on the characteristics of formal mode statistical thinking in the present study. I have also pointed out that the patterns of thinking I characterized as examples of formal mode statistical thinking may be characterized by others as beginning levels of new cycles still within the concrete symbolic mode. Still others may avoid putting labels associated with modes on such patterns of thinking (there is precedent for this approach, as Watson and Moritz (1999c) chose to call two levels of thinking outside of a cycle in the concrete symbolic mode simply "application of average in one complex task" and "application of average in two complex tasks"). In order to make the Biggs and Collis (1982, 1991) model more useful for investigating advanced levels of statistical thinking, the nature of statistical thinking within the formal mode needs to be further investigated and clarified.

Implications for Teachers

The interview protocol used in this study could be used by practicing teachers in order to help assess students' levels of statistical thinking. Teachers could administer all

or parts of the protocol to students and then compare the responses received to the framework constructed as a result of this study in order to assess students. Teachers with knowledge of where a student's thinking fits in with other levels of thinking observed in this study would then have an idea of the levels of thinking through which the student may need to progress in order to attain the desired level. Knowledge of students' levels of thinking can be used in order to design appropriate and effective instruction (Fennema & Franke, 1992).

It was also observed during this study that age and experience in school are not always accurate predictors of the levels of sophistication at which students will approach statistical thinking tasks. In some cases, students who had not taken a statistics course in high school responded to interview tasks at a higher level of sophistication than those students who had experienced an entire high school statistics course. Likewise, the younger students in the study were able to respond to tasks at higher levels of sophistication than their older counterparts in some instances. While experience may be necessary for progressing to increasingly more sophisticated levels of thinking, it does not automatically lead to more sophisticated levels of thinking. Teachers need to bear this in mind as they design instruction.

The study also alerts teachers that students who have experienced one semester, or even one year, of high school statistics do not automatically construct powerful and versatile meanings for statistical tools and statistical principles. This is evidenced in the failure of the former AP Statistics students interviewed to consider the use of tools and principles of formal inference, even when they had the opportunity to do so. These same

students also failed to incorporate formal tools such as best fit lines in some cases where they would have been applicable. They also did not invoke principles of experimental and survey design in some instances where they would have been helpful. Teachers need to be aware that students who successfully pass a high school statistics course do not necessarily apply formal statistical principles and statistical tools in cases where they would be useful.

Finally, teachers need to note that students who have had exposure to software that can be used for the purpose of statistical data analysis do not automatically use that software in situations where it would be appropriate. Several of the students I interviewed for the present study indicated that they were familiar with graphing calculators and computer software programs for statistical data analysis. However, just as in Biehler's (2001) study, few of the students I interviewed made good use of the available computer software and graphing calculator to execute statistical data analysis. Teachers need to be aware that time spent instructing students on the use of technology for statistical data analysis does not automatically translate to the use of the technology outside of classroom tasks.

Closing Remarks

This study has the potential to contribute to the field of statistics education in several ways. The first way is by informing the practices of researchers and teachers. As a result of this study, there is now more information available focusing upon high school students' statistical thinking. This is important to researchers seeking to compare and contrast statistical thinking at the high school level to statistical thinking at other levels,

and to those wishing to further investigate high school students' statistical thinking. The study is also important to high school teachers because it provides a base for assessing students' statistical thinking within several subprocesses, and hence also a starting point for effective statistics instructional design. Another way that this study contributes to the field of statistics education is that it has raised a number of questions for further investigation. Much work remains to be done in investigating statistical thinking within the formal mode and in adding to and refining the framework constructed as a result of this study. By providing the answers to some questions and raising important new questions in several areas, it is hoped that this study will be a catalyst for both researchers and teachers to continue to do substantial and productive work in the area of high school statistics.

REFERENCES

- Alper, L., Fraser, S., Fendel, D., & Resek, D. (1998). Interactive mathematics program. Emeryville, CA: Key Curriculum Press.
- 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.
- Australian Education Council. (1994). Mathematics: A curriculum profile for Australian Schools. Carlton, VIC: Curriculum Corporation.
- Batanero, C., Estepa, A., Godino, J., & Green, D.R. (1996). Intuitive strategies and preconceptions about association in contingency tables. Journal for Research in Mathematics Education, 27, 151-169.
- Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. Mathematical Thinking and Learning, 2, 127-155.
- Ben-Zvi, D. & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. Educational Studies in Mathematics, 45, 35-65.

- Berg, C.A., & Phillips, D.G. (1994). An investigation of the relationship between logical thinking structures and the ability to construct and interpret line graphs. Journal of Research in Science Teaching, 31, 323-344.
- Biehler, R. (1997). Software for learning and doing statistics. International Statistical Review, 65, 167-189.
- Biehler, R. (2001, August). Students' difficulties in practicing computer-supported data analysis: Some hypothetical generalizations from results of two exploratory studies. Paper presented at The Second International Research Forum on Statistical Thinking, Reasoning, and Literacy. Armidale, Australia.
- Biggs, J.B. (1999). Teaching for quality learning at university. Philadelphia: Open University Press.
- Biggs, J.B., & Collis, K.F. (1982). Evaluating the quality of learning: The SOLO taxonomy. New York: Academic.
- Biggs, J.B. & Collis, K.F. (1991). Multimodal learning and quality of intelligent behavior. In H.A.H. Rowe (Ed.), Intelligence: Reconceptualization and measurement (pp. 57-66). Hillsdale, NJ: Erlbaum.
- Blackburn, J.A., & Paplia, D.E. (1992). Adult cognition from a Piagetian perspective. In R.J. Sternberg & C.A. Berg (Eds.), Intellectual development (pp. 141-160). New York: Cambridge University Press.
- Bogdan, R.C., & Biklen, S.K. (1992). Qualitative research for education: An introduction to theory and methods. Boston: Allyn and Bacon.

- Bright, G.W. & Friel, S.N. (1998). Graphical representations: Helping students interpret data. In S.P. Lajoie (Ed.), Reflections on statistics: Learning, teaching, and assessment in grades K-12 (pp. 63-88). Mahwah, NJ: Erlbaum.
- Burger, W.F., & Shaughnessy, J.M. (1986). Characterizing the van Hiele levels of development in geometry. Journal for Research in Mathematics Education, 17, 31-48.
- Cai, J. (2000). Understanding and representing the arithmetic averaging algorithm: An analysis and comparison of US and Chinese students' responses. International Journal of Mathematical Education in Science and Technology, 31, 839-855.
- Chance, B., Garfield, J., & delMas, R. (2001, August). Developing simulation activities to improve students' statistical reasoning. Paper presented at The Second International Research Forum on Statistical Thinking, Reasoning, and Literacy. Armidale, Australia.
- Chazan, D. (2000). Beyond formulas in mathematics and teaching. New York: Teacher's College Press.
- Cobb, G., & Moore, D. (1997). Mathematics, statistics, and teaching. American Mathematical Monthly, 104, 801-823.
- Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. Mathematical Thinking and Learning, 1, 5-43.
- Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., & Perlwitz, M., (1991). Assessment of a problem-centered second-grade mathematics project. Journal for Research in Mathematics Education, 22, 3-29.

College Board (2001a, November 1). AP Statistics. Retrieved November 1, 2001 from

<http://www.collegeboard.org/ap/statistics/html/grade01.html>

College Board (2001b). Course description: AP Statistics. New York: College Board.

College Board (2002, May 4). AP Statistics 2001 solutions and scoring guidelines.

Retrieved May 4, 2002 from

http://apcentral.collegeboard.com/repository/sg_statistics_01.pdf

Collins, W. (1998). Glencoe algebra 1 : Integration, applications, connections. New

York: Glencoe/McGraw-Hill.

Confrey, J., & Makar, K. (2001, August). Secondary Teachers' Inquiry into Data. Paper

presented at The Second International Research Forum on Statistical Thinking,

Reasoning, and Literacy. Armidale, Australia.

Curcio, F.R. (1987). Comprehension of mathematical relationships expressed in graphs.

Journal for Research in Mathematics Education, 18, 382-393.

Derry, S.J., Levin, J.R., Osana, H.P., Jones, M.S., & Peterson, M. (2000). Fostering

students' statistical and scientific thinking: Lessons learned from an innovative

college course. American Educational Research Journal, 37, 747-773.

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, NY: Macmillan.

Ferrini-Mundy, J., & Gaudard, M. (1992). Secondary school calculus: Preparation or

pitfall in the study of college calculus? Journal for Research in Mathematics

Education, 23, 56-71.

- Fontana, A., & Frey, J.H. (2000). The interview: From structured questions to negotiated text. In N.K. Denzin & Y.S. Lincoln. (Eds.), Handbook of qualitative research (2nd ed., pp. 645-672). Thousand Oaks, CA: Sage.
- Friel, S.N. (1998). Comparing data sets: How do students interpret information displayed using box plots? In S.B. Berenson, K.R. Dawkins, M. Blanton, W.N. Coulombe, J. Kolb, K. Norwood, & L. Stiff (Eds.), Proceedings of the Twentieth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education: Volume 1 (pp. 365-370). Raleigh, NC.
- 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, 124-158.
- Fuys, D., Geddes, D., & Tischler, R. (1988). Journal for Research in Mathematics Education Monograph Number 3: The van Hiele levels of thinking in geometry among adolescents. Reston, VA: National Council of Teachers of Mathematics.
- Gal, I., & Garfield, J. (1997). The assessment challenge in statistics education. Amsterdam: IOS Press.
- Garfield, J., & Chance, B. (2000). Assessment in statistics education: Issues and challenges. Mathematical Thinking and Learning, 2, 99-125.
- Goldin, G.A. (2000). A scientific perspective on structured, task-based interviews in mathematics education research. In A.E. Kelly & R.A. Lesh (Eds.), Handbook of research design in mathematics and science education (pp. 517-546). Mahwah, NJ: Erlbaum.

- Groth, R.E. (2002a). Construction of Thought-Eliciting Statistical Tasks for High School Students. Unpublished manuscript, Illinois State University.
- Groth, R.E. (2002b). Characterizing secondary students' understanding of measures of central tendency and variation. In D.S. Mewborn, P. Sztajn, D.Y. White, H.G. Wiegel, R.L. Bryant, & K. Nooney (Eds.), Proceedings of the twenty-fourth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education: Volume 1 (pp. 247-257). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.
- Heaton, R.M., & Mickelson, W.T. (2002). The learning and teaching of statistical investigation in teaching and teacher education. Journal of Mathematics Teacher Education, 5, 35-59.
- Heid, M.K. (1988). Resequencing skills and concepts in applied calculus using the computer as a tool. Journal for Research in Mathematics Education, 19, 3-25.
- Hirsch, C.R., Coxford, A.F., Fey, J.T., Schoen, H.L. (1998). Contemporary Mathematics in Context: A unified approach. New York: Glencoe/McGraw-Hill.
- Jones, G.A., Langrall, C.W., Thornton, C.A., Mooney, E.S., Wares, A., Jones, M.R., Perry, B.P., Putt, I.J., & Nisbet, S. (2001). Using students' statistical thinking to inform instruction. Journal of Mathematical Behavior, 20, 109-144.
- Jones, G.A., Thornton, C.A., Langrall, C.W., Mooney, E.S., Perry, B., & Putt, I.J. (2000). A framework for characterizing children's statistical thinking. Mathematical Thinking and Learning, 2, 269-307.

- Konold, C., Lohmeier, J., Pollatsek, K., Well, A., Falk, R., & Lipson, A. (1991). Novice views on randomness. In R.G. Underhill (Ed.), Proceedings of the thirteenth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education: Volume 1 (pp. 167-173). Blacksburg, VA.
- Konold, C., Well, A., Lohmeier, J., & Pollatsek, A. (1993, September). Understanding the law of large numbers. In J.R. Becker & B.J. Pence (Eds.), Proceedings of the fifteenth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education: Volume 2 (pp. 299-305). Pacific Grove, CA.
- Lajoie, S.P., & Romberg, T.A. (1998). Identifying an agenda for statistics instruction and assessment. In S.P. Lajoie (Ed.), Reflections on statistics: Learning, teaching, and assessment in grades K-12 (pp. xi-xxi). Mahwah, NJ: Erlbaum.
- Lester, F., & William, D. (2002). On the purpose of mathematics education research: Making productive contributions to policy and practice. In L.D. English (Ed.), Handbook of international research in mathematics education (pp. 489-506). Mahwah, NJ: Erlbaum.
- Maykut, P., & Morehouse, R. (1994). Beginning qualitative research: A philosophic and practical guide. London: The Falmer Press.
- McClain, K., & Cobb, P. (2001). Supporting students' ability to reason about data. Educational Studies in Mathematics, 45, 103-129.
- Merriam, S.B. (1988). Case study research in education: A qualitative approach. San Francisco: Jossey-Bass.

- Mevarech, Z.R. (1983). A deep structure model of students' statistical misconceptions. Educational Studies in Mathematics, *14*, 415-429.
- Mevarech, Z.R., & Kramarsky, B. (1997). From verbal descriptions to graphic representations: Stability and change in students' alternative conceptions. Educational Studies in Mathematics, *32*, 229-263.
- Miles, M.B., & Huberman, A.M. (1994). Qualitative data analysis. Thousand Oaks, CA: Sage.
- Mokros, J., & Russell, S.J. (1995). Children's concepts of average and representativeness. Journal for Research in Mathematics Education, *26*, 20-39.
- Mooney, E.S. (2002). A framework for characterizing middle school students' statistical thinking. Mathematical Thinking and Learning, *4*, 23-64.
- Mooney, E.S., Langrall, C.W., Hofbauer, P.S., & Johnson, Y.A. (2001, October). Refining a framework on middle school students' statistical thinking. Paper presented at the twenty-third annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Snowbird, UT.
- Moore, D. (1990). Uncertainty. In L. Steen (Ed.), On the shoulders of giants: New approaches to numeracy (pp. 95-137). Washington, D.C.: National Academy Press.
- Morse, J.M. (1991). Approaches to qualitative-quantitative methodological triangulation. Nursing Research, *40*, 120-123.

- National Council of Teachers of Mathematics (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.
- National Council of Teachers of Mathematics (2000). Principles and standards for school mathematics. Reston, VA: Author.
- National Research Council (2001). Adding it up: Helping children learn mathematics. Washington, D.C.: National Academy Press.
- Orton, A. (1983). Students' understanding of integration. Educational Studies in Mathematics, 14, 1-18.
- Padilla, M.J., McKenzie, D.L., & Shaw, E.L. (1986). An examination of the line graphing ability of students in grades seven through twelve. School Science and Mathematics, 86, 20-26.
- Patton, M. (1990). Qualitative evaluation and research methods (2nd ed.). Newbury Park, CA: Sage.
- Pegg, J., & Davey, G. (1998). Interpreting student understanding of geometry: A synthesis of two models. In R. Lehrer & D. Chazan (Eds.), Designing learning environments for developing understanding of geometry and space (pp. 109-135). Mahwah, NJ: Lawrence Erlbaum Associates.
- Piaget, J. (1983). Piaget's theory. In P. Mussen (Ed.) Handbook of child psychology (pp. 103-128). New York: John Wiley & Sons.
- Pollatsek, A., Lima, S., & Well, A.D. (1981). Concept or computation: Students' understanding of the mean. Educational Studies in Mathematics, 12, 191-204.

- Reading, C., & Pegg, J. (1996). Exploring understanding of data reduction. In L. Puig & A. Gutierrez (Eds.), Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education: Volume 4 (pp. 187-194). Spain: Universitat de Valencia.
- Reber, A.S. (1995). Dictionary of psychology. London: Penguin Books Limited.
- Resnick, L.B. (1983). Toward a cognitive theory of instruction. In S.G. Paris, G.M. Olson, & W.H. Stevenson (Eds.), Learning and motivation in the classroom (pp. 5-38). Hillsdale, NJ: Erlbaum.
- Saldanha, L.A., & Thompson, P.W. (2001, August). Students' reasoning about sampling distributions and statistical inference. Paper presented at The Second International Research Forum on Statistical Thinking, Reasoning, and Literacy. Armidale, Australia.
- Schaeffer, 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: Learning, teaching, and assessment in grades K-12 (pp. 3-31). Mahwah, NJ: Erlbaum.
- School Curriculum and Assessment Authority & Curriculum and Assessment Authority for Wales. (1996). A guide to the national curriculum. London: School Curriculum and Assessment Authority.
- Sfard, A., & Linchevski, L. (1994). The gains and pitfalls of reification - The case of algebra. Educational Studies in Mathematics, 26, 191-228.

- Shaugnessy, 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 (Part 1, pp. 205-237). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Shaugnessy, J.M., Watson, J., Moritz, J., & Reading, C. (1999, April). School mathematics students' acknowledgement of statistical variation. In C. Maher (Chair) There's more to life than centers. Symposium conducted at the Research Pre-session of the 77th Annual National Council of Teachers of Mathematics Conference, San Francisco.
- Shaugnessy, J.M., & Zawojewski, J.W. (1999). Secondary students' performance on data and chance on the 1996 NAEP. Mathematics Teacher, 92, 713-718.
- Smith, J.K., & Deemer, D.K. (2000). The problem of criteria in the age of relativism. In N.K. Denzin & Y.S. Lincoln. (Eds.), Handbook of qualitative research (2nd ed., pp. 877-896). Thousand Oaks, CA: Sage.
- Sowder, J.T. (2000). Editorial. Journal for Research in Mathematics Education, 31, 1-4.
- Strauss, S., & Bichler, E. (1988). The development of children's concepts of the arithmetic average. Journal for Research in Mathematics Education, 19, 64-80.
- Strauss, A.L., & Corbin, J. (1990). Basics of qualitative research: Grounded theory procedures and techniques. Newbury Park, CA: Sage.
- Tall, D. (1992). The transition to advanced mathematical thinking: Functions, limits, infinity, and proof. In D.A. Grouws (Ed.), Handbook of research on mathematics teaching and learning (pp. 495-511). New York: Macmillan.

- Thompson, P.W. (1994). Images of rate and operational understanding of the fundamental theorem of calculus. Educational Studies in Mathematics, 26, 229-274.
- Torok, R., & Watson, J.M. (2000). Development of the concept of statistical variation: A pilot study. Mathematics Education Research Journal, 12, 147-169.
- Usiskin, Z. (1982). Van Hiele levels and achievement in secondary school geometry. (Final report of the Cognitive Development and Achievement in Secondary School Geometry Project.) Chicago: University of Chicago. (ERIC Document Reproduction Service No. ED 220 288).
- van Hiele, P.M. (1986). Structure and insight. New York: Academic Press.
- Vinner, S. (1985). Concept definition, concept image, and the notion of the function concept. International Journal of Mathematical Education in Science and Technology, 3, 114-162.
- Wainer, H. (1992). Understanding graphs and tables. Educational Researcher, 21 (1), 14-23.
- Wallman, K.K. (1993). Enhancing statistical literacy: Enriching our society. Journal of the American Statistical Association, 88, (421), 1-8.
- Watson, J., & Collis, K. (1994). Multimodal functioning in understanding chance and data concepts. In J.P. da Ponte & J.F. Matos (Eds.), Proceedings of the 18th Conference of the International Group for the Psychology of Mathematics Education: Volume 4 (pp. 369-376). Lisbon, Portugal.

- Watson, J.M., Collis, K.F., Callingham, R.A., & Moritz, J.B. (1995). A model for assessing higher order thinking in statistics. Educational Research and Evaluation, 1, 247-275.
- Watson, J.M., & Moritz, J.B. (1999a). The beginning of statistical inference: Comparing two data sets. Educational Studies in Mathematics, 37, 145-168.
- Watson, J.M., & Moritz, J.B. (1999b). The development of concepts of average. Focus on Learning Problems in Mathematics, 21, 15-39.
- Watson, J.M., & Moritz, J.B. (2000a). Developing concepts of sampling. Journal for Research in Mathematics Education, 31, 44-70.
- Watson, J.M., & Moritz, J.B. (2000b). Development of sampling for statistical literacy. Journal of Mathematical Behavior, 19, 109-136.
- Watson, J.M., & Moritz, J.B. (2000c). The longitudinal development of understanding of average. Mathematical Thinking and Learning, 2, 11-50.
- Well, A.D., Pollatsek, A., Boyce, S.J. (1990). Understanding the effects of sample size on the variability of the mean. Organizational Behavior and Decision Processes, 47, 289-312.
- Wild, C.J., & Pfannkuch, M. (1999). Statistical thinking in empirical inquiry. International Statistical Review, 67, 223-265.
- Woolfolk, A.E. (1993). Educational psychology (5th ed). Boston: Allyn & Bacon.
- Yates, D.S., Moore, D.S., & McCabe, G.P. (1998). The practice of statistics: TI-83 graphing calculator enhanced. New York: W.H. Freeman.

Zawojewski, J.S., & Shaughnessy, J.S. (2000). Data and chance. In E.A. Silver & P.A.

Kenney (Eds.), Results from the seventh mathematics assessment of the national assessment of educational progress. (pp. 235-268). Reston, VA: National Council of Teachers of Mathematics.

Zimmerman, G.M. (2002). Students' reasoning about probability simulations during instruction. Unpublished doctoral dissertation, Illinois State University.

APPENDIX A
BACKGROUND DESCRIPTIONS AND ANECDOTAL INFORMATION FOR
STUDY PARTICIPANTS

Category 1: Students who had recently completed a semester-long high school statistics
course

Lisa, one of the college freshmen interviewed, had completed a semester-long probability and statistics course during her junior year in high school. She indicated that she was not familiar with any technology for doing statistics beyond a simple pocket calculator. She ranked her own statistical ability as a 3 out of 5, stating, “I took the course two years ago. It was a semester course but I feel I still remember most concepts.” At the time she was interviewed, Lisa was engaged in the first month of study of a semester-long data and chance course, which she described as a “math course for art majors.” Topics included in the description for the course are data representations, curve fitting, interpretation of polls and experiments, central tendency, statistical reasoning, and applications of probability.

Kristen, another of the participants in this category, was also in her first month of study in the same semester-long data and chance course. She had taken a semester-long probability and statistics course during her senior year in high school. She indicated that the high school course had incorporated the use of the TI-83 calculator and the Excel software program. She ranked her own statistical ability as 3 out of 5, stating, “Math is

not my best subject. However, since statistics are different I understand it more. Statistics make a little more sense to me than [sic] regular math courses.”

The last of the interview participants in this category was Laura. Laura had taken a one-semester probability and statistics course during her senior year in high school. She indicated that she was familiar with the TI-83 calculator. She ranked her own statistical ability as 2 out of 5, stating, “I forgot a lot of the information taught. I remember some of the important things.” At the time Laura was interviewed, she was engaged in her first month of study in a pre-calculus course. Topics in the description for the course included polynomial, absolute value, rational, exponential, logarithmic, and trigonometric functions and their graphs; properties of trigonometric and inverse trigonometric functions and their applications; conics, translation, and rotation of axes.

Category 2: Students who had recently completed a year-long high school statistics course

Jeff was a college freshmen who agreed to be interviewed for the study. He had completed an AP Statistics course during his senior year in high school. He indicated that the technology used during that course included TI-83 and TI-86 calculators, the Fathom software program, and the Minitab software program. When asked to rank his own statistical ability, he gave himself a ranking of 4 out of 5, stating, “I feel I know a lot about statistics, but I’m sure there’s more to know.” At the time he was interviewed, he was in the first month of a semester-long finite mathematics course at the university. The topics included in the course description were linear functions, matrices, systems of linear equations, sets and counting, probability, statistics, and mathematics of finance.

Hillary, another of the interview participants, was a college freshman enrolled in the same finite mathematics class as Jeff. She also had completed an AP Statistics course during her senior year in high school. During that course, she indicated that the instructor had incorporated the use of the TI-83 calculator, the Excel software program, and the Minitab software program. When asked to rank her own statistical ability, she gave herself a 4 out of 5. She stated that she had consistently received a score of 3 out of 5 on AP Statistics practice examinations she had taken the previous year. At the time she was interviewed, Hillary was also engaged in the first month of study in the finite mathematics course.

Another interview participant in this category, Jonathon, was involved in his first month of study in the same course as Kristen and Lisa. He had completed an AP Statistics course during his senior year in high school. He indicated that the TI-83 calculator had been used in his high school course. He ranked his own statistical ability as 4 out of 5, stating, "I have taken a few statistics courses and I feel that I understand its many concepts and can put them into practice."

The last of the interview participants in this category was Julie. At the time of the interview, she was in her first semester as a college freshman at a Midwestern university different from that of the previously described study participants. She had completed an AP Statistics course during her senior year in high school, and was enrolled in a semester-long non-calculus based statistics class at the university. Julie was recommended for participation in the study by her high school statistics teacher. She indicated familiarity with the TI-83, Excel, and Fathom. She ranked her own statistical

ability as a 3 out of 5, stating, “I’ve taken one stats class and am in the process of completing my second, so I feel I have a good start but I’m still learning.”

Category 3: Students enrolled in high school at the time of the interview

Crystal, a high school senior who agreed to be interviewed, was enrolled in an AP Statistics course at the time of her interview. Her class had just completed the study of the first three chapters of a popular AP Statistics text (Yates, Moore, & McCabe, 1998). She was described by her AP Statistics teacher as being a student of average ability who had a strong work ethic. Crystal indicated that she was familiar with the TI-83 calculator, Excel, Fathom, and Data Desk. She ranked her own statistical ability as 3 out of 5. She supported her ranking with the statement, “I have just begun taking a stats course so I do not know everything yet. However, with all of my education and work experience I believe I have a good statistical ability.”

The next interview participant in this category, Bill, was a senior in the first semester of an AP Calculus course when his interview took place. He had not taken a statistics course while in high school. He indicated that he was familiar with the TI-89 calculator and Excel. He ranked his own statistical ability as 3 out of 5, stating, “I feel I can get by with my statistical skills.”

Luke, another student in this category, was a senior enrolled in the same AP Calculus class as Bill. He also had not taken a statistics course while in high school. Like Bill, he indicated familiarity with the TI-89 calculator and Excel. When asked to rank his own statistical ability, he gave himself a 4 out of 5. He supported his ranking by stating, “I have always been very strong in math and have a good grasp on mathematical concepts.”

The fourth student in this category, Daniel, was a high school freshman enrolled in an honors geometry class at the same high school as Crystal, Daniel, and Luke. He indicated that he was familiar with scientific, non-graphing calculators. He ranked his statistical ability as only 1 out of 5, stating, “I haven’t done very much work with statistics.”

The fifth student, Rick, was a high school sophomore enrolled in an algebra class. The algebra class in which he was enrolled was the first year of a two-year sequence covering the topics traditionally included in an introductory one-year algebra course. Rick indicated familiarity with a number of graphing calculators, including the TI-82, TI-83, TI-85, TI-86, and TI-89. He also indicated familiarity with Excel. He ranked his own statistical ability as only 2 out of 5, stating, “I know how to make statistics. I know how there [sic] used. I just don’t think I still know everything.”

The next student, Jessica, was a high school sophomore enrolled in a geometry class. She indicated familiarity with scientific, non-graphing calculators and Excel. She ranked her own statistical ability as only 1 out of 5, stating, “I’ve got no idea what I’m doing!”

The seventh student, Nancy, was a high school sophomore enrolled in a geometry class. She indicated familiarity with scientific, non-graphing calculators and Excel. She ranked her own statistical ability as 3 out of 5 because she had some spreadsheet experience.

The final student in the category, Brooke, was a high school sophomore enrolled in an algebra class. She indicated familiarity with the TI-82, TI-83, and Excel. She ranked

her own statistical ability as 2 out of 5, stating, “I’m just starting to work with statistical software and I need a little more time to develop my skills.”

APPENDIX B

STATISTICAL THINKING TASKS

Task 1: The State of Florida

Suppose that the governor of Florida puts you in charge of finding answers to the following questions:

- a. What is the typical income of adults in the state?
 - b. Will I be re-elected in the election this fall?
 - c. What percentage of the state is computer-literate?
 - d. Does the new drug for treating the West Nile virus actually work?
 - e. How successful was the law which raised the minimum driving age from 16 to 18?
- 1) Describe a plan for gathering the information you would need in order to answer each of the questions, and how you would carry out each plan and report the results to the governor (C).
 - 2) Joe went to a shopping mall in Florida from 8:00 a.m. until 9:00 a.m. on Feb. 20 and stopped people for interviews. He asked each person he stopped, “May I please interview you for a study I am doing about people who live in Florida?” He asked for the following information from all 20 people who agreed to do interviews:
 - a. Age
 - b. Annual income
 - c. Favorite color
 - d. Political party affiliation
 - e. Whether or not they considered themselves computer-literate
 - f. How many hours per week each person sleeps.

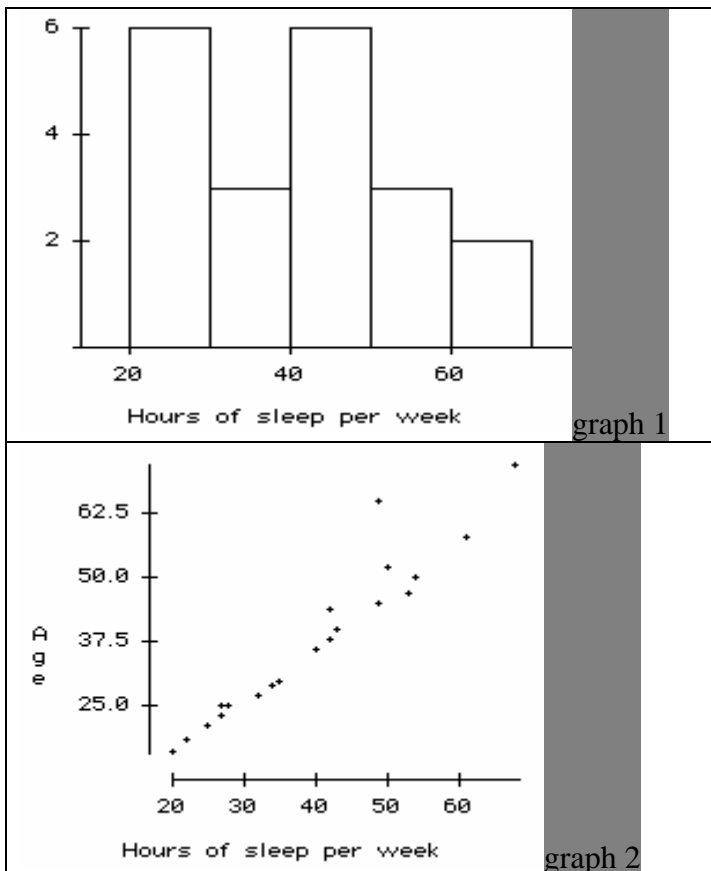
Comment on Joe’s method of collecting data. How useful do you think the information is which he obtained? Why? (C).

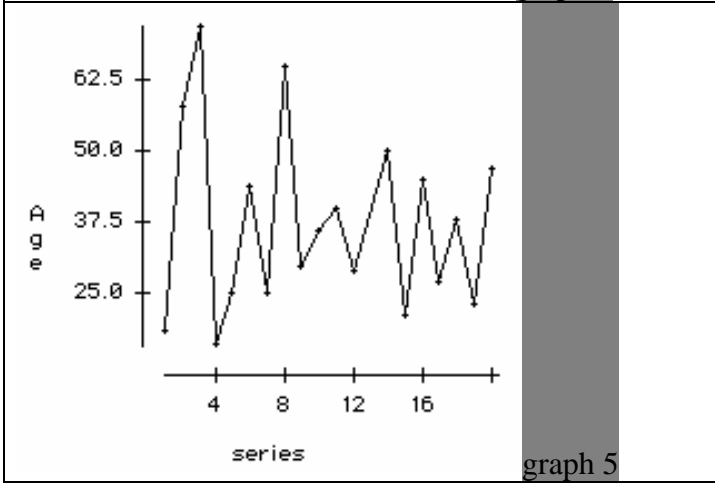
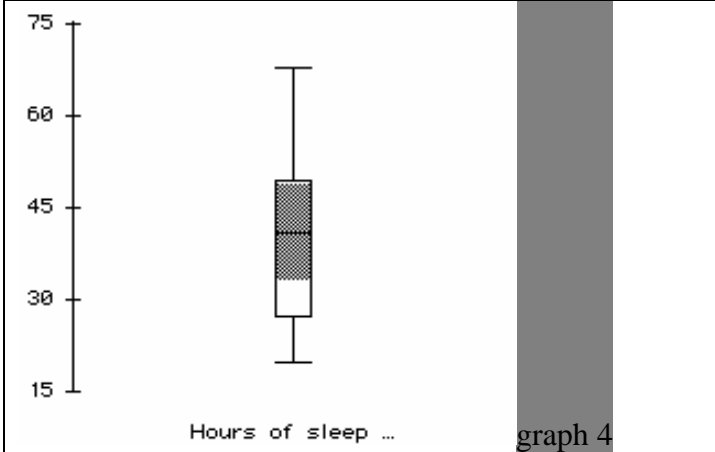
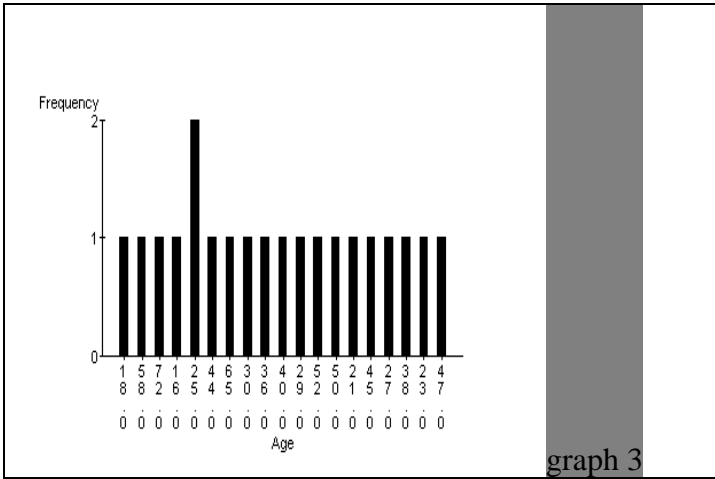
- 3) Here are Joe’s data. For each person interviewed, it shows, in order, the responses to each of these questions: “What is your age?” (Age); “What is your annual income?” (Income); “What is your favorite color?” (Favorite color); “What is your political party affiliation?” (Political party); “Are you computer literate?” (Computer lit.); “How many hours per week do you sleep?” (Sleep). Organize the data cards into different groups (a number of your choosing) in some logical manner that would make the information contained on them easier to read, and might help Joe see patterns in the data he has collected. Explain why you grouped the cards as you did. What are some other ways to organize the data cards? (O)

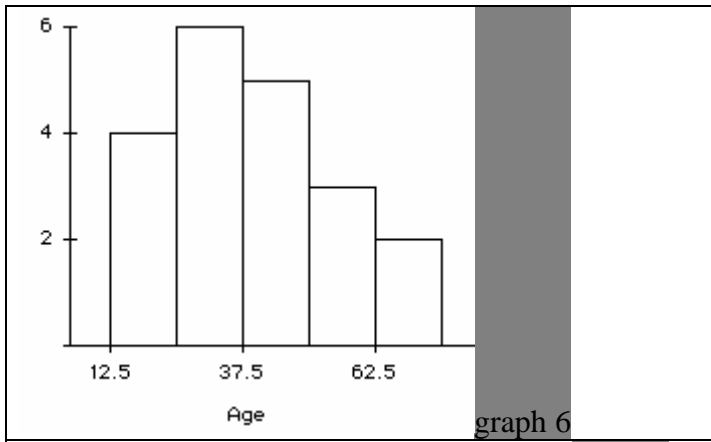
A Age: 18 Annual income: \$16,000 Favorite color: Red Political party: D Computer lit.?: No Sleep: 22	B Age: 58 Annual income: \$56,000 Favorite color: Yellow Political party: R Computer lit.?: Yes Sleep: 61	C Age: 72 Annual income: \$80,000 Favorite color: Blue Political party: R Computer lit.?: No Sleep: 68	D Age: 16 Annual income: \$5,000 Favorite color: Yellow Political party: D Computer lit.?: Yes Sleep: 20
E Age: 25 Annual income: \$25,000 Favorite color: Blue Political party: D Computer lit.?: No Sleep: 28	F Age: 44 Annual income: \$54,000 Favorite color: Green Political party: D Computer lit.?: Yes Sleep: 42	G Age: 25 Annual income: \$24,000 Favorite color: Red Political party: R Computer lit.?: Yes Sleep: 27	H Age: 65 Annual income: \$53,000 Favorite color: Green Political party: I Computer lit.?: Yes Sleep: 49
I Age: 30 Annual income: \$37,000 Favorite color: Blue Political party: R Computer lit.?: Yes Sleep: 35	J Age: 36 Annual income: \$45,000 Favorite color: Yellow Political party: D Computer lit.?: Yes Sleep: 40	K Age: 40 Annual income: \$110,000 Favorite color: Red Political party: R Computer lit.?: Yes Sleep: 43	L Age: 29 Annual income: \$38,000 Favorite color: Blue Political party: D Computer lit.?: No Sleep: 34
M Age: 52 Annual income: \$46,000 Favorite color: Red Political party: R Computer lit.?: Yes Sleep: 50	N Age: 50 Annual income: \$2,000,000 Favorite color: Green Political party: R Computer lit.?: Yes Sleep: 54	O Age: 21 Annual income: \$26,000 Favorite color: Green Political party: I Computer lit.?: Yes Sleep: 25	P Age: 45 Annual income: \$64,000 Favorite color: Red Political party: R Computer lit.?: Yes Sleep: 49
Q Age: 27 Annual income: \$28,000 Favorite color: Green Political party: D Computer lit.?: No	R Age: 38 Annual income: \$40,000 Favorite color: Red Political party: I Computer lit.?: Yes Sleep: 42	S Age: 23 Annual income: \$14,000 Favorite color: Blue Political party: R Computer lit.?: No	T Age: 47 Annual income: \$52,000 Favorite color: Yellow Political party: R Computer lit.?: No

Sleep: 32		Sleep: 27	Sleep: 53
-----------	--	-----------	-----------

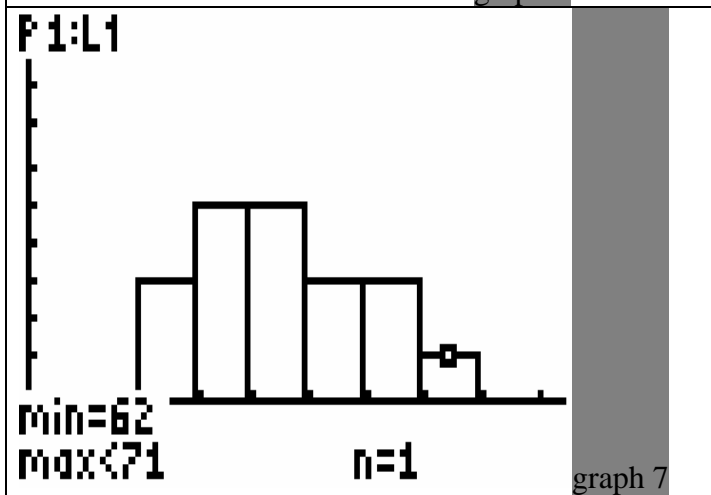
- 4) A) Assuming that people told the truth about their annual incomes, what is the typical annual income of people in Joe's sample? B) Assuming they told the truth about political party affiliation, what is the typical political party affiliation of people in Joe's sample? (O).
- 5) Show graphically how people with different favorite colors answered his question about income. What are some other graphs you could construct to display this information? Which graph do you prefer, and why? (R)
- 6) Show graphically how people considering themselves computer-literate answered the question about political party affiliation. What are some other types of graphs which might display this information? Which graph do you prefer, and why? (R)
- 7) Show graphically how people of different ages answered the question about income. What are some other types of graphs which might display this information? Which graph do you prefer, and why? (R)
- 8) Joe constructed these graphs for the data in his set, but did not label some of them completely. Sort them into families so that each family has only graphs that could possibly contain the same data (R). Explain why they could possibly contain the same data. Also provide labels for graphs that need them.



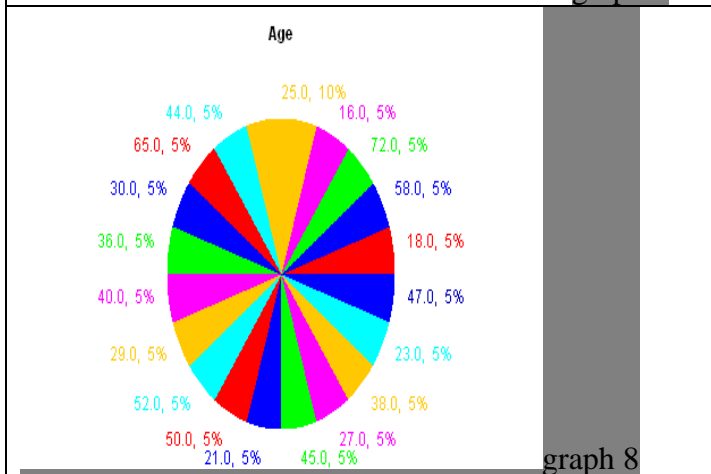




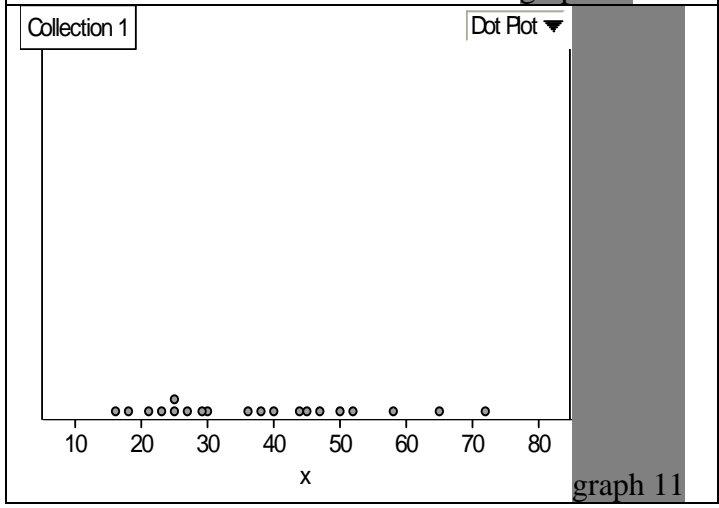
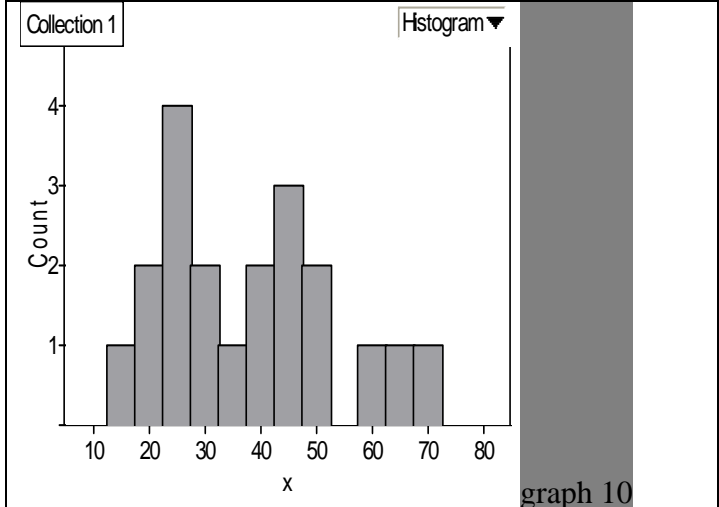
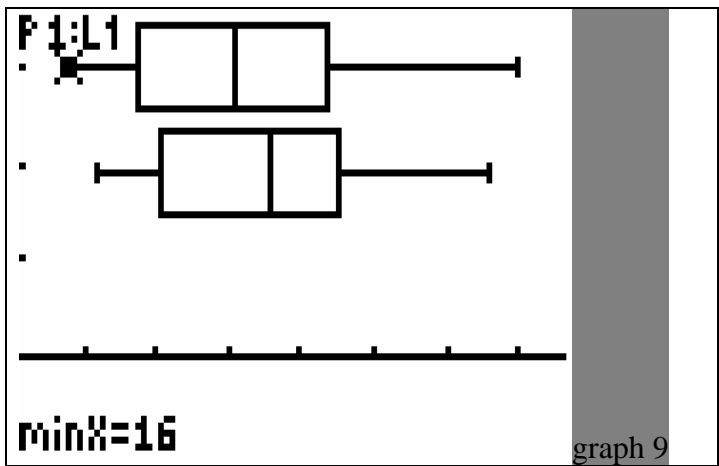
graph 6

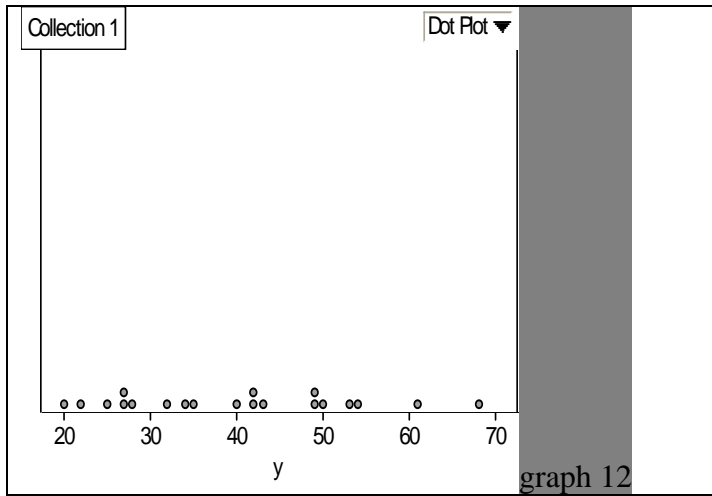


graph 7



graph 8





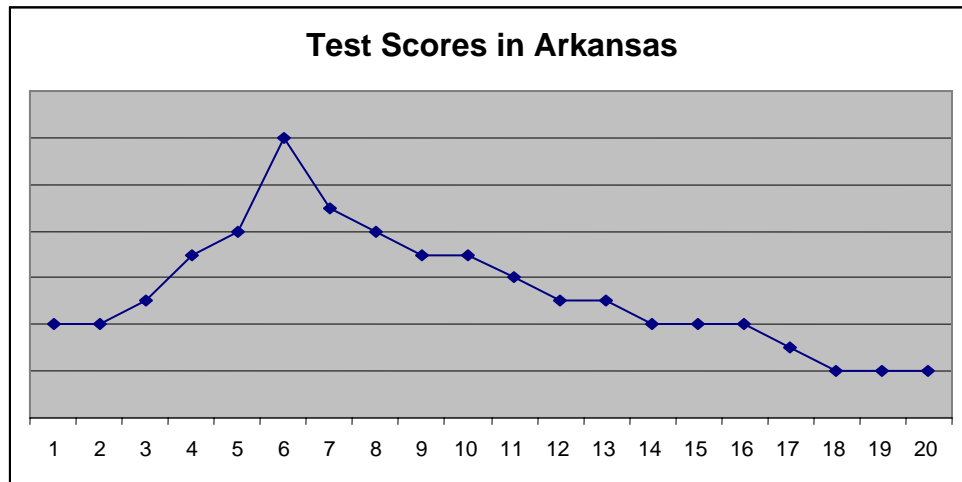
- 9) What are some of the strengths of each of the data displays in each family? What are some of the weaknesses of each? If presenting your findings from this study in a newspaper, which display from each family would you pick? (R)
- 10) Ask them to identify what each of these represents, and give units:
- the leftmost point on the scatterplot
 - the highest horizontal bar on the boxplot
 - the rightmost bar on each parallel boxplot
 - the leftmost point on the dotplot
 - the bar signifying the median on the boxplot (D).
- 11) Point to lines in line plot and ask them to identify what they represent. Point to whiskers of the boxplot and ask what they represent. Point to “collection” label and “max” and “min” labels and ask what they represent. Point to cursor on TI-graph and ask what it represents. (D)

Task 2: Test Scores

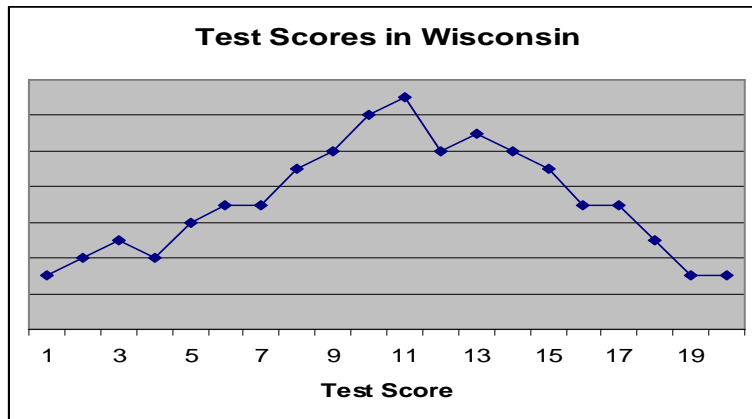
We have the following information on test scores for three different classes in a school building that took the same 20 point test (Class B has two less students than Classes A and C):

Class A	12	13	14	14	15	16	16	17	18	18
Class B	7	9	13	14	15	19	19	20		
Class C	7	11	12	14	15	18	19	19	20	20

- 1) Which two classes in the school do you think were the most similar in performance? Why? (A)
- 2) In which class are the scores the most spread out? How do you know? (O)
- 3) What is the highest score of any student in the three classes? What is the lowest score for a student in class A? (D)
- 4) Are there any unusually high or low test scores in any of the classes? How can you tell? (A)
- 5) This same test was given to all the students in the state of Arkansas. A graph showing the distribution of scores for students in the state is shown below:



- What is the highest score received by any student in Arkansas? How do you know? (D)
- 6) What is the typical score for students in the state of Arkansas? How do you know? (O)
 - 7) The same test was also given to all students in the state of Wisconsin. The graph showing the distribution of scores for students in the state is shown below:



How does the performance of students in the state of Wisconsin compare to the performance of students in the state of Arkansas? Explain your reasoning (A)

- 8) Suppose that teacher A takes a random sample of 5 students' scores from the state of Arkansas and computes the average. Teacher B takes a random sample of 15 students' scores from the state of Arkansas and computes the average. Do you think that there will be any difference between the averages they compute? Why or why not? (A) Do you think that one of the teachers is more likely to obtain an average greater than 17? Why or why not? (A)
- 9) Joe took a random sample of 10 scores from the state of Iowa. Suppose the average score for the state of Iowa is 11. Which of these sets of scores could be Joe's random sample? (possibility of more than one answer) (A)
- 9, 11, 13, 6, 14, 8, 12, 21, 10, 9
 - 4, 5, 6, 5, 5, 5, 7, 6, 7, 5
 - 6, 11, 16, 14, 13, 8, 10, 7, 9, 16
 - 11, 15, 17, 16, 15, 15, 15, 14, 16, 16
 - 7, 16, 12, 12, 15, 14, 10, 7, 6, 11
- 10) Go back to the very first table in the task and look at the data for class A. Suppose the teacher adds 15 points to each student's score. How is the typical score for the class affected? How is the spread of the scores affected? (O)

Task 3: Department Store Worker Salaries

Suppose that in 1999, a newspaper reporter took a random sample of 15 department stores from the state of Illinois. For each department store he sampled, he found out how much the highest paid man and the highest paid woman in the department store were paid per hour. In the table are the results of his survey.

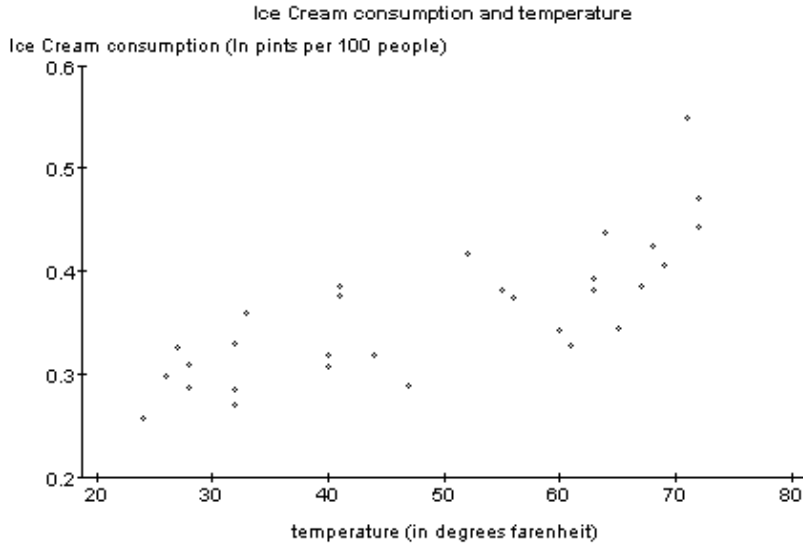
Salaries (in dollars per hour) of the people in the reporter's sample

Department Store	Women	Men
A	\$18.50	\$17.50
B	15.00	15.00
C	17.00	20.00
D	15.50	20.00
E	8.50	20.00
F	18.00	19.00
G	9.00	20.00
H	10.00	18.00
I	10.00	5.50
J	8.50	6.00
K	12.00	10.00
L	14.50	16.50
M	10.50	12.50
N	9.00	10.00
O	16.00	7.00

- 1) Assuming that this is truly a random sample, how do you suppose the reporter went about collecting the data in the table above? (C)
- 2) Make a graphical display that could accompany a newspaper article about the reporter's survey (R). Are there any other types of graphical displays you can think of that would be fitting? (R).
- 3) Are there any unusually high or low values in the column showing women's salaries? How can you tell? Are there any unusually high or low values in the column showing men's salaries? How can you tell? (A)
- 4) How spread out are the salaries in the men's column? (O) How spread out are the salaries in the women's column? (O) In which column are the salaries more spread out: In the men's column or in the women's column?
- 5) After doing this survey, the reporter wrote, "There is no difference between the average hourly salary of the highest-paid male employees of department stores in Illinois and the average salary of their female counterparts." How accurate do you believe this claim to be? Explain. (A)
- 6) The reporter also wrote, "I didn't find any department stores where the highest paid woman made \$17.00 per hour. However, if I had found such a department store, the highest paid man working there would have made \$13.30 per hour." How accurate do you believe this claim to be? Explain. (A)
- 7) The reporter wants to repeat this survey again in 2005 and expand it to the population of the entire United States. Describe a plan he could use for getting the information needed for the survey. (C)
- 8) Describe some of the difficulties which might arise while carrying out the 2005 survey, and how those difficulties could be overcome. (C)

Task 4: Ice Cream

Below are a graph and a table showing the ice cream consumption of people in a city in Northern Canada on 30 randomly selected days during the 1990's. ¹



Temperature (in degrees Fahrenheit)	Ice cream consumption (in pints per 100 people)
41	.386
56	.374
63	.393
68	.425
69	.406
65	.344
61	.327
47	.288
32	.269
24	.256
28	.286
26	.298
32	.329
40	.318
55	.381
63	.47
72	.443
72	.386
67	.342
60	.319
44	.307
40	.284
32	.326
27	.309

28	.359
33	.376
41	.416
52	.437
71	.549
64	.4

1. Data reprinted with permission from Econometrica. Copyright 1970, the Econometric Society. All rights reserved.

- 1) (Circle the leftmost point) What does this point represent in terms of this situation? (D)
- 2) Describe as fully as you possibly can the relationship between ice cream consumption and the outside temperature. (A)
- 3) Suppose an ice cream salesman asks you, “How much ice cream does this group consume if the temperature hits 50 degrees Fahrenheit?” Give the best answer you possibly can, and explain your answer. (A)
- 4) Suppose the same salesman asks, “How much ice cream does this group consume if the temperature hits 10 degrees Fahrenheit?” Give the best answer you possibly can, and explain your answer. (A)
- 5) Are there any unusual data points in the set? Explain. (A)
- 6) Where might unusual points in a data set like this come from? (A, C) How might unusual points change answers you have given to questions above? (A)
- 7) Do you think the graph would have looked different if we had taken a different random sample? If so, how do you think it would have looked different? If not, why do you think it would not have looked different? (A)
- 8) Suppose we get rid of the first column in the table above. Construct a graph to display the information in the second column (R). Discuss other types of graphs you might make in order to display the information in the second column. (R)
- 9) What type of information do you easily get from your graph from #8 that you didn't easily get from the scatter plot? (R)
- 10) Suppose you are the editor of a newspaper, and one of the reporters asks you to comment on the following headline for a story based on the data above: “Higher temperatures cause ice-cream consumption to increase” (A)

Task 5: Going Fishing

This table shows how many of each type of fish were in Lake Del Sol on May 1. Lake Del Sol is a catch-and-release lake, meaning that you must release each fish immediately after you catch it. Game wardens are at the lake to enforce this rule.

Type of fish	Males	Females
Trout	20	40
Bullhead	35	42
Perch	14	10
Sunfish	41	12

- 1) Is catching a fish from Lake Del Sol the same thing as picking a fish from the lake at random? Why or why not? (C)
- 2) Which of the four types of fish in Lake Del Sol is the most common? How do you know? Which of the five types of fish in Lake Del Sol is least common? How do you know? (D)
- 3) If you went fishing at Lake Del Sol on May 1 and caught 10 fish, how many Perch would you expect to have caught? Why? (A)
- 4) How likely do you think it would be that among the 10 fish you caught on May 1 at Lake Del Sol, there would be 2 or less perch? Why? (A)
- 5) How likely do you think it would be that among the 10 fish you caught on May 1 at Lake Del Sol, there would be exactly 1 perch? Why? (A)
- 6) Does this table give any statistical evidence that there is anything unusual or unexpected about the population of fish in Lake Del Sol? Explain your answer. (A)
- 7) Suppose there is a new bait designed to help catch trout. Last Friday, the fisherman in boat A did not use the new bait, while the fisherman in boat B did use it. Here are the results from the day they spent fishing in Lake Del Sol:

	Trout caught	Non-trout caught
Boat A	10	20
Boat B	15	17

Do you think the new bait helps one to catch more trout? Why or why not?

- 8) The fisherman who caught the first fish in Boat A weighed the first fish he caught 7 different times on the same scale. Here are the measurements (in lbs.) that he came up with: 19.2, 21.5, 10.1, 23.1, 22.0, 20.3, 21.8. (O)
 - a. How spread out are the measurements he obtained?
 - b. What do you think the true weight of the fish was? Why?

APPENDIX C

ALIGNMENT OF STATISTICAL THINKING TASKS WITH STATISTICAL THINKING PROCESSES

Statistical thinking process and definition	Task:subtask designed to assess thinking within this process	Code
Describing Data: The explicit reading of data presented in tables, charts, or graphs.	1:10, 1:11, 2:3, 2:5, 4:1, 5:2	D
Organizing and Reducing Data: Arranging, categorizing, or consolidating a given set of data into summary form.	1:3, 1:4, 2:2, 2:6, 2:10, 3:4, 5:8	O
Representing Data: Displaying a given set of data by using graphs.	1:5, 1:6, 1:7, 1:8, 1:9, 3:2, 4:8, 4:9	R
Analyzing Data: Identifying trends and making inferences or predictions from a data display or set, using formal inferential methods when appropriate.	2:1, 2:4, 2:7, 2:8, 2:9, 3:3, 3:5, 3:6, 4:2, 4:3, 4:4, 4:5, 4:6, 4:7, 4.10, 5:3, 5:6	A
Collecting Data: Planning, conducting, and critiquing surveys, experiments, and observational studies.	1:1, 1:2, 3:1, 3:7, 3:8, 5:1	C

APPENDIX D

ALIGNMENT OF STATISTICAL THINKING TASKS WITH NCTM PRINCIPLES AND STANDARDS FOR SCHOOL MATHEMATICS

NCTM (2000) Data Analysis Content Standards (p. 324). ¹	Relevant task:subtask on interview script.
Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them	
-understand the differences among various kinds of studies and which types of inferences can legitimately be drawn from each.	1:1, 1:2, 4:10
-know the characteristics of well-designed studies, including the role of randomization in surveys and experiments.	1:1, 3:1, 3:7, 3:8, 5:1
-understand the meaning of measurement data and categorical data, of univariate and bivariate data, and of the term variable.	1:3
-understand histograms, boxplots, and scatterplots and use them to display data.	1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:11, 3:2, 3:6, 4:1, 4:8, 4:9
-compute basic statistics and understand the distinction between a statistic and a parameter.	1:1, 1:4, 2:8, 5:8
Select and use appropriate statistical methods to analyze data	
-for univariate measurement data, be able to display the distribution, describe its shape, and select and calculate summary statistics.	1:4, 2:1, 2:2, 2:4, 2:5, 2:6, 2:7, 3:3, 3:4, 5:8
-for bivariate measurement data, be able to display a scatterplot, describe its shape, and determine regression coefficients, regression equations, and correlation coefficients using technological tools.	3:6, 4:2, 4:3, 4:4, 4:5, 4:6
-display and discuss bivariate data where at least one variable is categorical.	1:5, 1:6
-recognize how linear transformations of univariate data affect shape, center, and spread.	2:10
-identify trends in bivariate data and find functions that model the data or transform the data so that they can be modeled.	3:6, 4:2, 4:3, 4:4
Develop and evaluate inferences and predictions that are based on data	

-use simulations to explore the variability of sample statistics from a known population and to construct sampling distributions.	5:3, 5:4, 5:5
-understand how sample statistics reflect the values of population parameters and use sampling distributions as the basis for informal inference.	1:1, 2:8, 2:9, 5:8
-evaluate published reports that are based on data by examining the design of the study, the appropriateness of the data analysis, and the validity of conclusions.	1:2, 3:1, 3:5, 4.10
-understand how basic statistical techniques are used to monitor process characteristics in the workplace.	3:5, 5:8

1. Reprinted with permission from Principles and Standards for School Mathematics. Copyright 2000 by the National Council of Teachers of Mathematics. All rights reserved.

APPENDIX E

TASKS DESIGNED TO ALLOW STUDENTS TO DEMONSTRATE
KNOWLEDGE OF SOME OF THE TOPICS INCLUDED ON THE
2001 AP STATISTICS SYLLABUS BUT NOT INCLUDED IN
THE NCTM PRINCIPLES AND STANDARDS FOR
SCHOOL MATHEMATICS

Task/subtask	Relevant topic from the AP Statistics course syllabus (College Board, 2001b)
3:5	Two sample (independent and matched pairs) t procedures.
4:7	Inference for the slope of least-squares regression line.
5:2	Exploring categorical data: frequency tables
5:3, 5:4, 5:5	Binomial probability distributions
5:6, 5:7	Chi-square procedures

APPENDIX F

STUDENT QUESTIONNAIRE

Name _____

Thank you for agreeing to participate in an interview! Please take a few moments to complete this survey.

- 1) Check off all of the following technology that you know how to use:
- a. TI-82 calculator _____
 - b. TI-83 calculator _____
 - c. TI-85 calculator _____
 - d. TI-86 calculator _____
 - e. TI-89 calculator _____
 - f. TI-92 calculator _____
 - g. Other graphing calculator (please list) _____
 - h. Scientific, non-graphing calculator _____
 - i. Simple pocket/handheld calculator _____
 - j. *Excel* software program _____
 - k. *Fathom* software program _____
 - l. *Data Desk* software program _____
 - m. Other software program that would help with statistics (Please list) _____.
 - n. Other technology, not listed here, which would help with statistics (please list) _____.

2) How would you rank your own statistical ability? (Circle one: 1 = low, 5 = high).

1 2 3 4 5

3) Please briefly explain the ranking you gave yourself in question 3.

4) Please list any questions or comments you have for me before the interview takes place: