UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

A MODEL FOR DEVELOPING AND ASSESSING COMMUNITY COLLEGE STUDENTS' CONCEPTIONS OF THE RANGE, INTERQUARTILE RANGE, AND STANDARD DEVIATION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

MIKHAIL TUREGUN
Norman, Oklahoma
2011

A MODEL FOR DEVELOPING AND ASSESSING COMMUNITY COLLEGE STUDENTS' CONCEPTIONS OF THE RANGE, INTERQUARTILE RANGE, AND STANDARD DEVIATION

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF INSTRUCTIONAL LEADERSHIP AND ACADEMIC CURRICULUM

BY

Dr. Stacy Reeder, Chair

Dr. John Chiodo

Dr. Michael Crowson

Dr. Neil Houser

Dr. Sacra Nicholas

In Memory of Jack Cain

An inspiring mentor, an esteemed colleague, a well-regarded teacher, an expert mountain climber, and a great friend

ACKNOWLEDGEMENTS

It is not as much about the destination as it is about the journey. The Ph.D. program has been a great intellectual and emotional journey for me. I take a great pleasure acknowledging many individuals, who have accompanied, encouraged, challenged, and guided me throughout this journey. I would like to express my gratitude to my advisor, Dr. Reeder, who has spent many hours helping me understand the fundamental theories and ideas about mathematics education research, teaching and learning, given me confidence in my own ideas, and allowed me the autonomy to pursue my own particular research interests which have paved the way for this dissertation.

I am grateful to Dr. Chiodo for encouraging, helping and providing me with an opportunity to publish my first paper in the doctoral program. I would like to thank Dr. Crowson for being a great mentor and helping me become a better statistics educator. I express my gratitude to Dr. Houser, who helped me view life, society and the oppressive structures of the “official world" from a different vantage point. I would like to thank Dr. Nicholas for her insightful suggestions during my research. I am very appreciative of my committee members for being my teachers and providing me with valuable feedback during the course of my studies at the University of Oklahoma.

I am particularly indebted to my parents, Doğan and Güneyyel Türegün, for instilling in me the importance of education from a very early age in my childhood, and their relentless encouragement and unconditional support in all my educational studies and endeavors, both professional and personal, throughout my life. I have learned a great deal from my parents about life and giving my best effort to live as a good person.

I acknowledge that the intellectual debt to the work of other researchers, both concordant and discordant with my ideas, is deep and extensive. I extend my thanks to my colleagues from whom I have learned a lot. Last but not least, I would like to thank especially my students from whom I have learned the most.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Traditional curricular materials and pedagogical strategies have not been effective in developing conceptual understanding of statistics topics and statistical reasoning abilities of students. Much of the changes proposed by statistics education research and the reform movement over the past decade have supported efforts to transform teaching practices to include an emphasis on students' development of conceptual understanding rather than a focus on mechanical calculations (Chance & Garfield, 2002). In recent years, major research interest in the statistical understanding of students seemed to be focused on the measures of centers in distributions (Shaughnessy & Ciancetta, 2002).

Despite some very recent research studies investigating conceptual understanding of variability among community college students in introductory statistics courses, there are still several unanswered questions regarding how we, as statistics educators, can help students develop conceptual understanding of measures of spread at the community college level (Budé, 2007; Slauson, 2008). The purpose of this study is to investigate students' conceptual understanding of measures of spread and provide learning experiences and environments to support students in developing conceptual understanding of measures of spread, as well as to assess community college students' conceptual understanding of measures of spread.

In order to gain a broader and deeper insight into to my research questions, both quantitative and qualitative data were gathered, analyzed, and embedded in an action research framework. The research focus was on investigating students' conceptual understanding of statistical measures of spread, and how students establish the connections between measures of spread and behavior of distributions of data sets.

A detailed analysis of the students' responses is presented to reveal the range of students' conceptions. My findings suggest that, despite the very limited mathematical and statistical backgrounds, this particular group of community college students exhibited at least some pre-existing conceptions of measures of spread, on which we, as statistics educators, need to capitalize in order to help students develop their conceptions further. This particular result both reinforces and is reinforced by constructivist/student centered theories, and indicates the applicability of these theories to college statistics students.

CHAPTER ONE

INTRODUCTION

The widespread consensus among statistics educators and statisticians has been that traditional curricular materials and pedagogical strategies used in introductory statistics courses have not been effective in developing conceptual understanding of statistics topics. Traditional approaches were also found by many researchers to play a rather limited role in improving statistical reasoning abilities of and promoting statistical literacy among students (Cobb, 1992; Gal, 2003; Garfield, Hogg, Schau, and Whittinghill, 2002; Hassad, 2008; Moore, 1997). Growing frustration and dissatisfaction experienced by students and faculty with introductory statistics courses over the years have led researchers to investigate alternative models and suggest new approaches to teaching and learning of introductory statistics. These alternative models and new approaches to teaching and learning of introductory statistics have paved the way for the statistics education reform movement.

At the heart of the statistics education reform lies the question of how teaching and learning of statistics can be improved. The reform movement over the past decade has supported efforts to transform teaching practices to include an emphasis on students' development of conceptual understanding rather than a sole focus on mechanical calculations (Chance & Garfield, 2002). The traditional methods of learning by way of students passively listening to lectures and working in isolation have been identified as leading causes for statistics, and also for mathematics, to be viewed as a sequence of disjoint topics padded with a series of techniques and rote memorization of fragmented facts (Begg, 2004; Meletiou-Mavrotheris & Lee, 2002; Turegun, 2009; Whitehead, 1929).

Organization of the Dissertation

My dissertation is composed of five chapters. Chapter One outlines the purpose of the study, and provides the research questions addressed within the study. Chapter Two gives a background and a literature review related to the research. Chapter Three describes the research methodology and the setting for the study. Chapter Four provides an analysis of the research findings and a summary of the findings. Finally, Chapter Five presents a discussion along with the implications of this study.

Statistics Education Reform and Constructivism

The guiding theory for statistics education reform is based on a learning theory arising from earlier ideas and writings of Jean Piaget on cognitive development and is referred to as constructivism (Garfield, 1995; von Glasersfeld, 1995; Wheatley, 1990). Constructivism views learning as an active process and learners as cognitively active human agents who construct their knowledge through interaction with the environment. The basic tenets of constructivism are in stark contrast with and pose a strong challenge to the earlier conceptions of learners as passive beings whose empty minds are to be filled with information transmitted directly from and controlled by a central knower. According to Dewey (1933, p. 261), the mind of the student is "treated as if it were a cistern" into which information is filled with one set of pipelines and pumped out by way of another set of pipelines in the form of regurgitation on demand. In his analysis of the teacher-student relationship, Freire (1970, p. 72), varying the metaphor, used "the banking concept of education" as another metaphor. This vivid metaphor also illustrates the above mentioned narrative character of teaching as a futile cycle of teacher, the depositor, making deposits which students, the depositories, who in turn receive, memorize, and repeat (Freire, 1970).

Several researchers indicated that in these traditional narrative-based settings students were not learning what statistics educators wanted them to learn (Chance, delMas, and Garfield, 2004; delMas & Liu, 2005; Gal, 2003; Garfield & Gal, 1999; Konold, 1995). The statistics education reform movement goes beyond a mere criticism of this narrative character of teaching. Statistics education reform embraces the idea that students construct knowledge by constantly negotiating and renegotiating new knowledge in relation to past experiences. To a large extent this negotiation process is influenced by social, cultural, and historical backgrounds of students, as well as of those around the students (von Glasersfeld, 1995). It is through this complex network of rich interactions that students are able to propose or challenge new ideas, engage in dialogue about those ideas, and try to make sense of those ideas. Constructivism, along with the above outlined Vygotskian perspective, known as social constructivism, which takes into account social interactions and historical or cultural influences on learning, have become the guiding theory for reform efforts, and for much research in mathematics, science and statistics education (Gordon, 1995; Hassad, 2008; Mvududu, 2005).

Despite the above described views, which guide the statistics education reform movement, there are many introductory statistics courses at the college level that continue to use the traditional lecture-and-listen format (Moore, 1997). In addition to using the lecture-and-listen format predominantly, many such courses also heavily rely on having students do assignments in textbooks or in computer labs, and take multiple choice or traditional tests emphasizing formulae, rote memorization skills and procedural knowledge, as opposed to conceptual knowledge of statistics (Garfield, 1995). Statistics education reform, guided by the views of constructivism and social constructivism, shifts the emphasis more toward conceptual knowledge, and focuses on helping students develop conceptual understanding of statistics topics.

Conceptual Understanding

Some of the basic goals of statistics education are often stated or referred to by the researchers in terms of statistical literacy, statistical reasoning and statistical thinking (Collins & Mittag, 2005; delMas, Garfield, and Chance, 1999; Reyes, 2002; Zieffler, Garfield, Alt, Dupuis, Holleque, and Chang, 2008). Even though there has been an ongoing discussion among researchers on how to define these concepts unanimously, there continues to be a great deal of discordance regarding the definitions and the nature of these concepts which have several different competing models (delMas, 2002). As a result, the concepts and definitions of statistical literacy, statistical reasoning and statistical thinking remain unclear, overlap with one another, and often are used by many researchers interchangeably (Ben-Zvi & Garfield, 2004; delMas, 2002; Garfield & Chance, 2000).

Motivated, in part, by the lack of clear and commonly used definitions of these terms as the outcomes of student learning, the recent research efforts have focused on conceptual understanding in statistics education (Broers, 2006; Budé, 2007; Slauson, 2008). However, the research on developing and assessing conceptual understanding of core statistical ideas has been and remains to be a complex endeavor (Budé, 2007; Hawkins, 1996).

Many educators regard conceptual understanding as one of the primary goals in education, even though Budé (2007) indicated that the term conceptual understanding might not be a well defined term in itself. At this point, I would like to consider the following questions: What does it mean to come to know or to understand a concept? And what is conceptual understanding? Since my focus in this study is on students' conceptual understanding, I am also concerned with how to tell if a student has achieved what is meant by conceptual understanding. Devlin (2007) posed a number of similar questions and tried to provide some insight.

Conceptual understanding is one of the five strands of mathematical proficiency included in the 2001 report of the Mathematics Learning Study Committee of the National Research Council (NRC). According to the NCR Committee report, conceptual understanding is achieved by an integrated and functional grasp of mathematical ideas. Devlin (2007) refers to this definition of conceptual understanding as functional understanding. Hiebert & Lefevre (1986) view knowing-why as an indicator of conceptual understanding, as opposed to knowing how-to, which they consider to be an element of procedural understanding.

Other researchers focus on the "implicit or explicit understanding of the interrelations between units of knowledge" as critical characteristics of conceptual understanding (Rittle-Johnson, Siegler, and Alibali, 2001, p. 346). The relational aspect of conceptual understanding is also noted by Dantonio & Beisenherz (2001). In their view, conceptual thinking and understanding require the learner to create patterns or relationships among the different pieces of information gathered by the learner. Rasmussen, Zandieh, King, and Teppo (2005) indicate that students' sense making and understanding of mathematics necessarily include content and connections within content. A common thread emerging from the attempts made by several of the above cited researchers to define conceptual understanding is the formation of cognitive connections among the related components of a cognitive entity.

Conceptual Understanding and Procedural Understanding

Although there are two distinct groups of researchers advocating either procedural or conceptual understanding, it is certainly not my intent to promote a dichotomy when I emphasize conceptual understanding as an important component of mathematics and statistics education. However, whether the educational aim treats procedural understanding as solely an end in itself or views it as an integral part of the teaching and learning process is a problematic issue. Procedural understanding as a skill acquired with the aid of conceptual understanding is essential, if the acquired skill is to be retained for an extended time period. It is just as necessary to acquire some procedural skills prior to developing conceptual understanding for some advanced mathematical subjects, such as calculus and beyond (Devlin 2007; Rittle-Johnson, Siegler, and Alibali, 2001).

False oppositions or dichotomies, such as information and understanding, phonics and meaning, procedural skills or understanding and conceptual understanding, are commonly set up and have a long history in the field of education (Devlin, 2007; Dewey, 1933; Wu, 1999). I argue that basic skills or procedural understanding and conceptual understanding are tightly intertwined in the fields of mathematics and statistics. For example, the mechanistic nature of the first-outer-inner-last (FOIL) method in multiplying binomials provides insights into to the pattern for factoring trinomials and the theory of the zeros of the of the polynomial functions. As another case, the need to know an automated skill, such as the multiplication table for single digits, is very beneficial when doing a standard multiplication problem involving more than single-digit numbers. By the same token, if procedural understanding or basic skills are avoided entirely, only a partial conceptual understanding may be achieved. As Wu (1999) stated, there is no "royal road to conceptual understanding", and the basic procedural skills should be taught as well as conceptual understanding.

Guidelines for Assessment and Instruction in Statistics Education (GAISE)

Over a decade of efforts focusing on reform in statistics education from the academic community, primarily in the United States, coordinated mainly by the American Statistical Association (ASA), the Mathematical Association of America (MAA), and the National Science Foundation (NSF), have culminated in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) project (http://www.amstat.org/education/gaise/GAISECollege.htm). In 2005, the ASA endorsed the GAISE recommendations which are considered to be a basis for statistics education reform in the introductory college statistics courses (Hassad, 2008). According to the GAISE recommendations, there is a need for statistics educators to stress conceptual understanding rather than mere knowledge of procedures in teaching and learning of introductory statistics. With the emphasis placed on the conceptual understanding there should be fewer procedures, recipes, calculations, and derivations (Lovett & Greenhouse, 2000; Turegun & Reeder, 2008).

The GAISE recommendations also articulated the point of view that numerous introductory statistics courses contain a great deal of material with a collection of ideas which are presented disjointly by instructors, understood superficially and forgotten quickly by students. The consensus among the contributors to the GAISE recommendations was that there is little value in knowing a set of procedures if students do not understand the important concepts.

Researcher Perspective

I have been teaching mathematics courses for twenty-three years and statistics courses for fourteen years at the community college level. In the early years of my teaching career, as a mathematics instructor at a community college, I was asked to teach an introductory statistics course offered by the mathematics department. Having been trained as an engineer and being a mathematician, I did believe that I could teach statistics, even though my formal statistics education might be summed up as such: the second time I was in a statistics course I was there to teach the course.

I was given a departmental course syllabus common to all the sections of the course. The course syllabus tightly outlined the general guidelines and listed the topics of the course with a timeline. Since it was my first time teaching statistics, I thought it was rather helpful for me to have such detailed guidelines to follow, especially during the first couple of semesters. A similar sentiment is shared by many introductory statistics instructors who happen to be teaching the course for the first time (Slauson, 2008). However, the more frequently I taught the course in the subsequent semesters the more confidence and competence I gained in the content knowledge and topics of an introductory statistics course.

The increased confidence and competence that I gained in the content knowledge combined with continuous reflections on my teaching had gradually built up my aspirations to venture into implementing small incremental changes in the course. These small incremental changes and my conviction to not continue to do business as usual when I recognized there were better ways to teach the course set in motion a transformation of the way I viewed the course, as well as my transformation as a teacher. The narratives of transformations of other teachers, such as Meier & Rishel (1998) and Slauson (2007), for having a desire to "do something different" in terms of teaching their particular courses are, in essence, similar to my narrative of transformation as a teacher.

However, my own narrative of transformation as a teacher can be traced back to a bifurcation point in my teaching as a limit cycle of "doing the same." According to systems theorists, when or how a bifurcation point occurs cannot be precisely predicted (Briggs & Peat, 1999). When small changes are fed back into a complex and non-linear system, such as teaching and learning, at some point, an interruption of a limit cycle and a bifurcation point or several bifurcation points occur. I believe that my incessant reflections on teaching and learning personally and collectively with colleagues at conferences and professional meetings, and with my fellow doctoral students in a Ph.D. program have led the way to the formation of several feedback loops in my mind regarding my own teaching practices.

I felt strongly, in the sense of Palmer's (2007) notion of "divided no more", about changing how I was teaching and about transforming the classroom environment in which my students and I were to take part. My personal need and desire, and the subsequent decision to live "divided no more" are centered on the concept of theory informing practice, not on the intent of assaulting or criticizing my colleagues, many of whom I happen to notice, when walking in the hallways, to be engaged in transmitting information via, almost exclusively, lecture-and-listen mode based classroom settings.

From a personal perspective, my transformation as a teacher is also a reflection of how strong and powerful it can be to name and claim one's identity and integrity.

Meanwhile, I was slowly beginning to realize that the course was designed with an excessive amount of material and contained too many topics. In order to get through the listed topics in the course syllabus within a semester, I had to follow the timeline suggested by the departmental syllabus very strictly. Even then, I was still forced to have the tests administered in the college's testing center to free up class time to be used for lectures. Lectures, without any time for group discussions, hands-on activities or any other innovative and progressive approaches seemed to be the only efficient way of getting through the long list of topics on time.

There were several indicators that the lecture-and-listen approach combined with the long list of topics to teach introductory statistics was not a very effective way of teaching the course. Some of these indicators were more general in nature, such as the growing frustration and dissatisfaction experienced by students and myself with the traditional methods of teaching and learning of introductory statistics by way of students passively listening to my lectures and working in isolation. Even though this traditional method where an all- or more-knowing teacher stands in front of a classroom and transmits a prescribed set of facts and procedures has been practiced widely for quite some time, it has been identified as a leading cause for mathematics and statistics to be viewed by most students as a lengthy sequence of disjoint topics padded with a series of techniques and rote memorization of fragmented facts (Begg, 2004; Meletiou-Mavrotheris & Lee, 2002; Moore, 1997; Turegun, 2009; Whitehead, 1929). In turn, this particular impression of mathematics and statistics which is prevalent among students can be interpreted as a sign of a lack of conceptual understanding of the core ideas of these two fields.

Some of the other indicators were more particular in nature. One of these indicators can be articulated in the following realization to which I have come while teaching the course. As I was teaching the course, I reflected and acknowledged that I had not developed any conceptual understanding of the statistical topics when I had previously studied these topics myself. I had little or no difficulty in performing the calculations using my procedural understanding of the formulae and recipes. Even so, I could not explain or describe the ideas behind those formulae and procedures. I realized that there were similar discrepancies between conceptual and procedural understanding of introductory statistics topics among my own students. For instance, the standard deviation of a data set is calculated by means of the following formula.

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

Using the above formula I could, as a student, manually calculate the standard deviation without any difficulties. However, I did not know what it meant conceptually. Likewise, my students had little or no difficulty calculating the standard deviation using their required calculators once they discovered and remembered which buttons to press or which menu of the calculator to use. But like me, they could not explain or describe what standard deviation meant in the context of a particular data set.

Although some students could regurgitate the generic definition of the standard deviation as a measure of spread, they were not able to explain or make the connection between the number they calculated and the behavior of the data set which they used to calculate it. This apparent difference between the levels of procedural and conceptual understanding of standard deviation among my students is not an isolated case or anecdotal evidence. The lack of conceptual understanding of core concepts and big ideas such as standard deviation among introductory statistics students is documented extensively in the statistics education research literature (Collins & Mittag, 2005; delMas, Garfield, and Chance, 1999; delMas & Liu, 2005; Reyes, 2002; Zieffler, et al., 2008).

Goal of the Study

Statistics is defined as a "science of data" (Moore, 2007), and variation, as a central concept in statistics, appears all around us in every imaginable context from gasoline prices, to test scores, to income levels. Variation is an inherent aspect of every data set and with every data set the fundamental task that a statistician is faced with is to explain, predict, or account for variability. Facilitating statistical understanding among students and enabling them to develop conceptual statistical understanding then naturally starts with carefully establishing conceptual understanding of variability in a number of different statistical topics such as, data analysis and graphs (i.e., descriptive statistics), sampling distributions (i.e., inferential statistics), and basic probability.

Generally, introductory statistics courses begin with descriptive statistics in which data analysis and graphical tools, such as dot plots, stem plots, box plots and histograms, to display and analyze data are investigated. Overall spread between the maximum and minimum in a data set, the range; central spread about the median in a data set, the interquartile range (IQR); and an average spread about the center in a data set, standard deviation (SD) are regarded as the three basic measures of spread which can be effectively used to initially introduce the concept of variation in descriptive statistics.

The goal of this study is to examine the individual perceptions and conceptual understanding that community college students have of statistical measures of spread in descriptive statistics. It should be made clear here that the focus of the study is not on individual computational skills or insights for calculating variance or standard deviation, but rather on the students' perceptions and conceptual understanding of statistical measures of spread.

Research Questions

In recent years, several research studies related to students' statistical understanding have seemed to focus on the measures of centers in distributions (Shaughnessy & Ciancetta, 2002). Although variation is the foundation of statistical reasoning process and thinking, until very recently there has been only a limited number of research studies conducted on students' understanding of variation. This particular lack of research on students' understanding of variation might be a reflection of the curricular emphasis, which has traditionally focused on measures of center rather than on measures of spread (Watson, Kelly, Callingham, and Shaughnessy, 2003).

The traditional curricular focus on measures of center can be attributed to procedural teaching practices which view variation and standard deviation to be messy and complex formulations. While much of the initial work focused on students' understanding of variation in a sampling environment (Shaughnessy & Ciancetta, 2002), one early research study concentrated on both American and Australian school children's understanding of and thinking about variation in probability and probability sampling settings (Shaughnessy, Watson, Moritz, and Reading, 1999).

In addition to my growing personal dissatisfaction with my introductory statistics students' misconceptions and lack of conceptual understanding of standard deviation, I was also motivated to conduct this research study by the future research suggestion made by Watson et al. (2003) to investigate older students' understanding of spread. There are a few investigative research studies available on the understanding of variation among university students in the United States by Meletiou & Lee (2002) and in Mexico by Sanches & Mercado (2006), but these type of studies are considered to be very limited (Inzuna, 2006). While a number of very recent research studies investigated community college students' conceptual understanding of variability in an introductory statistics course, there are still several unanswered questions regarding how we, as statistics educators, can help students develop conceptual understanding of measures of spread in a descriptive statistics setting at the community college level (Budé, 2007; Slauson, 2008).

The fact that almost half of the university and college students are enrolled in community colleges nationwide combined with the open admission system of community colleges make this particular student population quite large, diverse, and considerably different than the traditional student body found at many colleges and universities across the nation. Teaching introductory statistics to such a diverse body of community college students is not an easy mission. Research on teaching and learning of introductory statistics at the community college level where students are, in general, educationally underserved might provide unique insight into the teaching practices of community college instructors.

While the general focus of my study is on the conceptual understanding of the statistical measures of spread among community college students in an introductory statistics course, I intend to seek answers particularly to the following questions:

1. What are some of the pre-existing conceptions of students regarding statistical measures of spread?
2. How do students articulate their conceptual understanding of statistical measures of spread?
3. How do students establish connections between statistical measures of spread and behavior of distributions of data sets?

Summary

Since the traditional curricular materials and pedagogical strategies used in introductory statistics courses have not been effective in developing conceptual understanding of statistics topics, many researchers were motivated to investigate alternative models and suggest new approaches to reform teaching and learning of introductory statistics. The recent research efforts have focused on conceptual understanding in statistics education (Broers, 2006; Budé, 2007; Slauson, 2008). The GAISE recommendations also emphasize a need for statistics educators to stress conceptual understanding rather than mere knowledge of procedures in teaching and learning of introductory statistics.

Since variation is considered by many statistics educators to be an inherent aspect of every data set, facilitating statistical understanding among students and enabling them to develop conceptual understanding of variability in descriptive statistics is an important issue. It is also a good starting point, since, generally, introductory statistics courses begin with descriptive statistics in which data analysis and graphical tools, such as dot plots, stem plots, box plots and histograms, to display and analyze data are investigated. Overall spread between the maximum and minimum in a data set, the range; central spread about the median in a data set, the interquartile range (IQR); and an average spread about the center in a data set, standard deviation (SD) are regarded as the three basic measures of spread which can be effectively used to initially introduce the concept of variation in descriptive statistics.

In this Chapter, I outlined the need and the purpose of this study along with a brief yet broad overview of the current research in statistics education. I then, via a personal perspective, led up to my research questions by progressively narrowing the research focus. The goal of this study is to examine the individual perceptions and conceptual understanding that community college students have of statistical measures of spread in descriptive statistics. Chapter Two provides a review of the literature related to statistics education reform, constructivism and conceptual understanding. A discussion of the Biggs and Collis cognitive development model is presented in Chapter Three.

CHAPTER TWO

A REVIEW OF THE RELATED LITERATURE AND BACKGROUND

The widespread consensus among statistics educators and statisticians has been that the traditional curricular materials and pedagogical strategies used in introductory statistics courses have not been effective in developing conceptual understanding of statistics topics and statistical reasoning abilities of students and in promoting statistical literacy among students (Cobb, 1992; Gal, 2003; Garfield, Hogg, Schau, and Whittinghill, 2002; Hassad, 2008; Moore, 1997).

Frustration and dissatisfaction experienced by students and faculty with introductory statistics courses over the years have led researchers to investigate alternative models and suggest new approaches to teaching and learning of introductory statistics. At the heart of the statistics education reform lies the question of how teaching and learning of statistics can be improved. The reform movement over the past decade has supported efforts to transform teaching practices to include an emphasis on students' development of conceptual understanding rather than a focus on mechanical calculations (Chance & Garfield, 2002). The traditional methods of learning by way of students passively listening to lectures and working in isolation have been identified as leading causes for statistics, and also for mathematics, to be viewed as a sequence of disjoint topics padded with a series of techniques and rote memorization of fragmented facts (Begg, 2004; Meletiou-Mavrotheris & Lee, 2002; Turegun, 2009; Whitehead, 1929).

The guiding theory for statistics education reform is based on a learning theory arising from earlier ideas and writings of Jean Piaget on cognitive development and is referred to as constructivism (Garfield, 1995; von Glasersfeld, 1995; Wheatley, 1990).

Constructivism views learning as an active process and learners as cognitively active human agents who construct their knowledge through interaction with the environment. According to constructivism, the learners, who possess their individual ideas and prior knowledge in the form of schemas, fit the actively constructed new knowledge into their own cognitive frameworks "through the processes of assimilation, accommodation, and equilibration" (Houser, 2006, p. 16). Assimilation takes place when an individual learner comes across a new idea and uses existing schema to make sense of the new idea or a phenomenon. If the new idea or phenomenon does not seem to conflict and reasonably fits with the existing schema then it is assimilated or fitted into the existing schema. However, if the existing schema cannot fully explain or help the learner reason through the new idea, it creates a perturbation for the learner. Then, an inquisitive review of the phenomenon begins which, in turn, may result in an accommodation of the new idea with the existing schema, either by altering or replacing existing schemas.

These tenets of constructivism are in stark contrast with and pose a strong challenge to the earlier conceptions of learners as passive beings whose empty minds are to be filled with information transmitted directly from and controlled by a central knower. In his analysis of the teacher-student relationship, Freire (1970) revealed this narrative character of teaching with a vivid metaphor. "The banking concept of education" illustrated the futile cycle of teacher, the depositor, making deposits which students, the depositories, who in turn receive, memorize, and repeat on demand (Freire, 1970, p. 72). Constructivism, along with a Vygotskian perspective, known as social constructivism, which takes into account social interactions and historical or cultural influences on learning, has become the guiding theory for reform efforts and much research in mathematics, science, and statistics education (Gordon, 1995; Hassad, 2008; Mvududu, 2005).

Many introductory statistics courses at the college level use the traditional lecture-and-listen format along with having students do assignments in textbooks or in computer labs, and taking multiple choice or traditional tests emphasizing formulae, rote memorization skills and procedural knowledge, as opposed to conceptual knowledge of various introductory statistics topics (Garfield, 1995; Moore, 1997). Several researchers indicated that in the above described traditional settings students were not learning what statistics educators wanted them to learn (Chance, delMas, and Garfield, 2004; delMas & Liu, 2005; Gal, 2003; Garfield & Gal, 1999; Konold, 1995).

Statistics Education Reform Efforts and Conceptual Understanding

The research in statistics education has considered and proposed a multitude of changes in introductory statistics course so that we, as statistics educators, can best help students learn. Much of these changes suggested by statistics education reform efforts have focused on the development of conceptual understanding of the underlying statistical ideas. These changes were made possible, in part, by recent trends in society towards quantitative literacy, and the abundance of advance technologies in computing and communicating. Based on a review of literature focusing on the changes proposed to reform statistics education, I identified the following four broad and interconnected areas: changes related to content, pedagogy, technology, and assessment.

Content Related Changes

Statistical content is broadly viewed as what we, as statistics educators, want students to learn in statistics. The traditional content and approach to teaching statistics as a sequence of linearly and hierarchically ordered disjoint topics padded with a series of techniques are being challenged widely in the literature. Meletiou-Mavrotheris & Lee (2002) stated that presenting statistical content as a linear hierarchically ordered list of topics might lead students to view statistics as a collection of fragmented formulae and procedures taught in isolation without any interconnectedness established among the various topics. This fragmented view of statistics, created and promoted in a large part by the traditional content and approaches, tends to impress upon our students an image of statistics as a collection of specific, factual and behavioral objectives (Begg, 2004).

In addition to their fragmentarily organized content, the traditional introductory statistics courses have been also largely based on probabilistic inference (Moore, 1997). The content related changes, which were suggested by the joint curriculum committee of the American Statistical Association (ASA) and the Mathematical Association of America (MAA), and approved by the ASA, advocated exploring and producing data, and statistical concepts (Cobb, 1992; Moore, 1997). Since introductory statistics courses are the first, and only course for many students especially at the community college level, these courses should offer opportunities for students to work with real, not just realistic, data arising from real problem settings. Interpretations of graphics, developing strategies for explorations of data, and informal inference need to be brought to the forefront of the course. The conceptual meanings of P-value, confidence, significance should be emphasized in an introductory college level statistics course where the students may be at any level of mathematical sophistication or ability.

Over a decade of efforts focusing on reform in statistics education from the academic community, primarily in the United States, coordinated mainly by the ASA, the MAA, and the National Science Foundation (NSF), have culminated in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) project (http://www.amstat.org/education/gaise/GAISECollege.htm). In 2005, the ASA endorsed the GAISE recommendations which are considered to be a basis for statistics education reform in the introductory college statistics courses (Hassad, 2008). According to the GAISE recommendations, there is a need for statistics educators to stress conceptual understanding rather than mere knowledge of procedures in teaching and learning of introductory statistics. With the emphasis placed on conceptual understanding, there should be fewer procedures, recipes, calculations, and derivations (Lovett & Greenhouse, 2000; Turegun & Reeder, 2008).

It is more important for students to grasp the reasoning of inference than the number of different inferential procedures students are taught, and some of these procedures might even be eliminated from an introductory statistics course. Moore (1997) pointed out that few students of introductory statistics might need to interpret analyses of variance (ANOVAs), fewer might need to carry them out computationally, either by manual or electronic means, and even fewer might need to understand the details behind the ANOVA software. He further argued that treating or designing an introductory statistics course as a first course to educate or train future statisticians might not be the right approach. Therefore, teaching some of the traditional topics could be too narrow and specific for a general body of students in introductory statistics courses.

The GAISE recommendations also articulated the point of view that numerous introductory statistics courses contain a great deal of material with a collection of ideas which are presented disjointly by instructors, understood superficially and forgotten quickly by students. The consensus among the contributors to the GAISE recommendations indicated that there was very little value in knowing a set of procedures if students do not understand the important underlying concepts.

In their literature review paper, Garfield & Ahlgren (1988) documented the nature of difficulties that students at the precollege and college level had in understanding probability and statistics. One of their many suggestions was to teach descriptive statistics alone without relating it to probability. They indicated that students' knowledge of or ability to master a computational probability rule did not imply an understanding of the underlying probability concept. Several misconceptions about probability might still persist, even though students were able to use appropriate terminology or formulae to answer questions correctly on a traditional test. Another recommendation they made was that researchers should consider how useful ideas of statistical inference could be taught from a conceptual point of view.

Rather than focusing on eliminating certain topics, Chance & Rossman (2001) debated the pros and cons of alternative ways of sequencing topics in an introductory statistics course. Their paper focused on the presentation order of the following four issues: data analysis versus data collection; descriptive analysis for bivariate data versus inference procedures for one variable; inference for proportions versus inference for means; and tests of significance versus confidence intervals. They provided two sets of arguments, both in favor and against, for each issue. What emerged from both sets of arguments were several common goals for reforming the content of introductory statistics courses. Identifying correctly, introducing early and revisiting often of the central ideas, building connections among different ideas, emphasizing common elements of analysis or interpretation, and minimizing time devoted to mathematical details were some of the points of agreement that emerged from the arguments presented in their paper.

In an attempt to explain and understand why reform efforts work, Lovett & Greenhouse (2000) presented five principles of learning based on cognitive theory, and applied those principles to an introductory statistics course design at Carnegie Mellon. Those principles were stated as: students learn best what they practice and perform on their own; knowledge tends to be specific to the context; learning is more efficient when students receive real time feedback; learning involves integrating new knowledge with existing knowledge; and learning becomes less efficient as the mental load students carry increases. The newly designed course goals no longer included derivation of statistical formulae or computation of certain statistics by hand, but instead included statistical literacy and ability to reason statistically about real-world problems (Lovett & Greenhouse, 2000). In organizing their new course content, they placed the emphasis on students' conceptual understanding and practical use of statistical reasoning rather than memorization of statistical formulae and procedures. A decade earlier, Wheatley's (1990) mathematics education research resulted in similar recommendations. He advocated a more problem-centered content focus, and an emphasis on practical and contextual use of mathematics.

Pedagogy Related Changes

Pedagogy can be broadly viewed as what we, as statistics educators, do to help students learn. Traditional pedagogic practices currently used in teaching introductory statistics courses is, to a large extent, based on the lecture-and-listen model. In this model, students are conceived to be passively listening to lectures so that their empty minds can be filled with information transmitted directly from a central knower. In our rigid adherence to the lecture-and-listen model, we, as educators, perpetuate the futile cycle of teacher making deposits and students receiving, memorizing, and repeating. We, as educators, need to remember that "we overvalue lectures" (Moore, 1997, p. 125).

We also need to remember that, as teachers, we tend to "underestimate the difficulty students have in understanding basic concepts of probability and statistics" (Garfield, 1995, p. 31). Based on a literature review, related to difficulties that students have in understanding probability and statistics, Garfield & Ahlgren (1998) concluded that a large proportion of students had no understanding of many of the basic statistical concepts they had studied. The difficulty of subject matter combined with the passive and oppressive character of the traditional teaching and learning practices based on lecture-and-listen model has been one of the leading causes of frustration and dissatisfaction experienced by students and faculty with the course over the years. The widespread frustration and dissatisfaction with the course have led researchers to investigate alternative models, and offer recommendations in order to reform the pedagogies and practices employed in introductory statistics.

The most common factor among pedagogical changes suggested was a move away from the traditional lecture-and-listen model toward activity-based courses which promote, and support active participation and interaction among all participants. Moore (1997) considered asking students to work in groups cooperatively, and having students communicate their findings orally and in writing to be necessary components of good pedagogy. He further stated that "the core of new pedagogy is genuinely active learning" (Moore, 1997, p. 130).

In an attempt to either supplement or replace traditional lectures with active learning activities, Garfield (1993) described the use of cooperative learning activities in teaching and learning of statistics. She identified different ways of using cooperative learning activities and outlined the reasons for implementing this type of an approach. Also given in her paper were the characteristics of good activities and some suggestions for evaluation of group work.

Cooperative learning, which falls in the more general category of collaborative learning, can be described as working in groups and mutually searching for understanding to improve one's own and each other's learning in the process. Garfield (1993) argued that since small group learning activities might be designed specifically to allow students opportunities to actively and individually construct their own knowledge, as opposed to copying down knowledge transmitted, the use of small group activities was aligned with the constructivist theory of learning. Among some of the characteristics of the activities she deemed as worthwhile, she included getting all members involved, not allowing only one or two students to always do the work, and giving very clear instructions for the activity to minimize confusion among students about what to do.

Some cooperative learning activities that worked reasonably well in classes of 40 to 100 students were also reported by Keeler & Steinhorst (1995). Data consisting of final grade distributions and the number of students retained indicated that the final scores and the retention rates in two experimental sections were higher than the final scores and the retention rates in a comparison section. They also collected responses on a questionnaire that was designed to assess students' attitudes towards the group activities. The responses analyzed suggested that students were "more engaged in the course material and learned the material better when involved in collaborative groups than when presented with traditional lectures" (Keeler & Steinhorst, 1995, paragraph 19). In a follow-up paper, Steinhorst & Keeler (1995) realized that the traditional training of many statistics instructors might be a road block in developing active learning material. They provided a set of representative active learning material to help the traditionally trained statistics teachers overcome the challenge of teaching introductory statistics courses from a more conceptual and active learning viewpoint.

Making changes from traditional lectures to active learning techniques were also viewed as considerable challenges to many statistics instructors by Garfield & Ben-Zvi (2008). The traditional training of many statistics instructors could be one of the reasons which might make the change from a teacher-centered approach of lecture-and-listen format a difficult and challenging one for them. Another reason for the reluctance to make the change was identified as the ease with which statistics instructors can prepare a lecture. Contrary to preparing a lecture, designing a student-centered learning environment where "students engage in activities, discussions, and collaborative projects supported by technological tools" is much more difficult, challenging, and time consuming (Garfield & Ben-Zvi, 2009, p. 74).

In an effort to alleviate such difficulties and challenges, Roseth, Garfield, and Ben-Zvi (2008) provided practical examples of how statistics educators might apply a cooperative framework to teaching and learning statistics. Based on the premise that statistics education ought to resemble the inherently cooperative nature of the practice of statistics, their paper described the specific ways of incorporating corroboration into teaching and learning of statistics. They hoped to address the concerns of some statistics educators who might be reluctant to change or move away from teacher-centered approaches toward more student-centered approaches. Roseth et al. (2008) advocated collaboration among statistics educators, as well as among students. They pointed out the potential of collaboration among statistics educators to enhance and sustain the efforts of cooperative teaching and learning of statistics in the classroom.

The proper choice of curricular material and textbooks that emphasize and structure activities through which to illustrate concepts may also alleviate difficulties and challenges faced by the statistics instructors in making the pedagogical changes. Keeler & Steinhorst (1995) identified a textbook written by Freedman, Pisani, and Purves (1978) as the first major introductory statistics textbook emphasizing concepts. Moore (1997) described the text titled Workshop Statistics written by Rossman (1996) as another innovative textbook with a collection of learning activities. Currently, these two textbooks are available in their latest editions. The introductory statistics course that I teach uses the 3rd edition of Workshop Statistics as the required textbook.

An empirical study conducted by Giraud (1997) revealed that students in a class that used in-class cooperative groups to work on assignments had higher test grades than students in a lecture class. Giraud used the test scores, obtained from two sections of an undergraduate statistics course, as dependent variables to examine the relative effects of cooperative versus lecture methods of instruction. The students in the cooperative class completed assignments and practice problems in groups during class. The students in the lecture class completed assignments and practice problems individually, outside of class. Giraud reported that the students in the cooperative learning class achieved higher test scores. Magel (1998), who implemented cooperative groups in a large lecture class, also found improvement in students' test scores compared to previous semesters.

In an effort to implement changes in pedagogy, Smith (1998), who is a firm believer in learning statistics by doing statistics, modified a traditional introductory statistics course by incorporating into his lectures a set of biweekly out-of-class projects with written and oral reports of the results. He reported dramatically improved test scores and overwhelmingly positive student assessments of the new approach.

I argue that the studies using cooperative learning activities may need to provide more descriptive details and more qualitative information regarding the exact nature of the group dynamics and organization of the student groups in order for the reader to be able to evaluate whether what was implemented could be considered a true cooperative learning experience for students. As one examines the available related literature, it is also important to realize what cooperative learning is not. As Garfield (1993) pointed out, a genuine cooperative learning experience is not having students merely sit side-by-side at the same desk or table and talk with each other as they do their individual assignments, or having students do a task individually and then have the students who finish first help the slower students.

Although some traditionally oriented statistics instructors tend to believe that increasing interaction and active learning in our classrooms might cause them to "cover" less material, I believe this can be taken as an opportunity to focus on "big" ideas, to go deeper with those ideas, and use them as threads to weave a curriculum matrix for introductory statistics (Palmer, 2007; Turegun, 2009; Whitehead, 1929). When examined from a paradoxical perspective, teaching less may lead to learning more. Hence, less is more. Arguments in favor of active learning, which might be perceived by traditional statistics instructors as an implication for a decrease in the material coverage, are actually attempts to increase learning by focusing on the "big" ideas of statistics.

Technology Related Changes

There are various technological tools which can be used to support the development of conceptual understanding and reasoning abilities of students. Computers, graphing calculators, the Internet, statistical software packages and Web applets are among the several forms of technology available to statistics educators. Several researchers have explored the use of these different forms of technology to improve teaching and learning of introductory statistics. For the most part, the decision regarding what form of technology to use might be dependent upon the issue of accessibility by students.

Using computers or calculators merely to generate statistics, to follow algorithms or to produce graphs of data are very limited views of technology use in statistics education reform. These types of limited technology use do not tend to extend beyond the notion of, what I refer to as, "using technology for the sake of using technology."

Technology use in that sense becomes a tool for doing statistics, not a tool for learning statistics (Meletiou-Mavrotheris, Lee, and Fouladi, 2007; Moore, 1997).

Using technology to help students visualize concepts and understand abstract ideas is considered to be far more important than using technology solely to automate messy statistical computations (Garfield & Ben-Zvi, 2009). For example, simulating drawing samples from various populations and observing the distributions of statistics computed from these samples are better ways of illustrating the Central Limit Theorem than providing a mathematical proof for it. In calculus reform, the interest in technology was focusing on off-loading the burden of calculation to technology so that students could focus more on learning the conceptual ideas. This is still controversial because we have not done anything to increase conceptual understanding, but rather simply took the computational burden away (Lovett & Greenhouse, 2000).

In a survey to evaluate the impact of statistics education reform efforts on the teaching of introductory statistics, Garfield et al. (2002) collected 243 statistics instructors' responses regarding particular uses of technology, along with other issues such as teaching methods, assessment methods, and faculty views on reform efforts. More than two-thirds of the faculty surveyed, from the mathematics, statistics and other departments (i.e., psychology, business, sociology, and economics) of two-year colleges, four-year colleges, and universities, reported making changes in their courses over the past few years. The most common changes made were the increased use of technology, followed by pedagogical changes, content related changes, and lastly, assessment related changes. While the survey results indicated a considerable increase in the usage of technology, many instructors were unaware of excellent Web resources such as applets.

One of the findings regarding technology use reported by Garfield et al. (2002) was the increased use of graphing calculator technology.

The increased use of graphing calculators generated some interest among researchers. For example, Collins & Mittag (2005) reported on a study that explored the effects of using graphing calculators on students' performance on an exam over hypothesis tests and confidence intervals, and on the final exam scores. The two introductory statistics sections, one with 22 students using graphing calculators capable of inferential statistics, and the other with 47 students using non-inferential graphing calculators, were taught by the same instructor. It was reported that there was no observed significant difference between the two groups of students in their performance on examinations on inferential topics in an introductory statistics course, once the adjustment was made for performance on previous tests.

The use of graphing calculators has certain advantages such as portability and suitability to active participation. However, several researchers expressed concerns regarding the use of graphing calculator technology (Forster, 2006; Moore, 1997). Limited amounts of data entry and small screens with static graphs are listed as some of the weaknesses of graphing calculators.

Although initial uses of technology were limited to or focused on the computational power only, the uses of technology seem to be shifting toward the conceptual power (Garfield & Ahlgren, 1988). Technological tools such as simulations and applets have become increasingly common in introductory statistics courses to illustrate abstract concepts, simulate data, and build understanding of statistical concepts (Chance & Rossman, 2006; Zieffler & Garfield, 2007). Instead of using calculators to generate z-scores, it is possible to have students explore the empirical rule of 68-95-99.7 by using a Normal distribution Java applet (Everson, Zieffler, and Garfield, 2008). The applets have been gaining an increasing importance because of their effectiveness in illustrating various statistical concepts visually. Everson et al. (2008) suggested another applet for having students examine how variability in bivariate data affected the magnitude of the correlation coefficient. They identified the website http://www.causeweb.org as a good resource for a variety of applets. The use of technology to quickly create sampling distributions from different populations for helping students develop conceptual ideas of the Central Limit Theorem was also a suggestion provided in their paper.

Likewise, several other researchers have found that the use of simulation and applets to make abstract concepts more concrete was an effective learning tool (Chance & Rossman, 2001; Chance & Rossman, 2006; Glencross, 1988). In their debate based paper Chance & Rossman (2001) were in complete agreement that the concepts of randomness, confidence, and significance should be introduced to students through the use of simulations and applets. However, they indicated the importance of having students perform physical simulations first with hands-on manipulatives such as coins, dice and cards prior to the use of the calculator or computer.

Having students perform physical simulations first with hands-on manipulatives is especially important in light of warnings of Lane & Peres (2006). They noted that one of the common ways to use a simulation was as a demonstration in class. In their study conducted at Rice University, they used a simulation activity to demonstrate various characteristics of a sampling distribution such as the relationship between sample size and the standard error of the mean. They reported that, although both the instructors and the students felt that the class was excellent, one-third of the students produced the wrong answer to a simple question about the effect of sample size on the standard error of the mean.

A similar discovery was made by Lunsford, Rowell, and Goodson-Espy in their (2006) study. They reported a conclusion that the use of a computer simulation, only for demonstration purposes, was not sufficient for developing deep graphical understanding of concepts. Unfortunately, using a simulation or applet for the sole purpose of demonstration by instructor in front of a class with students being only passive learners does not ensure active learning, and can in fact lead to poor learning.

Along the lines of using technology with simulations, in an earlier study, delMas, Garfield, and Chance (1999) examined the development of students' understanding of sampling distributions using a simulation program and research-based activities. Their research was conducted in collaboration among the three authors at their respective institutions and in three different classroom settings. delMas et al. (1999) made a strong case for using technology with simulations in a predict-and-test environment to create cognitive dissonance in developing students' understanding of sampling distributions. They reported that an activity, which asked students to test their conjectures and confront their misconceptions, was found to be more effective than one based on guided discovery. Based on their initial findings, delMas et al. (1999), similar to Lane & Peres (2006), concluded that while software could provide the means for a rich classroom experience, computer simulations alone did not guarantee conceptual change. They indicated a need for a conceptual change model of learning which suggested having students make their own predictions and then test them out by using simulations.

The conceptual change theory had been used previously in the field of science education and is being used by other researchers in social studies education (Posner, Strike, Hewson, and Gertzog, 1982). According to the conceptual change theory of learning, students who have misconceptions or misunderstandings need to experience an anomaly or go through effective discrediting experiences which would create cognitive dissonance between their expectations or predictions and observed outcomes. Once they revised the assessment instruments accordingly, they found that student performance improved as the students were asked to make and test conjectures about different empirical sampling distributions from various populations.

Technology, in general, is viewed to serve content and pedagogy. The use of graphing calculators ties in with and encourages active participation. But, to a certain degree, technology has changed content and makes possible or allows new forms of pedagogy. Even though not all statistics teachers agree on what is simply a rule, automating anything that is simply a rule is considered good pedagogy (Moore, 1997). Yet, the of use technology in statistics education reform goes beyond the simple automation of rules. What makes the use of technology an effective learning tool, in addition to computing and producing static graphs, is the capability of technology to illustrate various statistical concepts visually and make abstract concepts more concrete.

Assessment Related Changes

The traditional method of assessing student learning consists of module tests, generally with multiple-choice questions, designed for ease of grading. The traditional exam questions place a strong emphasis on the procedural or computational aspects of statistics and do not evaluate high-level cognitive and conceptual understanding of students (Cobb, 1993; Garfield, 1994; Zieffler, Garfield, Alt, Dupuis, Holleque, and Chang, 2008). These traditional testing methods tend to evaluate the rate of defects in the final product. The connection of the traditional testing methods to the actual statistical practice is an unexamined assumption. The efforts of statistics education reform in assessment, which is sometimes referred to as “authentic assessment”, starts with challenging and questioning this assumption. The statistical education reform efforts have responded to this challenge and critically evaluated the traditional testing practices to offer some alternative forms of assessing students’ statistical reasoning and conceptual understanding (Chance, 1997; Gal & Garfield, 1997; Garfield, 1994; Garfield & Gal, 1999).

As the statistics education reform efforts begin to change the teaching and learning of introductory statistics at the college level, there is even a greater need for appropriate assessment materials to evaluate whether the changes help improve students’ statistical reasoning and conceptual understanding (Garfield, 1994). As a result, in order to adequately address and respond to the reform efforts, we must also search for new and innovative ways of assessing what our students know about statistics. Most statistics teachers tend to view assessment as separated from teaching, and as limited to testing, grading exams, quizzes and homework assignments. This traditional view of assessment and its related forms are considered to be too narrow and too specific to provide useful information about student learning (Garfield & Gal, 1999). Even worse, several misconceptions about statistical topics, such as probability, may still persist, even though students are able to use appropriate terminology or formulae to answer questions correctly on a traditional test (Garfield & Ahlgren, 1988).

The GAISE recommendations outline an emerging view of assessment as an ongoing evaluation of students' learning over the course of the semester with constant gathering of information and providing feedback. The use of this emerging view of assessment to guide assessment practices in introductory statistics courses can be very valuable in informing our teaching. In this emerging view of assessment, teaching and assessment no longer appear to be in a dichotomous relationship, but rather in a continuous cyclical relation of informing one another with the ultimate goal of improving student learning. As part of the changes related to assessment, instructors are encouraged to collect a variety of assessment information from sources other than individual student tests, the results of which traditionally were used to assign grades and rank students.

Among some of the alternative forms of assessment were cooperative group activities, computer lab exercises, portfolios, projects/reports, presentations, essay questions, journal entries, and open-ended writing assignments (Chance, 1997; Garfield, 1994; Onwuegbuzie & Leech, 2003). These alternative forms of assessment may be structured to provide some rich information in assessing the nature of student learning. Walking around the class to observe students as they work in small groups on an activity, and having students explain their answers are some of the ways to informally assess students' statistical reasoning. Being able to hear students express their understanding of what they have learned is important because it provides teachers with an ongoing, informal assessment of how well students are learning and understanding statistical ideas. Written reports on group activities are useful sources of information in assessing students' ability to solve a particular problem, apply a skill, demonstrate understanding of an important concept, or use statistical reasoning.

If the reform efforts focus on teaching our students what we, as statisticians or statistics educators, value most, then we must also assess what we value most. Chance (2002) refers to this as "the number one mantra" to remember when designing assessment instruments. On a final note on assessment, I, in complete agreement with Ben-Zvi (1999), have become accustomed to regarding assessment as being a continual and recursive process, as opposed to being a sporadic and conclusive one; students as being active participants in this process, as opposed to being the objects of the assessment; and assessment outcomes as an opportunity for all students to achieve their potentials, as opposed to filtering and selecting students out of the opportunities to learn statistics.

In conclusion, the changes proposed by statistics educators and researchers to reform introductory statistics education can be gathered in the following four categories: changes related to content, pedagogy, technology, and assessment. I regard these changes as the four pillars of statistics education reform. We, as statistics educators, need to be cognizant of the tension among these four pillars while pursuing our statistics education reform efforts.

My Theoretical Orientation and Beliefs on Teaching and Learning

From a holistic stance, the researcher can also be viewed as a subject in the process of generating new knowledge. In my triple role as the learner, teacher and researcher during the study, I was engaged in creating new knowledge and learning form the process along with my students. Since the process of reflection-action-reflection cycles were shaped, in part, by factors such as my theoretical orientation and my beliefs on teaching and learning, I will present a synopsis of those factors next.

Inspired by Dewey's views of knowing and knowledge as a product of continuous cycles of action and reflection, I understand teaching as a reflective practice with the ultimate goal of improving teaching and learning. Relying on iterative reflection-action-reflection cycles as a developmental approach, I aim at improving my teaching practice, students' learning, and my understanding of the specific educational contexts in which I teach. Then, the improved understanding of practice can be used to inform theory and contribute to the knowledge base of teaching and learning.

Having taught at the community college level for a number of years, I am keenly aware of the socio-economic, cultural, educational, and historical background differences between my students and myself. Despite these differences, I try to establish at least some commonalities with my students. It is these commonalities that help me better identify and form a strong bond with my students. Additionally, I seek opportunities to interact further with my students by promoting a community of learners based on mutual respect and trust.

I am also inspired by the ideas of complexity theory in envisioning and creating such a community. Complexity science points out that hardly any event or activity can be reduced to being just a "thing" in isolation. I make a strong commitment in my teaching to an open and inclusive pedagogical approach which values, seeks, and welcomes students' contributions to the classroom community of learners, where I am but one voice among many. By providing emotional comfort and intellectual safety, I aim to promote an environment in which sustained interaction, conversation, and discussion take place while being respectful to and responsible for one another. I believe in such an environment students can present themselves and express their ideas freely and naturally without the trepidation of insensitive ridicule or the apprehension for dismissive remarks.

Weaving the fabric of community of learners with these values used as strong cohesive threads, I believe, encourages students to respond to one another's ideas rather than responding directly to me, and developing an appreciation for the strength of diversity in a community of learners. From my constructivist and social constructivist perspective, shaped, for the most part, by reading the works of Piaget, von Glasersfeld, and Vygotsky, establishing a strong community of learners paves the way for students to construct their knowledge by way of negotiated meanings in relation to their past experiences through a web of social, cultural, and historical interactions.

In addition to the views outlined in the proceeding paragraphs, I hold particular thoughts which are reflected in my adherence to a set of certain principles in the following four broad and interconnected areas: content, pedagogy, technology, and assessment. Next, I will articulate my beliefs on these four areas in order to depict another dimension of my teaching philosophy. In terms of content, teaching and learning or mathematics should not be about memorizing formulae and identities. I place more of an emphasis on students' conceptual understanding and practical use of mathematical reasoning rather than memorization of formulae and procedures. I try to achieve that by using a more problem-centered content focus and an emphasis on practical and contextual use of mathematics. In the courses I teach, I try to introduce the few and important main ideas, and investigate these main ideas by throwing them into every imaginable possible combination, as Whitehead recommended nearly a century ago.

As far as pedagogy is concerned, traditional pedagogy for teaching mathematics courses, based on lecture-and-listen model, views students as passive learners. In our rigid adherence to the lecture-and-listen model, I believe, we overvalue lectures. I create my own curricular supplements that emphasize and structure activities to support activity-based courses which promote active participation and interaction among all participants. I ask students to work in groups cooperatively through these activities. I also have students communicate their findings orally in the form of short presentations in my calculus and statistics classes, and in writing in the form of semester-long portfolios or reflective journals. As students reflect on the different topics of the course, I encourage them to support their ideas and arguments with evidence. Frequent feedback is provided on students' reflections, through written comments and class discussions, in order to further facilitate their development as learners. The journals, portfolios, and short presentations serve not only as an ongoing assessment for me in order gain insights into my students' evolving understandings of mathematics and statistics, but also to help me make informed decisions about how to further support their learning.

My beliefs on the use of technology in the classroom focus on using technology not as a tool for doing mathematics or statistics, but as a tool for supporting the development of conceptual understanding and reasoning abilities of students. I believe using computers or calculators merely to generate statistics, to follow algorithms or to produce graphical representations of data are very limited views of technology use in mathematics and statistics. Using technology to help students visualize concepts and understand abstract ideas is far more important than using technology to solely automate messy computations or algorithms. For example, in my statistics classes, I have my students use applets in order to simulate drawing samples from various populations. Having students observe the distribution of statistics computed from these samples make an abstract concept more concrete and is a better way of illustrating the Central Limit Theorem than working through a mathematical proof for the theorem.

Traditional assessment practices are, for the most part, deficit-based evaluative procedures administered in a compartmentalized discrete fashion. These types of practices place a strong emphasis on the procedural or computational aspects of mathematics and statistics, and do not necessarily, in a useful way, inform us about the high-level cognitive and conceptual understanding of students. I believe teaching and assessment are not in a dichotomous relationship, but rather in a continuous cyclical relation of informing one another with the ultimate goal of improving student learning, as well as my teaching. In a continuous search for new and innovative ways of assessing what my students know about mathematics and statistics, I collect a variety of assessment information from a variety of sources, such as cooperative group activities, computer lab exercises, portfolios, projects/reports, presentations, essay questions, journal entries, and open-ended response assignments.

In conclusion, my views on the areas of content, pedagogy, technology, and assessment, collectively, have led me to regard these four areas as the four pillars of mathematics and statistics education. We, as educators, need to be cognizant of the tension among these four pillars while pursuing our mathematics and statistics education research efforts. I continually aim to improve my teaching by reflecting on and assessing my own teaching practices. I exchange ideas with colleagues at professional conferences or informal meetings about teaching mathematics and statistics. On a final note, although my passion is for mathematics and statistics in particular, a more general aim of my teaching philosophy is to instill in my students the value of being self-motivated life-long learners and thinkers.

CHAPTER THREE

RESEARCH DESIGN AND METHODOLOGY

The purpose of this study was to examine the individual perceptions and conceptual understandings that community college students have of statistical measures of spread in an introductory statistics course. Conducting the study in the classes I was teaching meant that I could assume the role of the teacher and researcher. I followed an action research model to investigate students' conceptual understanding of measures of spread. In order to gain a broader and deeper insight into to my research questions, I gathered and analyzed both quantitative and qualitative data. The study aimed at answering the following research questions:

1. What are some of the pre-existing conceptions of students regarding statistical measures of spread?
2. How do students articulate their conceptual understanding of statistical measures of spread?
3. How do students establish the connections between statistical measures of spread and behavior of distributions of data sets?

In order to address these questions, this chapter first provides a rationale for the choice of action research model, and the suitability of embedding both quantitative and qualitative approaches. Then, the chapter offers an outline of the data collection procedures and a brief overview of the data analysis process. A contextual background along with community college and instructional settings for the study is also provided in this chapter. Finally, it concludes with a summary of the chapter.

Action Research

I concur with a number of researchers' views that the research questions are to guide the choice of a particular research methodology that leads to answer those questions in the best possible way (ASA, 2007; Johnson & Onwuegbuzie, 2004; Wilson, 1977). According to Dewey's (1933) view, inquiry can be depicted as a process linking reflection and action in a unified way for the creation of new knowledge in a given context. Action research treats inquiry as a process that is not fragmented or separated (Greenwood & Levin, 1998). Hence, from a holistic stance, the researcher or the inquirer is also viewed as a subject in the process of generating new knowledge. My triple role as the learner, teacher and researcher in the study led me to choose action research. The term action research is often used interchangeably with teacher research or classroom-based research in the literature. Classroom research, although initially conducted in elementary and secondary classrooms has, in recent years, become recommended for college level research (Cross & Steadman, 1996).

Even though action research has grown in popularity and gained acceptance as legitimate research, it has been criticized by some researchers as a sub-standard form of research and in the academic circles of conventional social scientists generally has been either degraded as unscientific or treated with disrespect (Greenwood & Levin, 1998). The main points of criticism are focused on teachers. The lack of qualification on teachers' part to conduct research properly and teachers' time taken away from the main task of classroom teaching are expressed as primary concerns. These points of criticism reflect a view that considers teachers as mere implementers of tightly scripted lesson plans and as disempowered individuals to critically examine their own practice.

However, when an individual or a group of individuals is involved in action research, which requires a systematic inquiry into their practices, the individuals get an opportunity to reflect on their practices and act accordingly in order to improve the status quo conditions, frame problems, and find solutions to problems they face (Kang, 2007). If learning is viewed as constructing meanings based on individual experience, then, by using an analogical extension, one can view teaching as a reflective practice (Schön, 1983). Inspired by Dewey's (1933) views of knowing and knowledge as a product of continuous cycles of action and reflection, Schön identified reflection-in-action and reflection-on-action as two reflective processes which can be used to enhance understanding. While the process of reflection-in-action can be viewed as a way or the ability to assess actions in the process of acting, the process of reflection-on-action can be considered as working through experiences gained from actions after the fact (Greenwood & Levin, 1998; Schön, 1983).

Since action research studies provide educators an opportunity to regularly assess and alter their instructional activities, to reflect on, improve, and use those practices to inform theory, several statistics education researchers have used action research model in their own work (Chance, Garfield, and delMas, 2000; delMas et al., 1999). Realizing the powerful and useful notions of a continuous cycle of praxis embedded in action research, and heeding the call of previous statistics education researchers such as delMas et al. (1999), I used an action research model to gain insight into community college students' conceptual understandings of measures of spread, the related problems and the sources of the problems. In this particular approach to action research, teachers examine their instructional practices in their own classrooms via the iterative and dynamic processes of reflection-in-action and reflection-on-action (Ball, 2000; Feldman & Minstrell, 2000; Greenwood & Levin, 1998; Kang, 2007; Noffke & Stevenson, 1995).

The proposed action research model with its action-reflection-action cycle is a developmental approach aimed at improving teachers' teaching practice in order to improve students' learning, as well as improving teachers' understanding of the specific educational contexts in which they teach. Then, the improved teacher practice can be used to inform theory and contribute to the knowledge base of teaching and learning. The strong context-bound nature of the knowledge generated in action research studies is recognized as one of the core characteristics of action research.

Even though action research intends to interpret what is observed in the classroom of the teacher/researcher within a particular context, the knowledge produced in one particular context can be valuable in another context as well. If a reader needs to transfer knowledge from one context to another one, then the researcher has the responsibility of providing the reader with detailed contextual factors. The reader, in turn, has the responsibility of developing an understanding of the given contextual factors in which the research was conducted, comparing and contrasting the two sets of context, and making an informed judgment about the suitability of linking the two sets of context before applying knowledge produced in one context to the next.

Techniques such as triangulation of a variety of data sources were employed in order to ensure and increase validity of the interpretation of the classroom data. Murphy (1989) used the phrase simultaneous triangulation to indicate that even though there might be limited interaction between quantitative and qualitative data during the data collection, the findings could complement one another at the end of the study. A form of simultaneous triangulation was achieved by conducting semi-structured student interviews in the form of mid-term conferences (MCI), a closed book journal assessment with an open-ended question (CBJ), weekly student journals (WJ), and online multiple-choice CAOS measures of spread (MOS) pre- and post-assessment. These items were regular components of the course assessment and used to solicit students' thoughts, understandings, and experiences in various ideas and topics, including but not limited to measures of spread, discussed throughout the course. The assessment structure used in the course along with a set of guidelines for successfully completing the course is given in Appendix A.

Embedding Quantitative and Qualitative Approaches

The seeds for the idea of integrating quantitative and qualitative approaches in order to enrich educational research considerably were being planted more than thirty years ago (Wilson, 1977). According to Batanero, Garfield, and Ottaviani (2001), many of the important research studies in statistics education do not necessarily involve the collection and analysis of solely quantitative data; and these studies instead, collect qualitative data in the form of clinical interviews and classroom observations. In recent years, mathematics educators have moved away from purely quantitative methods and relied heavily on qualitative methods. However, the recent reports by the American Statistical Association (ASA), recommend finding a middle ground between quantitative and qualitative methods (ASA, 2007).

The above stated research questions are best answered by an expansive and creative form of research, not a limiting form of research methodology situated in one paradigm only. Several researchers shared the view that "research methods should *follow* research questions in a way that offers the best chance to obtain useful answers" (Johnson & Onwuegbuzie, 2004, p. 17). Many research questions are best and most fully answered through mixed research solutions. For instance, adding interviews with a purposefully selected sample from a quantitative study would allow a researcher to gain additional insight into students' understandings, perspectives, and meanings. According to Johnson & Onwuegbuzie (2004), the reverse ordering might also be useful. As an example, in a qualitative research study the researcher might consider a qualitative observation or an interview, but compliment it with a quantitative instrument. The quantitative results of my research, in addition to identifying pre-existing conceptions, were used to provide a foundation and background for a more in-depth analysis of the qualitative results.

If the research findings are corroborated across different approaches, then greater confidence can be placed in the subsequent conclusions. On the other hand, if the findings are in conflict, then the researcher has still greater knowledge and can modify interpretations and conclusions accordingly. The goal of mixing is not solely to search for corroboration but rather to expand the researcher's understanding of the data and the findings (Onwuegbuzie & Leech, 2004). I used a mixed-model design by mixing qualitative and quantitative approaches within and across stages of my action research (Johnson & Onwuegbuzie, 2004).

Having acknowledged the importance of both quantitative and qualitative approaches and data types for this study, I identified the above mentioned data sources not only to achieve a form of simultaneous triangulation but also to generate a combination of qualitative and quantitative data in helping me conduct my research with an added scope and breadth. Hence, I chose to embed both quantitative and qualitative approaches in an action research model to investigate students' conceptual understanding of measures of spread in an introductory statistics course at a community college.

Setting

The fact that almost half of the university and college students and more than half of all first time college freshmen attend community colleges nationwide combined with the open admission system of community colleges make this particular student population very unique in many ways. The student body at community colleges is quite large, diverse, and considerably different than the traditional student body found at many of the nations' colleges and universities.

Research on teaching and learning of introductory statistics at the community college level where students are, in general, educationally underserved might provide unique insight into the teaching practices of community college instructors. This particular research study was based on such a community college instructor and his students as they were engaged in a collaborative effort to develop conceptual understanding of measures of spread.

The Community College and the Instructional Setting

The study took place at a mid-size, urban, and single-campus based community college which is situated on a 143-acre site located within the southwest section of a city in the Midwestern region of the United States. Initiated by the city's chamber of commerce, the community college started its first academic year with an enrollment of 1,049 students on September 25, 1972 as a junior college. As a result of a name change in 1983, the college became referred to as a community college. The current size of enrollment, reported to be nearly 22,000 students by the college's Planning and Research

Department, ranks the community college as the second largest community college and the fourth largest institution of higher education in the state.

The college offers 36 Associate in Arts and Associate in Science degree programs, and 24 Associate in Applied Science degree programs. In addition to the student population enrolled in these degree programs, there are also students who are enrolled in transfer courses to four year college or universities and some other students enrolled in technical/occupational tracts and 18 Certificate of Mastery programs in various fields. While this institution attracts students internationally and nationally from many cultures, ethnicities, and backgrounds, a great majority of students is from within the state. The student body largely consists of Caucasian students, with slightly over 32% merged by the college's Department of Planning and Research into a single category of Ethnic Minority.

In addition to introductory statistics courses offered by other departments at the college, a multi-section introductory statistics course is offered also by the mathematics department. The course is listed at the sophomore level. The student enrollment is limited to thirty-five students per section for all of the on-campus, face-to-face and on-line sections of the introductory statistics courses offered by the mathematics department. While the introductory statistics course is offered as a sophomore level course, the prerequisite for the course is not listed as a college level mathematics course, but it is rather listed as a remedial or non-college credit intermediate algebra course.

The generic departmental course syllabus provides a list of topics which is not atypical for a college level and non-calculus based introductory statistics course. Among the listed topics are descriptive statistics, graphical displays of distributions, sampling distributions, central limit theorem, probability, normal probabilities, confidence intervals, hypothesis testing, correlation, and regression. The course focuses on active learning and lists the third edition of Workshop Statistics - Discovery with Data and the Graphing Calculator by Rossman, Chance, and Von Oehsen as the required textbook. The textbook deemphasizes computational aspects of statistics and favors conceptual understanding. In addition to the required textbook, the course also requires a TI-83/84 graphing calculator.

Participants

While there are other introductory statistics courses offered at the college by business and psychology departments for the students who major in those fields of study, the students who enroll in the introductory statistics course offered by the mathematics department have a wide variety in the backgrounds and in their major fields of study. The major fields of study of the students in the introductory statistics course included, but not limited to, Nursing, Elementary Education, Early Childhood Education, Biology, Psychology, Political Science, Marketing, Criminal Justice, and Engineering. The open-enrollment practices at the community college, the wide variety of major fields of study, and a non-college credit intermediate algebra prerequisite for the course result in considerable differences among students in terms of their mathematics skills. Teaching introductory statistics to such a diverse body of community college students is not an easy mission.

I taught two sections of the introductory statistics course during a spring semester. These two sections were scheduled to meet on Monday and Wednesday from 1:00 pm to 2:20 pm, and on Tuesday and Thursday from 1:30 pm to 2:50 pm for fifteen weeks. In these two sections, there were a total of twenty-nine students of which twenty-two were female. Various course assignments and assessment items used in this study were regular components of the course work outlined in the course information and policy, of which the students were made aware on the first day of classes.

Almost a third (ten) of the students were Nursing majors. There were four Elementary Education majors and three Early Childhood Education majors. The other students listed various majors such as Biology, Psychology, Political Science, Marketing, Criminal Justice, and Engineering. Almost half of the students (fourteen) indicated being sophomore students. The distribution of class standings is summarized in Table 1. The class group labeled as "Other" included a college graduate, a student without a response, and a high school senior with a concurrent enrollment status.

Table 1: Class Group Distribution

Class Group	Freshmen	Sophomore	Junior	Other
Number of Students	5	14	7	3

The course load among students in these two sections varied somewhat. All of the students were enrolled in at least 6 credit hours of course work during the semester. Twenty-three of the twenty-nine students were enrolled in 12 or more credit hours, with two students carrying the largest course load of 15 credit hours. The work load also varied among students. Twenty-two students indicated that they worked anywhere from 5 to 50 hours a week. Fourteen students stated that they worked 20 to 30 hours per week. Two students indicated that they worked 50 hours or more per week. There was one student who reported working 5 hours per week.

Most of the students self-reported their college GPAs. Twenty-six students reported their college GPAs, which ranged from 2.4 to 4.0. The average GPA was 3.2 with a standard deviation of 0.5. Eighteen students self-reported their college mathematics GPAs, which ranged from 2.0 to 4.0. The average college mathematics GPA was 3.2 with a standard deviation of 0.5.

Nearly half of the students self-reported their ACT scores. Fourteen students reported their ACT scores, which ranged from 20 to 29. The average ACT score was 23.8 with a standard deviation of 2.5. ACT-M scores self-reported by fourteen students showed greater variation, with a range from 17 to 32. The average ACT-M score was 22.4 with a standard deviation of 4.3.

The students in the two sections ranged in age from the youngest at eighteen to oldest at over forty. The distribution of age groups is given below in Table 2. Many of the students were traditional college age students, who were 25 years old or younger. There were a few non-traditional students who were returning to school after a hiatus of several years. One of the students, who was a high school senior, had a concurrent enrollment status.

Table 2: Age Group Distribution

Age Group	18-20	21-25	26-30	Above 30
Number of Students	1	20	4	4

Regarding the high school and college mathematics course backgrounds, there were two students who indicated not having remembered taking a high school mathematics course, and two additional students did not respond. All but three students self reported having taken a college mathematics course previously. Almost all of the students (twenty-six) reported having taken at least a mathematics course within the last two years; for one student, it had been a number of years, and two students did not respond. The background information regarding the number of years since the last mathematics course taken by students is summarized below in Table 3.

Table 3: Number of Years Since Last Mathematics Course

Number of years	< 1 year	1-2 years	3-5 years	6-8 years	No response
Number of Students	12	14	0	1	2

Students had very limited prior exposure to high school or college level statistics courses. Only two students had prior exposure to a statistics course. One of these two students indicated taking a college statistics course two years ago without a successful completion, and the other student took a high school statistics course three years prior to enrolling into the course. In addition to these two students, one other student mentioned being in an ad-hoc college statistics prep course during the last week of the school year in his senior year in high school. All but three students indicated that they were taking the course because it was required by their respective majors. One of these three students indicated that it might be "possible" that the course was required by his major. The other two students did not state why they were taking the course while the course was not required by their respective majors, one of which was "undecided."

Considering the fact that the student population of the college is reported to be 68% Caucasian, it is not a surprising outcome that nearly three-fourths (twenty-one) of the twenty-nine students in the two sections I taught were Caucasian. There were three students who had Asian or Pacific Islander ancestry. Two students were of Hispanic or

Latin American backgrounds. There were also two students one of whom with Native American, and the other one with African American heritage.

Because of my intertwined role as the teacher, learner, and researcher, I was also a participant in the study. According to Wilson (1977), the quality of many participant observation studies can be in part judged on the basis of answers to a set of questions probing the researchers' ability to move beyond her or his own perspectives. A list of these types of questions was used by Wilson (1977) in judging several studies. Therefore, at this point, I would like to address some of these questions regarding my training and educational background, my previous experience in the field, and my theoretical orientations and personal feelings about relevant issues. I believe the subsequently offered answers to these questions might provide others with some very tentative guidelines by which to judge my research study.

At the beginning of the study, I had been teaching for twenty-three years of which eighteen were at this community college. All my teaching experience was at the community college and university levels. Prior to teaching at this community college, I had taught several mathematics courses at other community colleges and at a university in the state of Oklahoma. I completed my Bachelor's of Science degree in Engineering in 1982. Subsequent to working as an engineer, I started to attend graduate school in 1984 and completed my Master's of Science degree in Geophysics in 1987. While teaching at a community college I enrolled into a Master's of Science degree program in Mathematics. Upon completing my Master's of Science degree in Mathematics in 1991, I started teaching at this community college. In 2007, I began a Ph. D. program in Mathematics Education at the University of Oklahoma.

Since I have taught at this community college for nearly two decades, I was aware of the economic, cultural, educational, social, and historical background differences between my students and myself. Despite these differences, holding a full-time job and being a part-time student myself, I felt I shared at least some commonalities with my students. I believe these commonalities helped me better identify and form a strong bond with my students. Additionally, I believe that my triple role as the teacher, learner, and researcher gave me an opportunity to interact further with my students in promoting a community of learners based on mutual respect and trust. This discourse is, in part, aligned with what Davis & Sumara (2007) describe as an open and inclusive pedagogy. Having acknowledged that establishing such a relationship was one of the essential elements in encouraging my students to reveal their thought processes and let me investigate their conceptual understanding of measures of spread, I paid particular attention to creating an environment conducive to having such relationships.

I was inspired by the ideas of complexity theory in envisioning and creating such an environment. Complexity science points out that hardly any event or activity can be reduced to being just a "thing" in isolation. Teaching, being not an exception, is not reducible to what a teacher does and/both does not do in a classroom (Turegun, 2009). I made a strong commitment to an open and inclusive pedagogical approach which values students' opinions, seeks, and welcomes students' contributions to the classroom community of learners. This open and inclusive pedagogy is described by Davis & Sumara (2007) as being "oriented toward unimagined and not-yet-imaginable possibilities." I aimed to provide emotional comfort and intellectual safety for my students to feel that their ideas were needed and valued by others in our classroom community of learners. As a result, my students could present themselves and express their ideas freely and naturally without the trepidation of insensitive ridicule or the apprehension for dismissive remarks.

In part, carrying out my research study relied on creating such an environment in which sustained interaction, conversation, and discussion took place while being respectful to and responsible for one another. I was committed to have an environment in our classroom toward helping individuals contribute and, maybe more importantly, have a sense of belonging to a community of learners. This particular classroom atmosphere, in turn, was aimed at helping my students accept me in my triply entwined role as the teacher, learner, and researcher, who genuinely cared about their perceptions and keenly interested in how they thought through and developed conceptual understanding of, in general, various introductory statistics topics, and in particular, statistical measures of spread.

Data Collection

This study aimed at assessing conceptual understanding of measures of spread among community college students in an introductory statistics course. Multiple data sources including both quantitative and qualitative approaches were used during the semester in order to achieve consistency of findings, and render a more holistic view with added richness and details. I began collecting data for the study at the beginning of the spring semester in January of 2010. Collecting, documenting, and analysis of the data obtained from various sources such as, pre- and post-assessments, student weekly journals (WJ), closed-book journals (CBJ), my own reflective journal, and digital audio recordings of mid-term conference interviews (MCI) overlapped and continuously evolved during the semester.

Quantitative Data

I implemented a one-group pre- and post-assessment design for the quantitative part of my study. Garfield & Chance (2000) and Garfield, delMas, and Chance (1999) reported that students' prior knowledge, pre-existing conceptions and intuitions affected student's understanding of measures of spread. Using the results of the pre-assessment, I was able to provide baseline data to evaluate prior knowledge, reveal some of the pre-existing conceptions of students regarding measures of spread. These types of study designs seem to be commonly utilized by researchers in statistics education research (delMas, Garfield, Ooms, and Chance, 2007; Garfield et al., 1999; Leavy, 2006). I also treated the data obtained from the pre-assessment as the basis for several discussions with students on the statistical measures of spread.

For the data collection phase of the quantitative component of my study, I considered a couple of assessment instruments. First, I considered using the Statistical Reasoning Assessment (SRA). The SRA, with its heavy focus on probability and limited emphasis on data sense, did not seem to be suitable for what I needed to assess. Instead, I chose the Measures of Spread (MOS) component of "Comprehensive Assessment of Outcomes in Statistics (CAOS)" pre- and post-assessment.

The CAOS test was developed as a part of the Assessment Resource Tools for Improving Statistical Thinking (ARTIST) project which was funded by the National Science Foundation (NSF). In response to address the assessment challenge, which can be outlined as the need to develop reliable, valid, practical, and accessible assessment instruments, the NSF funded the ARTIST project. The ARTIST website (https://app.gen.umn.edu/artist/) provides resources for evaluating students' statistical literacy (e.g., identifying words and symbols, being able to read and interpret graphs and terms), reasoning (e.g., reasoning with statistical information, explaining why or how results are produced), and thinking (e.g., applying reasoning in context, asking questions to critique and evaluate decisions involving statistical information).

These resources were designed to assist faculty in assessing student learning of statistics, to evaluate and improve their courses, and to assess the impact of reform-based instructional methods on important learning outcomes. According to delMas, Ooms, Garfield, and Chance (2006), a collection of over one thousand high quality assessment items and tasks was coded according to content (e.g., normal distribution, measures of center, bivariate data), type of cognitive outcome (e.g., statistical literacy, reasoning, or thinking), and type of item. These assessment items were made available to users for searching and reviewing. Moreover, the users currently may register on the ARTIST website and then download the assessment items of their choice to their own computers.

The CAOS test was developed over a three-year period and through an iterative process of acquiring and writing items, testing, and revising items. This process also involved gathering evidence of reliability and validity. For example, a sample of 1470 introductory statistics students, taught by 35 instructors from 33 higher education institutions from 21 states across the United States was used to gather data for a baseline psychometric analysis in 2006. This sample was used in an analysis of internal consistency of the 40 items on the CAOS post-test and yielded a reported Cronbach's alpha coefficient of 0.82 (delMas et al., 2007).

Over the three-year development phase, the advisory board, which consisted of a group of 18 members of the advisory and editorial boards of the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE), had provided ratings for the analysis of the content validity of the CAOS test. Although some raters indicated topics that they felt were not included, there was a unanimous agreement by the raters with the statement “CAOS measures basic outcomes in statistical literacy and reasoning that are appropriate for a first course in statistics,” and 94% agreement with the statement “CAOS measures important outcomes that are common to most first courses in statistics” (delMas et al., 2007, p. 31).

Based on above outlined evidence for internal reliability of the CAOS test, and the fact that it has been shown to be a valid measure of important learning outcomes for introductory college statistics courses, I used the CAOS measures of spread topic scale that was designed to measure students’ conceptual understanding of measures of spread in a first course in statistics. According to the ARTIST website, the CAOS measures of spread topic scale focuses on conceptual understanding, and its items were written to require students to think and reason, as opposed to merely calculate, use formulae, or memorize definitions. The ARTIST website (https://app.gen.umn.edu/artist/) provides access to the CAOS measures of spread topic scale.

Qualitative Data

In an effort to collect qualitative data from multiple sources, I relied heavily on student weekly journals (WJ), closed-book journals (CBJ), my own reflective journal, and digital audio recordings of semi-structured, mid-term conference interviews (MCI), which I conducted with twenty-eight students. Since the mid-term conference interviews were part of the regular course assessment, all students initially signed up to participate in the interviews. One student did not show up for the mid-term conference interview. Two students did not give consent for their direct interview quotes to be used in this study. Hence, in its final form, the data obtained from semi-structured, mid-term conference interviews consisted of twenty-six digital audio recordings. These twenty-six digital audio recordings of the interviews, with an average duration of 14:45 minutes and a standard deviation of to 6:45 minutes, were transcribed and analyzed.

I avoided using a list of pre-determined interview questions because I wanted to probe until no more information was forthcoming. I also realized that different students might require or need different types of probing or different amount of probing to reveal their conceptual understanding. A scripted interview format might have impeded the richness of data. However, the interviews were not without any structure at all. They all started with some focused questions and most of the interviews then turned into a process of using emerging findings to guide subsequent inquiry. I was able to probe deeper into conceptual understanding by pursuing avenues created by responses. This type of semi-structured interview format, in which the interview questions are modified or prompted by the researcher, as needed, for students to expand or elaborate on their responses, is based on the concept of theoretical sampling (Chance, delMas, and Garfield, 1999). The interview protocol is given in Appendix B.

In addition to the semi-structured, mid-term conference interviews, I used an open-ended response closed book journal (CBJ) to assess students' conceptual understanding of measures of spread qualitatively. Conceptual understanding requires students to develop relatively systematic, integrated or holistic understandings, as opposed to being able to regurgitate only fragmented pieces of information. Garfield & Ben-Zvi (2005) used the term deep understanding characterized by knowing the components of a concept and the connections among various components of the concept. They indicated that deep understanding was akin to conceptual understanding or relational understanding, as opposed to procedural understanding, and further pointed out that outcomes were best assessed in research studies by using open-ended assessment items or questions. A sample open-ended response assessment item is presented in Appendix C.

Data Analysis

The qualitative data for this study were analyzed via two methods, if you will. The first involved an ongoing, iterative process of constant comparison of data as suggested by Glaser & Strauss (1967). This ongoing process of data analysis and emerging theme searching is nicely aligned with the reflection-action-reflection cycle inherent in teacher action research (TAR). Second, likewise, although not continually throughout the semester, the qualitative data were analyzed using a particular taxonomy developed originally by Biggs & Collis (1982).

The qualitative data for this study were collected by means of weekly student journal entries (WJ), the transcripts of semi-structured, mid-term conference interviews (MCI) conducted with twenty-six students, and an open-ended response closed book journal (CBJ). Techniques such as open coding and constant comparison from grounded theory, which is defined by Glaser & Strauss (1967) as an approach to inductively developing ideas, concepts, and models that is grounded within a corpus of data, was used for the analysis of weekly student journals. In implementing this method, I started out with a quick sift-through of the journals to bring forth an overall picture and have a feel for the whole data in order to begin the process of examining the data. This initial review of the data helped me gain some knowledge of the pre-existing conceptions of students regarding measures of spread. Through this reading, I also started to form some ideas as to how students were articulating their conceptual understandings of measures of spread, in particular, and their thoughts on some other topics of the course, in general.

In accordance with the basic idea of the grounded theory, then, I systematically went through a process of searching for and identifying categories, concepts and properties, reoccurring themes and their interrelationships, or common groupings grounded in the data as I carefully read and re-read the weekly journal entries. Essentially, in this open coding part of the analysis, I read each line, sentence, and paragraph in search of the answer to the repeated question "what is this about? What is being referenced here?" in order to explain or develop an understanding of the phenomena found in the text. While reflecting on the journals and subsequent readings, I made anecdotal notes and maintained an inventory of codes with their descriptions along with color coded pointers to the text that contained them. Additionally, as the codes were developed, I wrote code notes and the literature related theoretical notes, which helped me set up categories into which to sort the data, and the categories were refined in light of iterative comparisons of the emerging categories.

As a result of the analysis described above, several emergent categories, such as cases of *faux amis* in statistics, procedural and relational aspects of measures of spread, negotiated meaning via group work, cognitive dissonance for a deterministic mindset, and non-dichotomy of assessment and teaching were identified. For example, the following comments by Macy indicated a typical observation of the non-dualistic nature of the relationship between continuous formative assessment and teaching:

> I think that the assessment presents the problems a little differently than the book does, or than we do in the class. Again the thing that stood out to me was IQR, I get it, that is just one thing that I will look at more. I actually learned something from the closed book journal and assessment. I have been struggling with standard deviation, but as i looked at the assessments today, it just clicked and I understand it now.

The cognitive dissonance experienced by many students, especially those with a deterministic mindset was articulated often. For instance, Mable commented:

> I am still a little fuzzy on the concept of statistics, so I am not very comfortable with this yet. I am very black and white, the variability part of statistics make me uneasy. I am very hopeful that this class will open my mind a lot more.

The socially constructed and negotiated meaning via group activities was another emergent theme. The comments similar to the following comments by Misha were widespread:

> Today's class showed me there is much more than just computing. With the class set up in groups, you get a sense of understanding, and learn with the others in my group, which helps me learn better. Our conversation did not turn away from the task at hand because we were actively taking part in our work.

The above comments also reflected the active learning approach used in the course, as opposed to a lecture-and-listen format. A more detailed data analysis is presented in the next chapter.

Structure of Observed Learning Outcomes (SOLO) Taxonomy

The analysis and evaluation of the twenty-six transcripts of the semi-structured, mid-term conference interviews and the student responses to open-ended assessment were informed by a form of Structure of Observed Learning Outcomes (SOLO) taxonomy. The SOLO taxonomy was developed originally by Biggs & Collis (1982) in order to identify how various components of a particular concept are used and integrated. Several studies investigating statistical reasoning or conceptual understanding based on students' responses to open-ended assessments used some form of the SOLO taxonomy to describe a hierarchy of levels of conceptual understanding (Bill & Watson, 2007; Jones, Thornton, Langrall, Mooney, Perry, and Putt, 2000; Reading & Reid, 2006; Reading & Reid, 2007; Watson et al., 2003). This cognitive development model was used by Reading & Reid (2006) to frame their study on reasoning about distribution with the integration of related statistical concept components.

The interconnectedness among various components of a statistical concept was also recognized by Budé (2006). He identified the recognition of interconnections on the part of students as an indication of profound understanding required in order for the concept to be used in a different setting. Statistical measures of spread as a statistical concept considers a data set as an entire aggregate, with its own characteristics of measures of the range, interquartile range (IQR), median, mean, standard deviation (SD), density and the overall shape of the distribution of a data set. This comprehensive conceptual entity requires simultaneous consideration and integration of all aspects of the data set and all components of the comprehensive concept. Computational aspects of this comprehensive concept might support analysis, but generally they replace or hinder the development of more intuitive notions about the measures of spread for the data set. Based on findings reported by Budé (2006), a profound conceptual understanding of statistical measures of spread enables students to use this concept in a different setting such as informal inference.

In analyzing the data for the qualitative component of my research study, I applied the following SOLO taxonomy with five different levels to describe students' conceptual understanding of statistical measures of spread. The five different levels are as follows: Prestructural, unistructural, multistructural, and relational and extended abstract. In general terms, I envisioned the prestructural level to convey disconnected bits of information, a "statistical dump", as suggested by Werner (1997), or a tautology without any reference to any of the components of statistical measures of spread. The unistructural level focused on components of statistical measures of spread in isolation without any or minimal consideration given to the interrelations among them. The multistructural level focused on components of statistical measures of spread with at least some consideration given to the similarities, differences, and interrelations among them, while the relational level developed interconnections among various components of statistical measures of spread. The relational level also required an ability to appreciate the significance of the all three measures of spread in relation to a coherent whole and to integrate them into a coherent whole. The fifth level, the extended abstract level, focused on an ability to make connections not only within a context or given area, but also to generalize, transfer, or judge the appropriateness of the components of statistical measures of spread beyond the underlying specific instance.

I coded and themed the qualitative data obtained from closed book journals (CBJ) and semi-structured, informal interviews (MCI) conducted during mid-term conferences according to these five levels of the SOLO taxonomy. In the coding process, the following ideas, which were discussed and consulted with expert colleagues, were kept as a set of guiding principles. In introductory statistics, the range, interquartile range (IQR), and standard deviation (SD) are commonly referred to as measures of spread. For instance, an appreciation or recognition that all three have a common characteristic of identifying a global, local or typical spread in data sets is one way of forming conceptual understanding of these measures of spread. As the connections among these discrete bits of information about the measures of spread are established and become stronger to form a cognitive web, the conceptual understanding of these measures are developed and modified as needed (Hiebert & Carpenter, 1992). For example, recognizing the fact that these measures of spread have different levels of sensitivity to the presence of outliers in a data set may lead to the cognitive development in deciding which one of these measures may be more appropriate to use in a particular data distribution.

The coding of the responses was based on how well these relevant aspects of measures of spread were discussed and how well the connections among them were established in responses to the closed book journal (CBJ) question, which was stated as follows: What are the measures of spread? Discuss their similarities and differences. In coding of the responses some of the aspects of a check-coding procedure outlined by Miles & Huberman (1994) were followed. After I grouped the responses into different levels, I asked an expert colleague to read and group responses within the levels whose characteristics I described. In the cases of discrepant coding of a response, the appropriate coding of the response was discussed and resolved. The discussion also revolved around some of the revisions of descriptions for characteristics at some of the levels.

The qualitative characteristics of each of the levels of responses are illustrated by discussing sample responses from each of the levels identified. If a student relied on discussing or giving definitions for measures of spread in isolation without any or minimal connections among the measures when responding to an open-ended question, then the response was coded as unistructural. For example, the following response by Myrobi was a typical response at the unistructural level:

> Range is the min and max of the endpoints of the data. IQR is the spread about the median. Standard deviation is the spread about the mean.

This particular response included descriptions for the three measures of spread. Even though the response did not indicate specifically similarities, differences, and connections among the three measures, there was an indication of minimal similarity between IQR and SD being both spreads, but about two different anchor values. Another student Marion wrote:

> Range is the difference between maximum and minimum. Interquartile range between the upper quartile in other word interquartile range is the whole. IQR is the middle of data. Standard deviation observation from mean.

Marion, in her particular response, attempted to give also descriptions for the three measures of spread in isolation without specifically mentioning any connections among them.

On the other hand, in addition to the definitions or descriptions, if a student's response provided at least a vague notion of the range, IQR, and SD as all three being used to measure spread in data sets, and some indication of similarities, differences and connections among these measures of spread, then the response was coded as multistructural. For example, the following response by Misha was a typical response at the multistructural level:

> The range is the measure of one end of the data to the other. (i.e – for data between 1 &10 is 10) The interquartile range doesn't measure the entire

> range, only the center 50% of the data rather than the entire range to compact your data set. The standard deviation is the average distance away from the median in your data. It is also kind of close to the interquartile range. They are all similar because they are all measures of spread.

Another student, Mabel, commented as follows:

> They are all ways of measuring information of a data set. Range is the total spread (from smallest to largest). The IQR is the difference of the upper quartile & the lower quartile and the standard deviation is the average of the distances of each observational unit from the mean. They differ in what each one measures. i.e. – entire spread, distance from mean on avg. IQR=50 percentile from mean – I guess.

In responses at this level, it was evident that students had acquired the notion that each of the three measures captures some of the characteristics of data sets, and reflected some indication of similarities, differences and connections among these three measures of spread.

At the relational level, in addition to similarities and differences, a response needed to discuss the resistance of these measures to the outliers by relating these measures of spread to mean, median, and outliers. For example, the following response by Michka was a typical response at the relational level:

> Range is the difference between the max. and min. of the dataset. IQR is the difference between the upper quartile range and the lower quartile range. Standard deviation is roughly 25% on the left of the mean of the dataset and about 25% right of the mean. 25% being about ¼ of the total dataset. There are also different in that IQR is more resistant to change due to outliers and standard deviation is less resistant. There are all similar because they are all deal with deviations from or between a set of data points.

Another student, Mikael, made the following comments:

> Range, interquartile range, and standard deviation are all measures of spread. We use these as tools to figure out what data set will consist of and to draw conclusions about our data. Range is calculated by taking the highest number and subtracting it from the lowest number in the data set.

> This is a good way to find the spread of the data, but it is non-resistant to outliers. Therefore if there is an extreme outlier then the data might be misrepresented. Interquartile range looks at the numbers in the middle 50% of the data to see the spread. This type is resistent to outliers because we are looking in the center 50 percent and excluding those that might fall as extreme outliers. I believe this is the best way to represent the measure of spread for the data. The standard deviation of a data set is computed by looking at the average distance from the center that each number is. This is a good technique, yet, it is non-resistent also to outliers. It does represent the data better than the range though, I believe.

At the extended abstract level, a response included all the characteristics of those responses at the relational level. In addition to the characteristics of the relational level, a response at the extended abstract level needed to indicate an ability to appreciate the significance of the three measures of spread in relation to the whole data set as an entire aggregate, with its own characteristics of measures of the median, mean, outliers, skewness, density and the overall shape of the distribution of a data set. A statement regarding the appropriateness of these three measures of spread in the case of skewed distributions was also expected at his level.

The quantitative data for this study were collected by means of the CAOS measures of spread (MOS) topic scale which was designed to measure students' conceptual understanding of measures of spread in a first course in statistics. In analyzing the data for the quantitative component of my study, I implemented a one-group pre- and post-assessment design. A matched-paired t-test was conducted to determine whether or not there was a significant difference between the pre- and post-assessment scores of students.

Starting with the data analysis phase and throughout the remainder of the study pseudonyms were used to help ensure the privacy and confidentiality of all students who consented to participate in the study. The use of pseudonyms was also instrumental in bridging any individual pieces of data back to raw data items, when necessary.

Summary

This chapter outlined the methodological components, such as research design, data collection, and data analysis, of my study focusing on community college students' conceptual understanding of statistical measures of spread. A discussion on reasons and appropriateness of framing the study within an action research design into which quantitative and qualitative approaches were embedded was also provided here. The chapter further described the settings at the community college in general, and the instructional setting, in particular. The details of the participants in the study were also provided in this chapter.

The particular research questions to which I sought answers can be reiterated as follows:

1. What are some of the pre-existing conceptions of students regarding statistical measures of spread?
2. How do students articulate their conceptual understanding of statistical measures of spread?
3. How do students establish the connections between statistical measures of spread and behavior of distributions of data sets?

An overview of the respective data sources and the types of data are given in Table 4.

Table 4: Research Question, Data Source, and Data Type and Analysis Method

Research Question	**Data Source**	**Data Type and Analysis Method**
Question 1	• Weekly Journals (WJ) • Own Journal • Pre-test CAOS (MOS Scale)	Qualitative (Constant Comparison) Qualitative (Constant Comparison) Quantitative (Descriptive Statistics)
Question 2	• Weekly journal (WJ) • Own Journal • Closed book journal (CBJ) • Mid-term Interviews (MCI) • Post-test CAOS (MOS) Scale	Qualitative (Constant Comparison) Qualitative (Constant Comparison) Qualitative (SOLO Taxonomy) Qualitative (SOLO Taxonomy) Quantitative (Descriptive Statistics and a matched paired t-test)
Question 3	• Weekly Journals (WJ) • Own Journal • Closed book journal (CBJ) • Mid-term Interviews (MCI) • Post-test CAOS (MOS) Scale	Qualitative (Constant Comparison) Qualitative (Constant Comparison) Qualitative (SOLO Taxonomy) Qualitative (SOLO Taxonomy) Quantitative (Descriptive Statistics)

To complete a visual summary of the data analysis used in answering the research questions, Table 4 also includes the analysis techniques used for each particular data source. In the next chapter, Chapter Four, I analyze the findings in response to the research questions of my study.

CHAPTER FOUR

RESULTS AND ANALYSIS OF FINDINGS

The purpose of this study was to examine the individual perceptions and conceptual understandings that community college students have of statistical measures of spread in an introductory statistics course. In order to gain a broader and deeper insight into to my research questions, I gathered and analyzed both quantitative and qualitative data. The study aimed at answering the following research questions:

1. What are some of the pre-existing conceptions of students regarding statistical measures of spread?
2. How do students articulate their conceptual understanding of statistical measures of spread?
3. How do students establish the connections between statistical measures of spread and behavior of distributions of data sets?

In the following section the quantitative and qualitative data analysis will be presented to seek answers to the first research question.

First Research Question

In answering the first research question, I analyzed the data obtained from the CAOS measures of spread (MOS) pre-test, weekly journal entries (WJ), my own journal entries, and closed book journals (CBJ). An analysis of the CAOS MOS pre-test results is presented first. The CAOS MOS pre-test was administered on-line in a computer lab setting. There was substantial variation in the number of appropriate responses provided by the thirty-three students who were present to take the CAOS MOS pre-test, which had a total of fourteen items. The Assessment Resource Tools for Improving Statistical Thinking (ARTIST) website at https://app.gen.umn.edu/artist/ provided data for the percentage of students who chose appropriate responses for each item along with a breakdown of the percentages for the remaining response choices for each item.

Providing a description of each item on the CAOS MOS pre-test might be helpful to identify what particular aspect of students' conceptual understanding of measures of spread that each item was intended to investigate. The descriptions of items in a narrative format are presented in the next paragraphs followed by a summary of item descriptions given in Table 5. Since the standard deviation (SD) is one of the most commonly used measures of spread, eight items out of the fourteen were specifically designed to focus on investigating the students' conceptual understanding of the SD as a measure of spread. These were items 1 through 8. Item 1 asked students to interpret the value of a zero SD for a distribution of data points. Item 2 was aimed at examining students' ability to interpret a non-zero value of the SD in a data set consisting of quiz scores. Items 3 and 4 asked students to compare the SD of two pairs of distributions represented by two sets of histograms.

In item 3, the students were asked to compare the SD values of a normal and a uniform distribution, whereas in item 4, the comparison of the SD values was based on a pair of normal distributions. Item 5 investigated students' ability to recognize the non-negative nature of the SD in the context of a data set consisting of test scores ranging from +15 to -15 points. Item 6 asked student to compare the SD of ages in the two populations of same size; all college students at a State university and all residents in a small town. Item 7 focused on determining students' ability to estimate, without calculating, the SD for a given data set of tightly clustered test scores with a known mean value. Item 8 was designed to examine the students' ability to estimate, without calculating, the value of the SD for a given data set of test scores with a known mean.

Item 9 was designed with a focus on the multi-relational aspects of measures of spread and measures of center for a skewed distribution represented by a histogram. This was an item where the highest level of cognitive development was necessary in order to respond appropriately. It might not be surprising that there were no students who chose the appropriate response for this item. Item 10 focused on investigating the students' ability to recognize the relationship between the outliers and the range for a skewed distribution represented by a histogram, and the inappropriateness of using range as a useful summary of variability for a skewed distribution. Item 11 was based on the same skewed distribution represented by a histogram and examined whether or not students could identify an appropriate measure of spread for a skewed distribution.

Item 12 investigated students' ability to recognize the multi-relational aspects of the extreme values of a mound-shaped and symmetric distribution, and measures of center and spread. Item 13 was focused on whether or not students could identify a distribution with the smallest IQR among a given set of several distributions represented by box plots. Item 14 was based on one of the box plots to examine the ability of students to recognize the relationships between the outliers and the three different measures of spread in terms of resistance to outliers.

In addition to the above item descriptions given in a narrative format, the descriptions of items on the COAS MOS test are also summarized in Table 5 below. The

CAOS MOS pre-test in its entirety is available at the ARTIST website located at https://app.gen.umn.edu/artist/.

Table 5: Items of CAOS MOS test and their descriptions

Item	Description of Concept Investigated
1	Interpreting the meaning of a zero standard deviation for a distribution of data points
2	Interpreting a given non-zero value of SD
3	Comparing two histograms for the size of SD
4	Comparing two mound-shaped histograms for the size of SD
5	Recognizing the non-negative nature of SD
6	Comparing the SD of ages in the two populations of same size; all college students at a State university and all residents in a small town
7	Estimating (without calculating) the SD of a tightly clustered set of data points with a given mean
8	Estimating (without calculating) the SD of a set of data points with a given mean
9	Choosing appropriate measures of spread and center for a skewed distribution represented by a histogram
10	Recognizing the relationship between the outliers and the range for a skewed distribution represented by a histogram
11	Choosing an appropriate measure of spread for a skewed distribution represented by a histogram
12	Recognizing the relationship between the extreme values of a mound-shaped and symmetric distribution, and measures of center and spread
13	Identifying a distribution with the smallest IQR given a set of several distributions represented by box plots
14	The relationships between the outliers and the measures of spread in terms of resistance to outliers

The mean number of items students responded appropriately on the CAOS MOS pre-test was 4.4 with a standard deviation of 1.8 and range from 2 to 8. The 95% confidence interval was given as CI=[3.9,5.2]. It is noteworthy that almost all students indicated that they had to guess on majority of the questions. Considering the fact that random guessing would lead to an expected mean appropriate response of 3.5, students did significantly better than choosing responses based on random guessing ($t(32)=3.3$, $p<0.01$).

This result may indicate that students do not give themselves credit for their already existent conceptual understanding or have confidence in their intuitive abilities. However, there were only six students (18%) who gave appropriate responses to at least half of the fourteen items. Four of the students (12%) provided appropriate responses to half of the items. Two students provided appropriate responses to eight items, which was the highest number of appropriate responses. The amounts of time that students took to complete the CAOS MOS pre-test varied widely with a median completion time of 14.8 minutes, mean of 15.2, SD of 6.3, and range from 5.7 to 28.4 minutes. The typical time was about a minute per item. A majority of the students indicated in their journal entries that they had guessed on almost all of items. However, four students spent more than 25 minutes on the CAOS MOS pre-test.

The next step in analyzing the CAOS MOS pre-test data was to examine the responses to fourteen individual items of the CAOS MOS pre-test. It was surprising to find out there were only four items on which at least 50% of the students demonstrated an initial ability or conceptual understanding by choosing the appropriate responses at the beginning of the course. These four items and the percentage of appropriate responses given by the students to each item are summarized in Table 6 below.

Table 6: Items of CAOS MOS pre-test appropriately responded by 50% or more students

Item	Description of Concept Investigated	% of appropriate responses
2	Interpreting a given non-zero value of SD	54.5 (18 out of 33)
4	Comparing two mound-shaped histograms for the size of SD	78.8 (26 out of 33)
6	Comparing the SD of ages in the two populations of same size; all college students at a State university and all residents in a small town	65.6 (21 out of 32)
8	Estimating (without calculating) the SD of a set of data points with a given mean	51.5 (17 out of 33)

More than half of the students did not choose appropriate responses for the remaining ten items. None of the students chose the appropriate response for one of those items, which was item 9. Furthermore, at least 75% of students did not choose appropriate responses for seven out of these ten items. In other words, half of the fourteen items were not responded appropriately by at least 75% of the students. These seven items with their percentages of appropriate responses are given in Table 7 below. Table 7 also displays the remaining three items and their respective percentages of appropriate responses. As indicated in Table 7, none of the students chose the appropriate response for item 9, which was followed by items 3 and 5 with only two students choosing the appropriate responses for each one of these two items on the CAOS MOS pre-test.

The percentage of appropriate responses given by students to the first group of five items in Table 7, which were items 9, 3, 5, 11, and 12, ranged from 0% (0 out of 33) to 9.4% (3 out of 32). The percentage of appropriate responses given by students for the second group five items, which were items 10, 14, 7, 1, and 13, were not as low as the first group of five items and ranged from 24.2% (8 out of 33) to 45.4% (15 out of 33). These percentages of appropriate responses indicated that less than half of the students were able to give appropriate responses to 71% (10 out of 14) of the items on the CAOS MOS pre-test.

Table 7: Items of CAOS MOS pre-test appropriately responded by less than 50% of students

Item	Description of Concept Investigated	% of appropriate responses
9	Choosing appropriate measures of spread and center for a skewed distribution represented by a histogram	0.0 (0 out of 33)
3	Comparing two histograms for the size of SD	6.1 (2 out of 33)
5	Recognizing the non-negative nature of SD	6.1 (2 out of 33)
11	Choosing an appropriate measure of spread for a skewed distribution represented by a histogram	9.4 (3 out of 32)
12	Recognizing the relationship between the extreme values of a mound-shaped and symmetric distribution, and measures of center and spread	9.4 (3 out of 32)
10	Recognizing the relationship between the outliers and the range for a skewed distribution represented by a histogram	24.2 (8 out of 33)
14	The relationships between the outliers and the measures of spread in terms of resistance to outliers	24.2 (8 out of 33)
7	Estimating (without calculating) the SD of a tightly clustered set of data points with a given mean	39.4 (13 out of 33)
1	Interpreting the meaning of a zero standard deviation for a distribution of data points	42.4 (14 out of 33)
13	Identifying a distribution with the smallest IQR given a set of several distributions represented by box plots	45.4 (15 out of 33)

Measures of Spread as a Comprehensive Conceptual Entity

Item 9, which was based on the recognition of a multitude of relationships among the measures of spread and measures of center, required a high level of cognitive development of students. As noted by Budé (2006), the interconnectedness among various components of a statistical concept and the recognition of these interconnections on the part of students is an indication of conceptual understanding. Statistical measures of spread as a comprehensive statistical concept considers a data set as an entire aggregate, with its own characteristics of measures of the range, interquartile range (IQR), standard deviation (SD), median, mean, density, outliers, and the overall shape of the distribution of a data set.

This comprehensive conceptual entity requires simultaneous consideration of all aspects of the data set as integrated with all components of the measures of spread. In order to provide an appropriate response to item 9, it was necessary for students not only to have a high level of cognitive development but also to demonstrate conceptual understanding of the measures of spread as a comprehensive conceptual entity with all of its interconnected aspects. Hence, it might not be surprising that there were no students who chose an appropriate response for this item at the beginning of the semester. A question similar to item 9 was included in the interview protocol and used during MCIs. The responses to this particular item and other items will be analyzed later in this chapter.

The percentage of particular choices of students for inappropriate responses in each item varied among the provided choices somewhat evenly. However, in one particular item, item 3, a single inappropriate response choice was chosen by 78.8% (26 out of 33) of the students. It is noteworthy that the choice of this particular inappropriate response might be a result of the fact that a great majority of students view the variation along the vertical axis of a histogram as a cause for a larger SD value in a distribution. This specific view, which seems to be prevalent among students, was well documented and reported by other research studies as well (delMas & Liu, 2005; Meletiou & Lee, 2002; Turegun & Reeder, 2008). In these research studies, it was also reported that when comparing the spreads of two distributions represented by histograms, students generally associated a smaller spread with a symmetric and mound-shaped distribution, regardless of the spread of the other distribution in comparison. However, in this study it was found that students did not seem to draw a similar association when they were comparing the SDs of a normal and a uniform distribution.

The percentage of appropriate student responses for item 5 tied with item 3 for the second lowest percentage of appropriate student responses among the fourteen items. It was rather surprising that 93.9% (31 out of 33) of the students' responses indicated that the students did not seem to recognize the non-negative nature of the SD in a data set. About 73% (24 out of 33) of the students' responses revealed that interpretations of a negative SD value as an indication of either most scores being negative scores or most scores being below the mean score were widespread among students. However, students' choices of these two inappropriate interpretations might be partially attributed to this particular item's context which consisted of the total test scores with positive and negative values ranging from +15 points to -15 points.

The percentage of appropriate responses given to item 11 tied with item 12 for the third lowest percentage of appropriate responses given by students. Item 11 was based on the same skewed distribution represented by a histogram as in item 10, and examined whether or not students could identify an appropriate measure of spread for a skewed distribution. Nearly one-fourth (8 out of 33) of the students' responses to item 10 seemed to indicate that the range was not a useful summary of the variability in a highly skewed data set because of the outliers easily influencing the range. These students were able to demonstrate conceptual understanding of measures of spread at unistructural and multistructural levels. However, it was not evident in students' responses to item 11 that students were able to make the transition from the unistructural level of their conceptual understanding of measures of spread to the next cognitive development level in the SOLO taxonomy, namely the multistructural level.

The percentage of appropriate responses given to item 12 tied with item 11 for the third lowest percentage of appropriate responses given by students. Item 12 investigated students' ability to estimate the extreme values of a distribution. It required recognition of the multistructural aspects of the extreme values of a mound-shaped and symmetric distribution, and measures of center and spread. Again, it was not evident in students' responses to item 12 that students were able to demonstrate their conceptual understanding of measures of spread at the relational level.

Because of its simple computational procedure and uncomplicated conceptual structure the range is regarded as the most basic measure of spread. Item 10 focused on investigating the students' ability to recognize the relationship between the outliers and the range for a skewed distribution represented by a histogram, and inappropriateness of using range as a useful summary of variability in the case of a skewed distribution. Yet, only 24.2% (8 out of 33) of the students provided an appropriate response to item 10. The low percentage of appropriate responses to this item might be attributed to the fact that the students were not merely being asked to only calculate or explain it as a concept at a unistructural level, but to demonstrate their conceptual understanding of the range as a measure of spread at the multistructural level by recognizing the relationship between the range and the extreme values in a skewed distribution.

Item 14, similar to item 10, focused on examining the ability of students to recognize the relationships between the outliers and the three different measures of spread in terms of resistance to outliers. Instead of being based on a skewed distribution represented by a histogram as in item 10, item 14 was based on a set of several distributions represented by box plots. As in item 10, only 24.2% (8 out of 33) of the students provided an appropriate response to item 14. Regarding the responses given by students to item 7, and also item 8, it was important to notice that students seemed to estimate the standard deviation of a set of tightly clustered data points with much better success than that of a data set which was not so tightly clustered.

Less than half (14 out of 33) of the students were able to give an appropriate response to item 1, which asked students to interpret the meaning of a zero SD for a distribution of quiz scores. However, when asked to interpret a small SD value of 1 point for the same distribution of quiz scores, in item 2, more than half (18 out of 33) of the students were able to choose the appropriate response. The students were able to determine that a typical score would be within 1 point of the mean quiz score with a better appropriate response rate than interpreting the meaning of a zero SD in a data set consisting of negative and positive quiz scores.

Item 13 was responded to appropriately by 45.4% (15 out of 33) of the students. Item 13 focused on whether or not students could identify a distribution with the smallest

IQR among a given set of several distributions represented by box plots. Item 13 also examined students' ability to identify the graphical representation of the IQR on a box plot. Since many of the students indicated that they were not familiar with the IQR, it was not surprising that less than half of the students were able to choose the appropriate response to this item.

While the results from the CAOS MOS pre-test provided a valuable overall view of the students' pre-existing conceptions, the multiple-choice format of the CAOS MOS pre-test made it very difficult to understand what students' particular thought processes were as they decided on their forced-choice responses. In order to triangulate some of the findings from the CAOS MOS pre-test and to gain additional insight into students' thought processes I examined the students' weekly journals (WJ) and my own journal entries for the indications of students' pre-existing conceptions of measures of spread. A qualitative analysis of these responses is presented next.

As students were working on the CAOS MOS pre-test using the computers in the Mathematics Lab they were asked to record the items for which they had to guess a response. The students were also instructed to jot down on a piece of paper the particular terms and concepts that they have never seen before or were not familiar with at all. A review of these collected data revealed that students guessed on almost every item on the CAOS MOS pre-test. Although it was explained to students the purpose of this assignment was to simply give me an idea about their statistical background and they were not going to be graded on this assignment, one of the main themes emerged during the analysis was the anxiety experienced by students. Some students expressed that they were tense and nervous about the idea of having to take the CAOS MOS pre-test. Not knowing or being familiar with the terms and the concepts on the CAOS MOS pre-test was a common source of intimidation for a number of students. For example, Medina expressed her feelings and thoughts as follows:

> I found that I had to guess at all the questions. This made me feel very intimidated and stupid. That is, until we got back to the classroom, and the other members of my group were saying they felt the same way. I am guessing at 1-14. Don't know these terms center, spread, outliers, quartile, and interquartile.

Feelings of intimidation because of not knowing the terms and concepts on the CAOS MOS pre-test was also addressed when Melis wrote,

> This test was extremely intimidating and I didn't know anything on the test. I had to guess on all of them mainly cause I didn't know the meaning of the terms they were using. After taking the test I do have to say this class seems a little more than intimidating, but after we returned to class and discussed some of the vocabulary it helped understand what exactly these questions were asking for. The questions I had: What is deviation? What is IQR? What are outliers?

Another student, Mabel, shared the following feelings and thoughts,

> Going to the Math lab to take the accessment test showed me how much I <u>don't</u> know about statistics. Guessed on #1, 2, 3, 5, 6, 7, 8, 9, 10, and 11, guessed on them all. It was painful, but once we returned to class and went over range, median, and graphed a few different figures I felt a little better.

A number of students found the items of the CAOS MOS pre-test to be challenging, difficult, and hard to understand. For example, Myrem wrote,

> When i took the test, it was actually difficult. I thought i would do better. When i came across problems, it was hard because i didn't understand the question it was asking me. I don't understand what it is asking me. I'm having a lot of trouble figuring out the right answer because I don't even understand what the question is asking me. It is a lot different from other math classes that I'd taken.

Another student, Mendela, indicated,

> I found the test to be somewhat challenging because it kept using the word standard deviation and I'm not really sure what that means. I guessed on #6, 7, 8, and 10, I guessed on the majority of the problems.

A number of students indicated either using some basic terms, such as mean, median and mode, or remembering them from previously taken high school or college classes. For example, Majide wrote,

> My biggest concern with today was that most people should have known these terms prior to this class. I mean on the assessment their were a couple of questions that I was unsure of but I think I knew the simplistic of these terms prior to class from high school and middle school. I guessed on #13 and #14 and IQR?

One of the students, Millie, wrote,

> While taking the test in the Math Center I realized I had forgot about mean, median, and mode. Before this class I had used mean, median, mode, and range before, but I have never used standard deviation. This is something new.

Another student, Myrobi, put down,

> We took an assessment on some terms that I have heard of but do not know what they mean. For example standard deviation. The assessment was only 14 questions long and it was okay for us to guess. On some I tried making an educated guess. I guessed on the questions #1, 2, 4, 5, 9, 10, and 14.

Some other students were familiar with and even used a number of the concepts from the items on the CAOS MOS pre-test, but had difficulties remembering the details. For example, Michka addressed her difficulties as follows:

> We went to the Math Lab to take an assessment for statistics. I found it incredibly challenging due to the fact that I couldn't recall the definition of standard deviation. I know I had learned this term in a previous class but I couldn't recall helpful tips to assist me on the assessment. We returned to class to discuss deviation and other unknown terms we saw in the assessment. I guessed on all questions.

Overall, most students indicated that they had guessed on almost all the items of the CAOS MOS pre-test. Although students were familiar and used some of the basic concept components of measures of center, such as mean and median, there were a number of common terms and concepts that the students listed as unfamiliar. These commonly listed unfamiliar terms and concepts consisted of standard deviation, interquartile range, followed by, deviation, outliers, and box plot. A number of researchers indicated that the curricular emphasis in high school and college has traditionally focused on measures of center, rather than measures of spread (Shaughnessy & Ciancetta, 2002; Watson, Kelly, Callingham, and Shaughnessy, 2003). The most commonly listed unfamiliar terms and concepts of measures of spread, such as standard deviation and interquartile range, might be an indication that the lack of curricular emphasis on measures of spread still continues to be a prevalent curricular practice at the high school and college level.

Reflection-in-Action

The findings from the CAOS MOS pre-test not only were helpful in identifying pre-existing conceptions of students regarding statistical measures of spread but they were also, as indicated by Kang (2007), valuable in informing my teaching in terms of framing problems to improve my teaching and students' understanding of measures of spread. I was able to use the findings of the CAOS MOS pre-test in the sense of Schön's (1983) idea of reflection-in-action in order to adjust my teaching and alter instructional activities in the process of teaching. For example, I used classroom activities and discussions which focused more on the terms and the concept components listed as unfamiliar by the students. In the next paragraph, a particular instance of such a case is presented.

Upon analyzing the student weekly journals regarding students' prior conceptions of measures of spread, an emerging common prior conception was the association of a large range with a large SD in a data set. In the spirit of reflection-in-action, reflecting back on this particular result, I decided to ferret out further and address this type of partially productive reasoning by searching for teaching activities that would create some cognitive dissonance among students. Although the range has an effect on the value of the SD in a data set, the locations with respect to the mean or distribution of data points within that range also affects the value of the SD in a data set.

I used the following particular item given in Figure 1 as a teaching activity to illustrate the concept that data sets with the same range not necessarily had the same IQR or SD. The item was downloaded from the ARTIST site and modified slightly. The three samples with the same range were used to discuss how the value of the SD was also dependent upon how the data points were distributed within the values of 10 and 15, and the larger range would not always result in a larger SD. In another group activity, various distributions were constructed from the data collected in class and the discussion was focused on guessing which data sets would have larger spread and larger SD values.

However, the following remarks in response to the assessment item in Figure 1 provided evidence that, after considerable discussion and group activities on the topic, there were still some students who did not give sufficient consideration to how the data points were distributed within the range. For example, Malia in responding to which sample has the smallest spread, wrote,

> Sample 2 because it only includes 2 numbers.

Malia's response did not seem to indicate a realization of how these two repeated data points were located at the far opposite ends of the range. Another instance of a similar response was given by Mogan who wrote,

> 2 [referring to Sample 2] because there's only 2 numbers.

In Mogan's response, there was no indication of any consideration given as to how the quiz scores were placed within the range.

Consider the following samples of quiz scores and answer the questions without doing any calculations.

Sample 1: 10, 11, 12, 13, 14, 15
Sample 2: 10, 10, 10, 15, 15, 15
Sample 3: 10, 12.5, 12.5, 12.5, 12.5, 15

Which sample has greatest spread? Why?
Which sample has the smallest spread? Why?

Figure 1. CAOS MOS assessment item from ARTIST site

Based on my past experience of teaching at various community colleges for more than two decades, the enrollment in most community college introductory statistics courses quite commonly consists of students with a wide variety in their backgrounds and in their major fields of study. As indicated in Chapter 3, this particular community college's introductory statistics course was not an exception. Hence, it was not surprising that the students' comments regarding their prior experiences with and use of statistics indicated a broad variation in the statistical backgrounds and abilities of students in the course. Out of the twenty-nine students from the two sections of the course, approximately one third indicated that they were not familiar with statistics at all. Some of the students in this group further indicated not even having heard of statistics until they were placed in the course. Slightly more than one third of the students indicated that their previous encounters with statistics were in their various college classes such as sociology, psychology, microbiology, government, genetics, biology, and pharmacology. However, further analysis of the student weekly journals (WJ) in conjunction with my own journal revealed that most difficult hurdles were not necessarily the new words that students encountered but various cases of *faux amis*.

Cases of Faux Amis in Statistics

Faux amis is a phrase used by the French to describe pairs of words or phrases that are either the same or look and sound similar, but whose meanings are different in two languages. Although an individual may not be aware of the fact that a word being used is a *faux ami*, a different country, a different language, or a different context may provide helpful clues for the individual to deduce the intended meaning. There are, however, cases of *faux amis* in the same language, same country and even within the same context. Skemp (1977) identified and provided a discussion on two such words in the context of mathematics and arguably placed these alternative meanings at the root of many of the difficulties in mathematics education today (p. 20).

Emerged from the analysis of students' weekly journals were three such cases of *faux amis* within the context of introductory statistics. One of the main themes, which became apparent during the analysis of students' comments, revealed that the word statistics itself was such a case. When students indicated a particular use of or a prior experience with statistics the focus was solely on looking up or searching for a statistic, or statistics in the plural sense of the word statistic, on a particular topic or item rather than the actual use of statistics as a body of knowledge to gather and analyze data. A student, Mevin, when asked to comment on his prior use of statistics, wrote the following phrase:

> The time when I read reviews, specifications, and personal experiences with tire selections via www.tirerac.com.

Another student, Myrem, similarly remarked on her prior use of statistics as follows:

> When i had to find graduate school i wanted to go i had to read and study the statistics in the academic area. I have recently looked on college websites at the statistics of the ages, races, and gender of students in each school.

Above remarks reflected that, for the most part, the difference in the meanings of these two words was not apparent in the prior conceptions of students. One of the students, not initially aware of this difference, resolved the issue later and reflected back on the distinction in the meanings of these two words by making the following remark:

> I also learned that statistics and statistic have very different meanings…

The second case of *faux amis* was regarding the use of the word range as a statistical term as opposed to its use based on a common vernacular. In statistics, the range is one of the measures of spread and defined as the distance or difference between the largest and smallest data values in a data set. A response given by Miranda to the item in Figure 1 illustrated a typical example for this case. Miranda's response was as follows:

> 1 [referring to Sample 1] because the sample ranges from 10 to 15 making the spread larger and even.

Similarly, Manuela gave the following reasoning to the same item.

> Sample 1 because the range of #'s.

Miranda's and Manuela's responses did not seem to indicate any acknowledgement that all three samples had the same value for their range. Instead, the use of the word range in their responses seemed to indicate a vernacular usage. Miranda and Manuela were trying to point out that all the consecutive integers between the endpoint data values of 10 and 15 were present in Sample 1. Slauson (2008) also reports similar findings regarding the word range being used rather liberally and not necessarily in a statistically correct way by her students.

There was considerable evidence for a third case of *faux amis* in the students' responses to the above item. This particular case was based on the words variation and variety being viewed by students as two interchangeable terms. Even though an association between variation and spread was acknowledged in the student responses, an appreciation for the difference between the variety, as in the case of different values in a data set, and variation within the values with respect to mean, as in spread, was not evident in many of the student responses. For instance, Myrobi, in response to the item in Figure 1, wrote the following reasoning as to why Sample 1 had the greatest spread.

> Sample 1, has more different values.

In a similar response to the same item, Mishka wrote the following.

> Sample 1, because it has the least repeated #'s

A couple of other similar responses given by Mona and Macy included the following lines of reasoning, respectively.

> 1 [referring to Sample 1] it has more variety in answers given; 5 different values from one another
>
> Sample 3, there are more numbers (data) & all 3 ranges for all 3 samples are the same so that has no effect.

Although Macy's response indicated an acknowledgment that all three ranges were the same, it was evident in all four responses that the students were using a vernacular for variety in lieu of the term variation.

During the analysis of the student responses, it was noted that students' use of their own terminology was indicative of students' construction of their own meaning.

However, it may become an issue to address if the students do not have appropriate statistical terminology built and developed enough to express their conceptual understanding of measures of spread. The foregoing samples of students' prior conceptions of spread revealed that not recognizing some of these statistical *faux amis* might prevent teachers from helping students eventually develop a more appropriate statistical terminology. Likewise, Makar & Confrey (2005) points out the importance of learning to recognize nonstandard language when students try to articulate their conceptual understandings of variation. In the next section of this chapter, when I examine how students articulated their conceptual understanding of statistical measures of spread, I will provide more samples for and discussion on the use of students' own terminology.

Second Research Question

In answering the second research question, I analyzed the data obtained from the CAOS measures of spread (MOS) post-test, weekly journal entries (WJ), my own journal entries, closed book journals (CBJ), and mid-term conference interviews (MCI). An analysis of the CBJ results based on the Structure of Observed Learning Outcomes (SOLO) taxonomy is presented first.

A SOLO Taxonomy Model for the Measures of Spread

The Structure of Observed Learning Outcomes (SOLO) taxonomy developed originally by Biggs & Collis (1982) to describe a hierarchy of levels of conceptual understanding and employed by several studies (Bill & Watson, 2007; Jones, Thornton, Langrall, Mooney, Perry, and Putt, 2000; Reading & Reid, 2006; Reading & Reid, 2007; Watson et al., 2003). The qualitative assessment of students' responses to open-ended questions was informed by the SOLO taxonomy and the statistical appropriateness of the response in order to identify how the different concept components of the measures of spread are used, related, and integrated into a whole conceptual entity.

A particular version of the SOLO taxonomy was envisioned for assessing the conceptual understanding of statistical measures of spread based on open-response items of the CBJs as well as the MCIs. The SOLO taxonomy, which I envisioned, consisted of five cognitive levels. The five cognitive levels of the SOLO taxonomy, the qualitative characteristics of each response level, and the number of student responses categorized at each cognitive level are presented in Table 8.

Table 8: The five levels of SOLO taxonomy and qualitative characteristics of responses

Level	Qualitative Characteristics of Response	Frequency
Prestructural	Provides a collection of disjoint information, use of a tautology, or "statistical dump" stimulated by the words in question, or misunderstood question	0
Unistructural	Relies on discussing, describing or giving definitions for measures of spread in isolation without any or minimal connections	11
Multistructural	In addition to the definitions or descriptions, provides at least a vague notion that the range, IQR, and SD are all used as measures of spread in data sets, and some indication of similarities, differences, and connections among these measures of spread	16
Relational	In addition to similarities and differences, describes which measure of spread may be more resistant to the outliers by relating measures of spread to mean, median, and outliers	2
Extended Abstract	In addition, generalizes as to which measure of spread may be appropriate in the case of a skewed distribution	0

A great majority of student responses (27 out of 29) was at the unistructural and multistructural levels. There were no responses indicating that students were off task, not engaged or misunderstanding or misinterpreting the question. Since these were some of the qualitative characteristics of the pre-structural level, no responses given by students was coded at the prestructural level. There were no student responses at the extended abstract level either.

Eleven student responses indicated unistructural-level understanding of the measures of spread. These responses did not go beyond describing or giving definitions for measures of spread in isolation without any or minimal connections among the measures. For example, the following response by Myrobi was a typical response at the unistructural level:

> Range is the min and max of the endpoints of the data. IQR is the spread about the median. Standard deviation is the spread about the mean.

The responses at the unistructural level generally included descriptions for the three measures of spread. Even though they did not indicate specifically similarities, differences, and connections among the three measures, there were hints of minimal similarity among the measures of spread.

There were sixteen student responses at the multistructural level. At this level, in addition to the definitions or descriptions, student responses provided at least a vague notion of the range, IQR, and SD as all three being used to measure spread in data sets, and some indication of similarities, differences and connections among these measures of spread. For instance, the following response by Misha was a typical response at the multistructural level:

> The range is the measure of one end of the data to the other. (i.e – for data between 1 &10 is 10) The interquartile range doesn't measure the entire

range, only the center 50% of the data rather than the entire range to compact your data set. The standard deviation is the average distance away from the mean in your data. It is also kind of close to the interquartile range. They are all similar because they are all measures of spread.

Another student, Mabel, commented as follows:

> They are all ways of measuring information of a data set. Range is the total spread (from smallest to largest). The IQR is the difference of the upper quartile & the lower quartile and the standard deviation is the average of the distances of each observational unit from the mean. They differ in what each one measures. i.e. – entire spread, distance from mean on avg. IQR=50 percentile from mean – I guess.

It was evident in student responses at the multistructural level that students had acquired the notion that each of the three measures of spread encapsulated some of the characteristics of data sets, and reflected some indication of similarities, differences, and connections among these three measures of spread.

At the relational level, in addition to similarities and differences, a response needed to discuss the resistance of these measures to the outliers by relating these measures of spread to mean, median, and outliers. There were two student responses at the relational level. Both of these responses are presented here. The following response by Michka was one of the responses at the relational level:

> Range is the difference between the max. and min. of the data set. IQR is the difference between the upper quartile range and the lower quartile range. Standard deviation is roughly 25% on the left of the mean of the data set and about 25% right of the mean. 25% being about ¼ of the total data set. There are also different in that IQR is more resistant to change due to outliers and standard deviation is less resistant. There are all similar because they are all deal with deviations from or between a set of data points.

The other student response at the relational level was by Mikael, who made the following comments:

> Range, interquartile range, and standard deviation are all measures of spread. We use these as tools to figure out what data set will consist of and to draw conclusions about our data. Range is calculated by taking the highest number and subtracting it from the lowest number in the data set. This is a good way to find the spread of the data, but it is non-resistant to outliers. Therefore if there is an extreme outlier then the data might be misrepresented. Interquartile range looks at the numbers in the middle 50% of the data to see the spread. This type is resistent to outliers because we are looking in the center 50 percent and excluding those that might fall as extreme outliers. I believe this is the best way to represent the measure of spread for the data. The standard deviation of a data set is computed by looking at the average distance from the center that each number is. This is a good technique, yet, it is non-resistent [sic.] also to outliers. It does represent the data better than the range though, I believe.

Even though there were no student responses at the extended abstract level, at this level a response would have included all the characteristics of those responses at the relational level. In addition to the characteristics of the relational level, a response at the extended abstract level was conceived in such a way that it needed to indicate an ability to appreciate the significance of the three measures of spread in relation to the whole data set as an entire aggregate, with its own characteristics of measures of the median, mean, outliers, skewness, density and the overall shape of the distribution of a data set. A statement regarding the appropriateness of these three measures of spread in the case of skewed distributions was also expected at his level.

When examining the various responses it became apparent that the lengths of student responses on the CBJs were becoming gradually longer as the cognitive level of responses increased. As indicated by the last two sample responses, students were inclined to articulate their conceptual understanding of measures of spread by writing longer responses and seemed to be naturally at ease with the use of terminology. In contrast, however, at the lower cognitive levels of the SOLO taxonomy student responses consisting of very short rationales were quite common. A similar pattern in students' responses regarding the use of language was observed and reported by Slauson (2008). She made following comments on the pattern of the language used by students in their written responses (p. 104):

> The students who did use correct statistical language tended to have longer responses. This may be an underlying factor to why so many responses had the correct answer but incorrect reasoning. In my experience, students tend to be hesitant to write more than they absolutely must to answer a question, especially if they were not confident about their answers.

Many students, when commenting on the similarities among the range, IQR, and SD indicated that these three were all measures of spread. However, it was not clear or I could not deduce from some of the students' responses whether this phrase had any significant meaning to them. There was a possibility that some students might be simply repeating a phrase that they heard being used, or even possibly relying on rote memorization. Reflecting back on this issue, I decided to include an item in the MCIs to ferret out students understandings of this phrase. The item in the MCI asked students to explain what the phrase measures of spread mean to them, or describe their understanding of the phrase measures of spread. If there was not a satisfactory indication of an understanding of the phrase in a particular student response during the MCIs, then the response to CBJ was categorized at a lower SOLO level than the original response. A discussion of the student responses to this question during the MCIs will be presented in the next section.

Conceptions of Measures of Spread

In analyzing student responses to describe their understanding of the phrase measures of spread one of the emerging themes was identified as the use of students' own words, terminology, and language in order to articulate their conceptions of measures of spread. I believe that students' use of their own terminology, which was meaningful to them, to articulate their conceptual understanding of measures of spread was a reflection of a constructivist perspective. Several researchers reported similar results on the types of non-standard or non-technical language pre-service teachers used in describing and analyzing various data sets (Makar & Confrey, 2005). Even though some students were using non-standard or non-technical language, they were still able to express a rich understanding of measures of spread. For example, Malia expressed her understanding of what it meant for a data set to have a small spread as follows:

> The data set is all crowded around the center.

However, it was not clear whether in her response she was referring to the mean or median as the center. Another student, Manuela, wrote the following sentence to describe a data set with a small spread:

> …because they are right by each other.

Mabel indicated her understanding of measures of spread as follows:

> Ahm... just to measure, I don't if you could say the most common, you know, whatever the observation units are you are wanting to measure, you know where everybody is at within, you know, the scale, you know, where it clusters I guess you are wanting to see where most of everybody is at. You know, the IQR will tell you your majority, then your lower quartile, your upper quartile kind of help you indicate your outliers. That's how I think of it anyway.

Although Mabel's response indicated a more comprehensive view of the measures of spread, her response showed an inclination of transitioning into describing procedures for calculations as a part of articulating the meaning of the phrase measures of spread. This particular tendency of either stating definitions or describing procedures for calculations was another theme identified in the analysis of students' responses. For instance, Macy articulated her understanding of measures of spread with the following comments which trails off into definitions and procedural descriptions:

> It is kind of the way the distribution is laid out. Like if your range is larger, if you have large range then your distribution is going to be wider. Hhm… I know the definition of all three of those. A range is your lowest point subtracted from your highest point. So that is going to include your outliers. So, it is not the best one to use. Hhm… your SD is how many units away from your mean you get. So, that is a very smaller range. I think that's the one I prefer to use. As much accurate because it gives a smaller range and it is giving you closest to the mean. IQR is like when you are in the highest 25% quartile and the lowest part.

Another student, Marion, made the following comments during her MCI:

> Kind of spread it just reminds ahm… SD like how the data is spread. And SD is kind of measure the variability of how spread the data is or how narrow it is going to be. And kind of ahm… wish I understood… kind of calculate ahm… calculate the different between… difference between the mean and the data value. And then the square root of the difference.

In order to triangulate the findings, I compared the student comments during the MCIs against the student responses to the open-ended CBJs. Many of the students who articulated their conceptual understanding of measures of spread with a sole focus on the procedural descriptions or definitions provided responses to the previously open-ended CBJs at the unistructural level of the SOLO taxonomy. Marion was one of these typical cases.

On the other hand, some students immediately started directing their comments towards viewing the measures of spread as a comprehensive entity with several interconnected components. A typical response for such a case was one given by Mishka who made the following comments during her MCI.

> How the data spread on a dot plot… the range and SD are affected by outliers and the IQR is not. Because it is more to do with how far each ahm… point is from the ahm… mid… the mean. From the median, yeah. One of them m-words. And that's not really going to be affected by… And it is really not going to be affected by the ones way out at the end.

Students, who made comments, such as Mishka's comments above, with an allusion to the several interconnections among the various components, also provided responses to the previously open-ended CBJs at the multistructural level of the SOLO taxonomy.

Mishka was not the only student who was using a graphical representation of a data set when articulating her conceptions of measures of spread. In fact, the tendency to use visual imagery to express conceptual understanding of measures of spread was another theme that emerged during the analysis of student comments. Although we constructed and discussed the uses of other types of graphical representations of data sets, such as stem plots, histograms, and box plots, when articulating their conceptual understanding of measures of spread by visual references, many students mentioned dot plots only. Reflecting back on students' exclusive use of dot plots as a visual aid in articulating their conceptual understanding, I realized that when we conducted a group activity using an applet for predicting and checking the IQR and SD for various randomly created data sets, the applet only generated dot plots for the graphical representations of the data sets.

Bakker, Biehler, and Konold (2005) indicated that students new to statistics generally have a tendency to attend to individual cases as opposed to having an aggregate view of data. They considered helping students in developing an aggregate view of data along with a case oriented view as a fundamental challenge of statistics education. The group of introductory statistics students who participated in this study showed similar tendencies when their articulations were solely relied on dot plots, which were based on individual cases. In the spirit of Schön's (1983) idea of reflection-on-action, I decided to investigate the possibility of using applets or software which would generate dot plots as overlaid on box plots for the subsequent semesters to help students develop the individual case and aggregate views of data sets concurrently.

Analysis of the CAOS MOS Post-test Results

The CAOS MOS pre-test was administered during the second week of the semester with the CAOS MOS post-test planned to follow four weeks later. However, since the college was closed several days due to the snow and ice storms in the area, the actual administration dates of CAOS MOS pre-test and post-tests turned out to be six weeks apart from each other during the semester. The mean number of items students responded appropriately on the CAOS MOS post-test was 6.5 with a standard deviation of 2.4 and range from 2 to 11. The 95% confidence interval was given as CI=[5.6,7.4].

Most students indicated in their WJs that they were, to the contrary of the CAOS MOS pre-test, not guessing in choosing the appropriate responses on the CAOS MOS post-test. For example, Mabel, who improved from seven appropriate responses on the CAOS MOS pre-test to nine appropriate responses on the CAOS MOS post-test, made the following remark:

> The assessment online went a lot better for me, but some parts were still hard for me to guess, like standard deviation.

Other students commented that they were not as intimidated, more comfortable with, and confident in their responses on the CAOS MOS post-test than they were with their responses on the CAOS MOS pre-test. For example, Melis wrote:

> Then we went to take our assessment. I was thinking about how difficult it was the first time around so I wasn't too confident, but starting it actually felt I understood what they were asking about this time around. This time around the assessment wasn't as intimidating.

Unfortunately, since the number of appropriate responses went from five to six for her on the CAOS MOS post-test, the amount of improvement she demonstrated did not seem to echo her perception adequately. Mona, who improved from four appropriate responses to seven on the CAOS MOS post-test, commented as follows:

> We also went to math lab to redo our assessments on the computer. I didn't know anything the first time around, but this time I felt more confident in what the questions asked. I would hope I did better the second time.

Another student, Mevin, went from three appropriate responses to six on the CAOS MOS post-test, shared the following reflections and insight:

> The assessment test did not seem easy after knowing the basic concepts and terminology. I still struggle through some questions, but I did notice that we completed the test twice as fast compared to the first time.

An analysis of the completion times, prompted partly by Mevin's remark, revealed that the amounts of time that students took to complete the CAOS MOS post-test varied widely with a median completion time of 10.8 minutes, mean of 11.7, SD of 4.4, and range from 4.1 to 25.0 minutes. The typical completion time per item was about three-fourths of a minute. However, there was one student who spent 25 minutes on the CAOS MOS post-test. A further analysis of comparison revealed a statistically significant difference between the completion times of CAOS MOS pre-test and CAOS MOS post-test ($t(29)=-3.04$, $p<0.01$). As indicated by several students, the significantly shorter completion times on the CAOS MOS post-test might be attributed to the fact that students were not as intimidated, more comfortable with, and confident in their responses in the CAOS MOS post-test.

Almost half of the students (14 out of 29) who participated in the CAOS MOS post-test chose appropriate responses to at least half of the fourteen items. Four of these students (14%) provided appropriate responses to half of the items. Two students provided appropriate responses to eleven items, which was the highest number of appropriate responses given on the CAOS MOS post-test.

The next step in analyzing the CAOS MOS post-test data was to examine the responses given by students to fourteen individual items of the CAOS MOS post-test. At least 50% of the students were able to choose appropriate responses to more than half of the items (8 out of 14) on the CAOS MOS post-test. These eight items and the percentage of appropriate responses given by the students to each item are summarized in Table 9 below.

Table 9: Items of CAOS MOS post-test appropriately responded by 50% or more students

Item	**Description of Concept Investigated**	**% of appropriate responses**
1	Interpreting the meaning of a zero standard deviation for a distribution of data points	55.2 (16 out of 29)
2	Interpreting a given non-zero value of SD	55.2 (16 out of 29)
4	Comparing two mound-shaped histograms for the size of SD	71.4 (20 out of 28)
6	Comparing the SD of ages in the two populations of same size; all college students at a State university and all residents in a small town	75.9 (22 out of 29)
7	Estimating (without calculating) the SD of a tightly clustered set of data points with a given mean	58.6 (17 out of 29)
8	Estimating (without calculating) the SD of a set of data points with a given mean	69.0 (20 out of 29)
10	Recognizing the relationship between the outliers and the range for a skewed distribution represented by a histogram	51.7 (15 out of 29)
13	Identifying a distribution with the smallest IQR given a set of several distributions represented by box plots	55.2 (16 out of 29)

More than half of the students did not choose appropriate responses for the remaining six items. Only three students (10.3%) chose the appropriate response for one of those items, which was item 3. Furthermore, at least 75% of students did not choose appropriate responses for half of these six items. These three items with their percentages of appropriate responses are given in Table 10 below. Table 10 also displays the remaining three items and their respective percentages of appropriate responses. As indicated in Table 10, smallest percentage of appropriate response was for item 3, which was followed by items 5 and 12 with only four students (13.8%) choosing the appropriate responses for each one of these two items on the CAOS MOS post-test.

The percentage of appropriate responses given by students to the first group of three items in Table 10, which were items 3, 5, and 12, ranged from 10.3% (3 out of 29) to 13.8% (4 out of 29). The percentage of appropriate responses given by students for the second group of three items, which were items 9, 11, and 14, were not as low as the first group of three items and ranged from 37.9% (11 out of 29) to 48.3% (14 out of 29). These percentages of appropriate responses indicated that less than half of the students were able to give appropriate responses to 43% (6 out of 14) of the items on the CAOS MOS post-test. A quick glance at the above described CAOS MOS post-test percentages revealed a considerable contrast to the CAOS MOS pre-post percentages. In the next section, a comparison between the CAOS MOS pre-test and CAOS MOS post-test results is presented.

Table 10: Items of CAOS MOS post-test appropriately responded by less than 50% of students

Item	Description of Concept Investigated	% of appropriate responses
3	Comparing two histograms for the size of SD	10.3 (3 out of 29)
5	Recognizing the non-negative nature of SD	13.8 (4 out of 29)
12	Recognizing the relationship between the extreme values of a mound-shaped and symmetric distribution, and measures of center and spread	13.8 (4 out of 29)
9	Choosing appropriate measures of spread and center for a skewed distribution represented by a histogram	37.9 (11 out of 29)
11	Choosing an appropriate measure of spread for a skewed distribution represented by a histogram	39.3 (11 out of 28)
14	The relationships between the outliers and the measures of spread in terms of resistance to outliers	48.3 (14 out of 29)

Comparison of CAOS MOS Pre-test and Post-test Results

A comparative analysis of the CAOS pre-test and post-test results revealed several different relationships which shed light on the development of students' conceptual understanding of measures of spread. Upon examining Tables 6 and 9, the number of items to which appropriate responses given by at least half of the students increased by a factor of two, from four to eight items. On the CAOS MOS post-test, in addition to the pre-test items of 2, 4, 6, and 8, there were four additional new items, which were items 1, 7, 10, and 13. As indicated by the descriptions of these four additional items, more than half of the students were able to develop conceptual understanding and meaning of the various values of SD for a data set, as in items 1 and 7, recognize the inappropriateness of the range for skewed distributions, as in item 10, and compare spreads given a set of box plots using IQR, as in item 13.

As the percentage of appropriate responses to item 4 indicated, students were able to identify and articulate which distribution had a larger SD when comparing two normal distributions, as in item 4. On the other hand, item 3, instead of comparing two normal distributions for their SD, used a normal and a uniform distribution for the comparison of SD. Even though there was a slight increase, from 6.1% to 10.3%, in the percent of appropriate response to item 3 from CAOS MOS pre-test to post-test, item 3 remained elusive for almost 90% of students for possibly three reasons. Reflecting back on the activities and class discussions, there were none, but one case of a uniform distribution. Students mainly viewed the histogram of a normal distribution as an empty box and several students asked as to where the distribution was. There were not enough teaching activities to help students develop a sense of what a uniform distribution might represent. In the spirit of Schön’s (1983) idea of reflection-on-action, reflecting back on this particular result after the semester was over, I decided to search for teaching activities in order to adjust my teaching and to include more contextual cases of uniform distributions in the future semesters.

The second reason for item 3 to be elusive for students might be as a result of the fact that a great majority of students tended to view the variation along the vertical axis of a histogram as a cause for a larger SD value in a distribution. This specific view, which seems to be prevalent among students, was well documented and reported by other research studies (delMas & Liu, 2005; Meletiou & Lee, 2002; Turegun & Reeder, 2008), and seemed to be the case here as well. The third reason might be closely related to the second reason. When students were asked to compare a uniform distribution displayed by a histogram with same bar heights across, to a normal distribution displayed by a histogram with the various bar heights across, students' inclination to view the variation along the vertical axis possibly became even more accentuated.

Examining Tables 7 and 10 revealed considerable increases from the CAOS MOS pre-test to post-test in the percentage of appropriate responses to several items. Although various considerable increases, ranging from about one-and-half times to quadrupling, were observed for the appropriate responses to items 5, 11, 12, and 14 in the CAOS MOS post-test compared to the pre-test, these items were responded to appropriately by less than half of the students. Item 9, to which no students responded appropriately in the CAOS MOS pre-test, was appropriately responded by almost 40% of students (11 out of 29) in the CAOS MOS post-test.

This result was very encouraging because item 9, which was based on the recognition of a multitude of relationships among the measures of spread and measures of center, required a high level of cognitive development of students. In order to provide an appropriate response to item 9, it was necessary for students not only to have a high level of cognitive development but also to demonstrate conceptual understanding of the measures of spread as a comprehensive conceptual entity with all of its interconnected aspects. Even though there were no students who chose an appropriate response for this item at the beginning of the second week of the semester, six weeks later it was evident in the CAOS MOS post-test results that eleven students were able to exhibit a high level of cognitive development and demonstrate conceptual understanding of measures of spread as a comprehensive and interconnected conceptual entity.

In an effort to triangulate findings of the CAOS MOS post-test, I decided to examine if any of the eleven students, who provided an appropriate response to item 9, also provided open-ended CBJ responses at the relational level of the SOLO taxonomy. I was expecting that the two students with open-ended CBJ responses at the relational level to be among the eleven students. Surprisingly, the data revealed that neither one of the two students, who appropriately responded at the relational level of the SOLO taxonomy, was included among the eleven students who appropriately responded to item 9. All eleven students, who appropriately responded to item 9, provided either unistructural or multistructural level responses to the open ended CBJs. Prompted by this unexpected result, even though an expert colleague and I reached a consensus on the final coding of responses on the CBJs, I wanted to reconsider the responses to CBJs to determine if there was a discrepancy in the coding of those student responses. A further investigation did not reveal any discrepancies which might have resulted in modifying of the categories or coding of the responses.

On the other hand, this unexpected result was rather encouraging when viewed not from a perspective of which students were not among these eleven students, but viewed from a perspective of which students were among these eleven students. Since the ability of the eleven students, whose responses were at the unistructural and multistructural level on the CBJ assessment, to choose an appropriate response to item 9 on the CAOS MOS post-test might be an indication of the cognitive development of these students to a higher relational level of conceptual understanding of measures of spread.

Even though the design of this study was not focused on a pre-test, intervention, and a post-test scheme, when comparing two sets of data in the form of CAOS MOS pre-test and post-test, a question of whether the observed differences between these two data sets were statistically significant might be worth investigating. The mean number of items students responded to appropriately on the CAOS MOS pre-test was 4.4 with a standard deviation of 1.8 and range from 2 to 8. The mean number of items students responded appropriately on the CAOS MOS post-test was 6.5 with a standard deviation of 2.4 and range from 2 to 11.

The 95% confidence interval for the mean number of items student responded appropriately on the CAOS MOS pre-test was given as CI=[3.9,5.2]. The 95% confidence interval for the mean number of items student responded appropriately on the CAOS MOS post-test was given as CI=[5.6,7.4]. A comparison of these two 95% confidence intervals indicated that there was no overlap for these two confidence intervals. Since there these two 95% confidence intervals did not overlap, at the significance level of $\alpha=0.05$, there was a significant difference in the mean number of items students responded appropriately on the CAOS MOS pre-test and post-test. Further analysis using a matched-pairs test revealed a significant difference in the mean number of items student responded appropriately on CAOS MOS pre-test and post-test ($t(28)=5.5$, $p<0.00001$).

Although statistically significant, the increase in the mean number of items students responded appropriately from the CAOS MOS pre-test to the CAOS MOS post-test was a mean increase of about 2 items out of the 14 items (95% CI=[1.3,2.9]). It was somewhat disappointing to find out that students were able to provide appropriate responses, on average, to about half of the items on the mid-semester CAOS MOS post-test. A more comprehensive research study involving twenty higher education institutions and several hundred students reported much smaller gains from the pre-test to post-test based on the entire CAOS test consisting of 40 items (delMas et al., 2007).

Third Research Question

Several researchers point out the importance of establishing the connections between behavior of distribution of data sets, and conceptual understanding of statistical measures of spread and center in order for students, ranging from seventh grade students to pre-service teachers, to successfully understand informal inferential statistics (Gal & Short, 2006; Groth, 2006; Leavy, 2006; Liu & delMas, 2005; Pfannkuch, 2006; Reading & Reid, 2006; Slauson, 2008). In particular, understanding how distributions behave requires a focus on the interconnected features of a distribution, such as center, variation and shape. Conceptual understanding of the behavior of a data set as a distribution comprised of several of these features plays an important role in investigating descriptions, comparisons, and inferences based on that data set. In answering the third research question by way of fleshing out students' conceptions of measures of spread and the relations of these conceptions to the data distributions, I analyzed the data obtained from the CAOS measures of spread (MOS) post-test, weekly journal entries (WJ), my own journal entries, closed book journal (CBJ), and mid-term interviews (MCI). An analysis of the results is presented next.

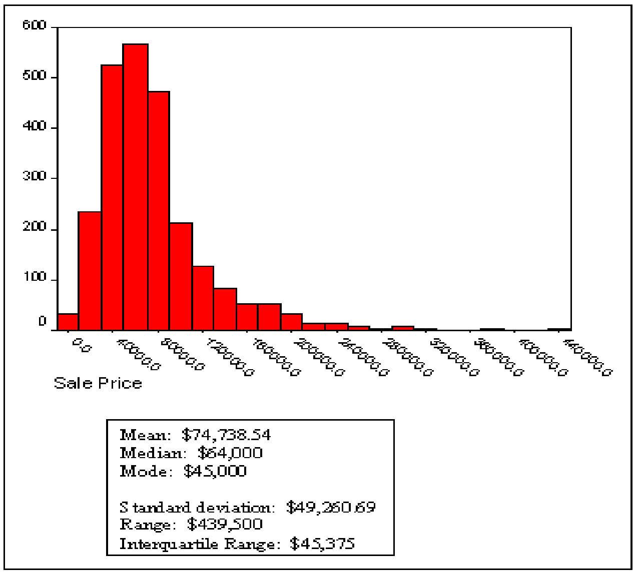
During the MCIs students were asked to comment on the connections between the statistical measures of spread and behavior of distributions of data sets. Initially students were asked to do so by being presented with an abstract setting without any particular context specified. For example, the following exchange between Michka and me illustrated this particular line of probing.

> Instructor (Inst): …Why would you want to choose one over another one of these three measures? …What kind of data sets would make you say that "oh, I see this particular aspect of the data set I think this data set would be better to be described by using the SD or IQR or… ".
>
> Michka: Right. Right.
>
> Inst: How would you respond to that?
>
> Michka: I am like refreshing my memory also. Ahm… because of the fact I think that IQR is more resistant to the change in outliers ahm… SD is … has less resistance, you know, to the outliers so I think it would be less accurate because you are throwing in all these weird plots on the outside and I think it affects kind of overall. So, I would say ahm… IQR would be… what was the other one? Range, IQR and SD. Ahm… whereas IQR is more resistant to outliers so I think it would be more accurate in kind of getting the basic concept of the information you are trying to… evaluate.

Michka was the only student who was able to respond to this line of probing which was couched in an abstract non-contextual setting. Her gain from the CAOS MOS pre-test to post-test was only minimal. Even though Michka did not provide an appropriate response to item 9 on the CAOS MOS pre-test and post-test, her response to the open-ended CBJ was one of the two highest level responses provided by students at the relational level of the SOLO taxonomy.

However, the rest of students needed a more contextual prompt and a less abstract line of probing. The search for a particular item conducive to achieving such a goal led me to the ARTIST site, from where the item given in Figure 2 was downloaded and included in the MCIs. The following passage was taken from a response given to this item by Masona. Not only did she need a more contextual prompt but she, similar to many students, also seemed to require more guidance as the following exchange took place during her MCI.

In January 2003, it was reported that 2440 homes were sold in the Twin Cities area in December of 2002. The sale prices of each home are displayed in the histogram, and measures of center and spread associated with this data set are also displayed.



Describe which measures of center and spread would be most appropriate to report for this distribution and why.

Figure 2. Assessment item on MCI

> Inst: …Is there a reason why we would want to choose one of these measures of spread, like the range, IQR, and SD over the others?
>
> Masona: Ahh, OK. Yeah. Well, when I first looked at that question I did not really understand what was asking…
>
> Inst: Hhm… Hhm… Can you describe the shape of this distribution?
>
> Masona: Skewed to the… left or… I mean skewed to the right … because tail is to the right. Right? I forget sometimes… I forget if the tail… wherever the tail goes… so, skewed to the right.
>
> Inst: Hhm… Hhm… And the next follow-up discussion is which one of these measures of spread would be more appropriate for this type of a data set?

Masona: I think the SD would be more off because of the big… the… a lot of outliers form the middle… ahm… and it says the inter quartile range is 45 and that's about in the middle… is… that's where the most of them are clustered at … ahm… I'll guess… I don't know… you see… that's why it's confusing because those two aren't really that different… or is that another comma or… I can't tell… is that… no it's not a comma… yeah, so it'll be the interquartile range.

Inst: And… do you think the IQR would be more appropriate for this data set in terms of describing the spread?

Masona: Ahm… you want the whole thing… whole spread or… or just interested in the… the… where most of them are… because if it's the whole thing… the range and the… ahm… the range has all the outliers and everything in it. But I don't know what… what the question is…

Inst: The question is given this distribution, which one of those three measures would be more appropriate to use to describe the spread of the distribution in the… home sale prices?

Masona: I'm not sure which one… would be… (long pause)… ahm… I guess I'm going to stick with the inter quartile range.

Inst: And… So, what's your reasoning for choosing that over the other two?

Masona: Because it represents the data better. If somebody asked me…

Inst: What do you mean it represents the data better?

Masona: That's where… I guess that's where the majority of the hoses fall… in that category…

Inst: Hhm… Hhm…

Masona: … you know, if somebody ask you what the average is they don't want to care about outliers.

Inst: Hhm… Hhm… So, why wouldn't you choose SD to represent or describe the spread for this data set?

Masona: Because there is just… one… really big, well, a few really big outliers that's more like a couple houses…

> Inst: Hhm… Hhm… and how is that going to affect your SD? What do you think?
>
> Masona: Makes it… a little higher… and it's not really representative of the whole.

Above comments made by Masona seemed to indicate a gradual shift from an individual case view of the distribution to that of an aggregate view as a whole. Bakker & Gravemeijer (2004) argue that one of the traits of an expert level statistical thinking and understanding is the ability to combine the local and global perspectives when using distributions. According to several researchers an understanding of a distribution as comprehensive entity with all of its connected components such as measures of spread and center hinges on perceiving data in a global way as opposed to focusing on only local features (Bakker, Biehler, and Konold, 2005; Pfannkuch, 2006).

Masona's above remarks suggested that she might be developing an ability to perceive a distribution by using local and global views alternately. However, she did not show an improvement from the CAOS MOS pre-test to post-test. In fact her number of appropriate responses decreased by one on the post-test. Masona did not provide an appropriate response to item 9 on the CAOS MOS pre-test and post-test, but her open-ended CBJ response was at the multistructural level of the SOLO taxonomy.

Another student, Miranda, had difficulties remembering some of the definitions, but with some prompting, she was still able to discuss her conceptions of measures of spread, and make some connections among the outliers and measures of spread. However, she did not seem to connect these further to the shape of the distribution. The following passage was taken out of her MCI to reflect her thought process in putting some of these connections together.

Inst: Which one of those three measures of spread like the range, IQR, and the SD would you say more affected by the presence of outliers?

Miranda: The presence of outliers? Ahm…

Inst: … and take your time think about it and then...

Miranda: OK.

Inst: … you need to kind of reason through it once you get your thoughts kind of collected.

Miranda: OK. Would be the SD? I mean… ahm… I remember that question being on the assessment…

Inst: Can you tell me why you went with SD?

Miranda: Because with IQR outliers on the outside… and IQR is in the middle. So, how would that affect the IQR…

Inst: Hhm… Hhm… That's good reasoning.

Miranda: … and ahm… I was thinking… see I'm stuck with range and SD. Ahm…

Inst: Is that going to help you if I say the range is between… the range is the difference between the smallest and largest values?

Miranda: OK. Hhm… Well, will it be range then because… the SD that's pretty much… that's like one SD 68… another one 95… I know that but… ahm… let's see… I'm sure the answer is like right in my face, I am sure. Ahm…

Inst: But, your reasoning about IQR not being affected is good.

Miranda: OK. Thank you.

Inst: So, you went with the SD as being affected by outliers. Range would be affected by outliers. Is that right?

Miranda: Yeah. It makes sense because range is the measure of variability from each side, you know. OK. It makes sense.

Miranda's last remarks alluding to the connection of the SD to the empirical rule of 68-95-99.7% was one of the common themes emerged throughout. In an effort to tie the SD as a measure of spread to the shapes of distribution, many students, similar to Miranda, were inclined to visualize a geometric representation of the SD in terms of the empirical rule of 68-95-99.7%. However, such a geometric representation of the SD can only be valid or developed in the context of normal distributions. When many students were visualizing the SD by tying it to the empirical rule of 68-95-99.7%, they were doing so without any regard given to whether or not the distribution was normal. Reflecting back on this, I realized again the perils of a purely rule-based teaching approach without any discussion of the specific set of limitations and conditions under which the rules might not be appropriate. Bakker, Biehler, and Konold (2005) argued in favor of a similar point and stated that the box plots and the interquartile range might be more useful for developing an initial conception of measure spread than other graphs or measures of spread.

During the later part of the discussion it became apparent that Miranda, similar to quite a few other students, had an unresolved issue in determining which way the distribution was skewed. The following passage presented below highlighted her confusion about the direction of skew.

> Inst: Let me show you something here. We have a distribution of home sale prices in a particular year at a particular location in the country. Can you tell me what the shape of this distribution is like?
>
> Miranda: OK. I think we just talked about this. So, it is skewed to the left…
>
> Inst: We talked about that at the beginning, yeah.
>
> Miranda: Yes. Because I was kind of confused I wanted to say the right because I would turn this side. But, it is skewed to the left. Yeah. OK. That makes sense.
>
> Inst: So, which way is it skewed?

Miranda: To the left. I am turned this way it is skewed to the left if I am turned this way it is skewed to the right. But, it is skewed to the left.

Inst: And why is that?

Miranda: Ahm… because it's more on a lower ends of the… of the chart.

Inst: Where do you think about the center of this chart located?

Miranda: The center… about 64. Yeah. Probably about there. [pointing to the tallest rectangle]

Inst: OK. So, now you are looking at each side of the center, right?

Miranda: Hhm… Hhm… Yes…

Inst: So, how far do we have to go on each side to go to the end of the data?

Miranda: About… to the end of the data… Further on the right.

Inst: So, which way is it stretched out more?

Miranda: To the right. You mean stretched out?

Inst: Hhm… Hhm…

Miranda: I think it's stretched out to the right more.

Inst: So, that's skewed to the right more?

Miranda: Ahm… ahm… OK… Tricky.

Inst: …from the center. Because if you look at from the center we go further to the right.

Miranda: Hhm… Hhm…

Inst: All the way 440K as opposed to…

Miranda: Oh… that way, yeah. So, it is skewed to the… Yeah. I see that now. So, it is skewed to the right. OK. I see what you are saying. OK. That's tricky. I mean… OK. I think I am making it harder than what it is.

Inst: Maybe you are thinking of like densities, you know, in… this is such a short distance but since we are sitting at the center here… …that means half of your data set is here… and the other half is over here.

Miranda: OK.

Inst: So, where are we loaded up more? Maybe, that's what you were thinking…

Miranda: Yeah. OK.

Inst: …more dense here.

Miranda: Over there, yeah.

Inst: Smaller area but you got half of your data set in… it's packed in here. This is more sparse. But, this is also 50% because that's where the median is, right?

Miranda: OK. Yeah. Makes sense.

Inst: But you are loaded up here more. Maybe that's why you are saying it is skewed to the left.

Miranda: OK. I'll have to remember that. That's why when I looked at that I see skewed to the left, so… I see what you are saying.

Inst: But, why? What's your reasoning?

Miranda: Because if I were to draw it be like that, you know, skewed to the left. You know, skewed to right… I mean that's how I see it. So, but when you explain that to me it makes sense. I don't know… I guess I see it differently.

Inst: Skewed to the left… I think you are thinking like most like… looks like most of the graph or the like bulk of the graph is…

Miranda: to the left. So, it is not like that at all. OK.

The above excerpt seemed to illustrate that since she was possibly viewing the distribution from a density point of view, as opposed to how sparsely the tail trailed off, she was considering the skewness to be in the opposite direction. In general, introductory statistics textbooks and teachers tend to take a rule oriented approach of telling students simply to point or follow the tail of a distribution when determining the direction of skew. But, for many students, such as Miranda, this approach might seem to be counter intuitive. According to Biehler (2004) and Pfannkuch (2006), one of the views that result from interpretation of spread is the view that focuses on regional densities. As indicated by the above excerpt, Miranda, like many other students, focused on the density view when looking at or visualizing spread and distributions.

Similar findings were reported by Lee, Zeleke, and Wachtel (2002) in their research on introductory statistics college students' conceptions of variation and its relation to other statistical concepts. They identified many students who had difficulties connecting different concepts of variation. One of the reasons for the difficulty was due to the fact that many students felt that there was an association between the direction of skewness and higher densities of values on distributions, as opposed to an association with outliers or sparsely placed data values.

We, as statistics educators, need to take advantage of this particular density view which students seem to bring in as a pre-existing view and use that particular approach, at least as an alternate approach, when discussing how to determine the direction in which a distribution might be skewed. Miranda did not show an improvement from the CAOS MOS pre-test to post-test. In fact, the number of appropriate responses she provided decreased by two on the post-test. However, Miranda provided an appropriate response to item 9 on the CAOS MOS pre-test and post-test and her open-ended CBJ response was at the multistructural level of the SOLO taxonomy.

Many students whose open-ended CBJ responses were at the unistructural level of the SOLO taxonomy seemed not to fully grasp and make the appropriate connections between statistical measures of spread and distributions, such as the one given in Figure 2. The following excerpt presented below was from Myrobi's MCI, and was a representative sample of the conceptual struggle of quite a few such students.

> Inst: So, let's say that… we have a… distribution of home prices in a given year at a certain place in the US.
>
> Myrobi: Hhm… Hhm…
>
> Inst: Can you describe the shape of this distribution of home sale prices?
>
> Myrobi: Describe it as in… I mean the first thing I can tell… I mean it peaks up over here… and then… there is the ahm… mean but it is greatly skewed to the right.
>
> Inst: Very good.
>
> Myrobi: I mean that's the first thing I see... it is just… skewed to the right, completely all the way over here [pointing to the right tail of the distribution]. And then it has a very I want to say, not large, or I guess large… in number… large of a number of an outlier… because that's way out here [pointing to the right tail of the distribution].
>
> Inst: Hhm… Hhm… That's very good.
>
> Myrobi: Ahm… that's the only thing I can kind of instantly see or observe… from the distribution. Would I also say that it is normal? I don't think it would be normal. It is skewed, right? So, that wouldn't be normal.
>
> Inst: Yeah. It is skewed to the right.
>
> Myrobi: So, it'll be skewed, not normal.
>
> Inst: Normal distributions tend to be more…
>
> Myrobi: Symmetrical.
>
> Inst: Hhm. Hhm. Given that the data set is skewed to the right, which one of the three measures of spread you would choose and considered to be most appropriate given that you have, like you said outliers here at the high-end and skewed to the right.
>
> Myrobi: I think the most appropriate measure would be the SD.

> Inst: And the reason would be...
>
> Myrobi: The reason would be… because it had to do with averaging… you are looking at how… how narrow or wide if you are looking at this… and I mean… I am not explaining this well… Because the median… in this would not be as important as in average when you are looking at all these prices and how many houses were sold. So… to me the most important would be the SD. Because you are trying to figure out the spread about the average, I guess, I am not sure how to say that.
>
> Inst: Hhm… Hhm…
>
> Myrobi: Or, what I am trying to say. But, I am going with the SD… as the most important one in this case.

Myrobi did not show an improvement from the CAOS MOS pre-test to post-test. He had the same number of appropriate responses on the pre-test and post-test. However, Myrobi provided an appropriate response to item 9 on the CAOS MOS post-test and his open-ended CBJ response was at the unistructural level of the SOLO taxonomy.

The importance of understanding how distributions behave in relation to the measures of spread and center has been stated by several researchers (Bakker & Gravemeijer, 2004; Ben-Zvi & Garfield, 2004). In answering my third research question, I have investigated the various ways of reasoning and approaches students exhibited in establishing the connections between the measures of spread and distributions of data sets. In the next chapter, I will summarize my conclusions, discuss the limitations of the study, and provide suggestions for further research.

CHAPTER FIVE

CONCLUSIONS AND IMPLICATIONS

> *No one can tell another person in any definite way how he should think, any more than how he ought to breathe or to have his blood circulate. But, the various ways in which men do think can be told and can be described in their general features.*
>
> Dewey (1933, p. 3)

Overview

The foregoing statement not only provided a challenge, but also constituted a major goal for me to achieve in this research study, which was conducted to investigate community college students' conceptual understanding of statistical measures of spread and how students establish the connections between measures of spread and behavior of distributions of data sets. The study, which was designed to collect data from a variety of quantitative and qualitative data sources in order to triangulate findings, focused on answering the following research questions:

1. What are some of the pre-existing conceptions of students regarding statistical measures of spread?
2. How do students articulate their conceptual understanding of statistical measures of spread?
3. How do students establish connections between statistical measures of spread and behavior of distributions of data sets?

In this chapter, I outline the significance of the study and the particular niche that the study aims to fill within the existing body of knowledge in statistics education community. I offer a discussion of my findings and describe their implications for teaching and learning of statistics. A summary of the limitations of and concerns with the study is presented, as well.

Why Focus on Community College Students?

One of the rare features of this study was its sole focus on community college students. Much of the research conducted at the undergraduate level has focused on students at the four-year colleges and universities. Yet, almost half of the first time college students are enrolled at two-year community colleges nationwide. These demographics combined with the open admission system of community colleges make the community college student population quite large, diverse, and considerably different than the traditional student body found at many of colleges and universities across the nation. Teaching introductory statistics to such a diverse body of community college students is not an easy mission. Even though there is evidence that community college students are not only an underserved population, and that they also perceive mathematics and statistics as a barrier to their academic success, there are only a few studies which have focused on community college students' understanding of introductory statistics (Brandsma, 2000; Prichard, 1995; Priselac, 1995).

One of the issues recognized in statistics education is the lack of research to guide practice. There is a need for continued scholarly activity to determine the nature of students' understanding of introductory statistics concepts at various levels. However, several researchers indicated that there was still is a disconcertingly small body of research regarding community college students' understanding of introductory statistics (Brandsma, 2000; Slauson, 2008). Hence, the focus on this underserved and heterogeneous population of community college students was important. I believe that the results of this research on teaching and learning of introductory statistics at the community college level has the potential to confirm and possibly expand on the previous results, and provide unique insight into the teaching practices of community college instructors so that they can improve teaching and learning of introductory statistics to the benefit of this educationally underserved and quite heterogeneous population of community college students.

Frequent Absenteeism and High Attrition Rates

In addition to answering the above presented research questions, the study also revealed some additional insight and shed light on other related aspects of teaching and learning introductory statistics at the community college level, in general, and statistics education, in particular. Having taught at various community colleges in the last twenty-three years, I observed that frequent absenteeism and one of its dire consequences in the form of drastically low retention rates were detrimental to student success among community college students, especially in mathematics and introductory statistics courses.

Whether the frequent absenteeism and high attrition rates are attributed to increased student work hours due to harsh economic conditions, changes in student attitudes, varied academic background and preparation as a result of the open enrollment policy of community colleges, different learning styles, and large classroom sizes, or more likely to a combination of all of the above, other researchers and the community college faculty have taken a notice of this disconcerting pattern as well. For example, Brandsma (2000) reported that high attrition rates were widespread in community college mathematics courses throughout the United States.

Having been keenly aware of these problems over the years, I decided to implement a very strict attendance policy to curtail absenteeism. The attendance policy I implemented was not simply based on penalizing absenteeism, but also focused on fostering continuity in the learning process and encouraging responsibility for one another in the classroom community of learners. The Course Information document, which is presented in Appendix A, provided to the students in the course included the following statement which outlined an explanation of roles, and articulated several justifications for class preparation, attendance and participation.

> Every member of our class is considered to be both a learner and a teacher. You will need to assume responsibility for both roles. A major responsibility, both as a learner and a teacher, is to attend class regularly, arrive on time, stay for the duration of the class, and prepare by studying all materials and completing all assignments carefully and on time. Thus, you should engage in class activities and conversations, contribute to collaborative learning activities in a meaningful, professional way, demonstrate respect for all participants, and help ensure that everyone is permitted and encouraged to share equally in class opportunities and responsibilities. Your attendance and contribution to the development of a healthy and active classroom community is essential. It is ***extremely*** unlikely that you will earn an "A" for this course if you miss more than five classes for any reason, or a "B" if you miss more than ten classes. Please note that even perfect attendance does not guarantee an "A" or "B." In addition, if you miss a class for any reason, you will be expected to write a make-up paper for any class session you may miss. It would be your responsibility to contact at least one classmate to discuss what occurred during the time you missed class in order to successfully complete the make-up paper. In the paper, you will include closed-book journal activities, journal activities, and a detailed summary of any other topic we discussed and worked on during the class you missed. The paper will be due at the beginning of class period that you return.

The Course Information document was handed out on the first day of class to inform students of the structure of the course. Students began the semester aware of these guidelines. There was hardly anybody missing any class periods, with the exception of a few students, who later withdrew from the course.

Typical retention and success rates are about 50% or so for introduction to statistics courses at community colleges in the nation (Brandsma, 2000). Retention rates were considerably high with only a total of four students (12%) withdrawing from the two sections of the course. In general, the retention rates in the mathematics department at this community college are reported to be around 45%. Student completion rates in the course were considerably high when compared to average completion rates of 50-60% in mathematics courses offered by the mathematics departments at this community college. Final course grades seemed to indicate that not only the retention rates, but also the success rates were quite high. Only one student did not complete the course successfully. Only four students (14%) performed below a final course grade level of C. Eight students (28%) received an A, ten students (34%) received a B, and seven students (24%) received a C for the final course grade. The two D grades were received by the students who did not turn in their final portfolios, which was worth 25% of the final course grade.

Conclusions

According to one of the ASA recommendations, there is a need for high school students to become engaged in various aspects and techniques of statistics from descriptive statistics to inferential statistics. Similarly, NCTM (1989 and 2000) recommended more emphasis on data analysis, in particular, and statistics as a subject, in general, at the secondary level. Groth (2003) pointed out that there seemed to be some evidence to indicate that introductory statistics as a subject was increasingly becoming widespread at the secondary level. However, my results revealed very little evidence to support the widespread emphasis on introductory statistics at the secondary level for the group of students who were in this study. Based on the self-reported backgrounds of students, approximately one-third of students indicated that they were not familiar with statistics at all.

There were only two students (6%) who had either taken or used statistics in a meaningful way when they were in high school. One of these students was engaged in collecting data to perform a statistical analysis in order to determine how often the moon was a full moon and the movement of the earth. The other student indicated being in a physics class in high school when she was engaged in collecting data for a linear regression analysis. It is quite possible that many of these high school students who were in the widespread regular or Advanced Placement (AP) statistics classes in high school either may not be attending a community college or at least the community college where this study was conducted or they received college credit for an AP statistics course in high school. On the other hand, students' very limited pre-existing conceptions of measures of spread aligned with the findings of a number of researchers who indicated that the curricular emphasis in high school and college has been traditionally placed on measures of center, rather than measures of spread (Shaughnessy & Ciancetta, 2002; Watson, Kelly, Callingham, and Shaughnessy, 2003).

Reflection-on-Action and Implications for Teaching and Learning

There are a couple of implications from a constructivist perspective. One of the findings of the study pointed out that students used their own terminology when articulating their conceptual understanding of measures of spread. Students' use of their own terminology to express their own understanding is the very essence of a constructivist perspective. Teachers need to be aware of the fact that students may use non-standard or non-technical terms or even make up terms that make sense to them in order to express their conceptual understanding of measures of spread. Students' prior conceptions of spread revealed that not recognizing several cases of statistical *faux amis* might prevent teachers from helping students eventually develop a more appropriate statistical terminology. Likewise, Makar & Confrey (2005) points out the importance of learning to recognize non-standard language when students try to articulate their conceptual understandings of variation.

Teachers should be not only familiar with, but also initially foster various forms of non-technical terminology use by students to express their conceptual understanding of measures of spread. However, student's persistent use of inappropriate terminology may become a problematic issue to address if we do not help students develop enough appropriate statistical terminology to express their conceptual understanding of measures of spread. Subsequently, the student use of technical terminology could be developed gradually based on the initial use of a student's own terminology.

Secondly, constructivism also emphasizes that a student's mind cannot be viewed as an empty cisterns into which information to be pipelined. To the contrary, constructivism recognizes that students bring their pre-existing conceptions of measures of spread into introductory statistics. The evidence of students' ability to organize the concepts of spread in way meaningful to them individually is one of the results revealed by this study. This particular result both reinforces and is reinforced by constructivist/student centered theories, and indicates the applicability of these theories to college statistics students. The results of this study also suggested that students tended to possess intuitive understanding of measures of spread at levels which were quite high.

This was rather an unexpected conclusion, since students' self-reported backgrounds in mathematics and statistics were weak or almost non-existent.

This result might indicate that students do not give themselves credit for their already existent conceptual understanding or have confidence in their intuitive abilities. Similar conclusions were drawn by other researchers in their own research studies (Garfield, 2007; Slauson, 2008). We, as statistics educators, need to find ways to frame, retain, and help students further develop this prior intuitive knowledge about the measures of spread. It is possible that one way to accomplish this might be focusing on not using formulae and procedures as much, since the findings here indicated that most students, especially at the unistructural level and some at the multistructural level, showed a tendency to think that they need or must use formulae to provide appropriate responses.

However, it proved to be a challenge, since most students at the unistructural level quite often showed evidence, in various data sources, of a computation and formula oriented mindset, which was persistent throughout the course. The challenge lies in the fact that we, as statistics educators, need to be aware of how much cognitive dissonance an individual student can withstand before disequilibrium, necessary for a shift in mindset, become, as Dewey (1938) put it, a miseducative experience to eventually bring the educational process and development of students to a halt. We need to be attentive to establishing this delicate balance between dissonance and safety. A computation and formula oriented mindset should not be upset completely, but just enough cognitive dissonance is needed to be introduced into the computational mindset of students.

On the other hand, the group of students whose responses were at the multistructural and relational levels of the SOLO taxonomy demonstrated a better ability to describe and connect various components of the measures of spread as a conceptual entity, and make further connections between the measures of spread and how distributions of data sets behave. Students whose responses were at the unistructural level rarely demonstrated that ability, and hardly made connections between the measures of spread and behavior of distributions of data sets. Similar conclusions were also drawn by Reading & Reid (2006) in their research on the effects of students' conceptual understanding of variability, spread and density on understanding the concept of sampling distribution of sample means.

The notion of density when viewing the graphical representations of distributions of data sets was one of the pre-existing conceptions students had coming into the course. Many students were able to view distributions from a density point of view. However, a traditional approach of discussing skewness and its direction in terms of stretched out tails of distributions seemed to be counter intuitive for many students. According to Biehler (2004) and Pfannkuch (2006), one of the views that result from interpretation of spread is the view that focuses on regional densities. Many students felt that there was an association between the direction of skewness and higher densities of data values on distributions, as opposed to an association with outliers or sparsely placed data values.

We, as statistics educators, need to take advantage of this particular density view, which students seem to bring in as a pre-existing view, and use that particular approach, at least as an alternate approach, when discussing how to determine the direction in which a distribution might be skewed. Similarly, as delMas (2001) pointed out, we need to help students view the SD as a measure of the density of values about the mean of a distribution, and become more aware of how clusters, gaps and extreme values affect the SD.

Despite the fairly common density view of distributions, the findings of the study revealed that uniform distributions, as straightforward as we might think of them to be, were troublesome for many students. We need to remember that, as teachers, we tend to "underestimate the difficulty students have in understanding basic concepts of probability and statistics" (Garfield, 1995, p. 31). It should not be taken for granted that uniform distributions are not a problematic topic for students. Since the uniform distributions represent the concept of uniform density, they should be addressed and used more in introductory statistics in order to support and further emphasize the density view which was prevalent among some students.

There was evidence that students' understanding of various graphical representations of distributions is an area which requires much needed development. Likewise, several researchers reported that students' conceptions of spread commonly included treating the "bumpiness" of a graph as a measure of spread (Bakker, 2004; Meletiou & Lee, 2002; Turegun & Reeder, 2008). Since many students in this study also seemed to be confusing horizontal and vertical axis of histograms when determining spread, we need to think about how students can be helped to improve their ability to recognize the difference between what the horizontal and/or vertical axes represent in a variety of different graphical representations, such as histograms, dot plots, and box plots.

Similar to the density view of distributions, an aggregate view of distributions should be emphasized and developed simultaneously with an individual case view.

Tendencies shown by some students in articulating measures of spread with a sole focus on individual cases indicate a need to help students develop an aggregate view of data along with a case oriented view. As pointed out by Bakker, Biehler, and Konold (2005), one of the fundamental challenges facing statistics education community is to facilitate the simultaneous development of aggregate and individual case views of distributions. The combined and simultaneous use of individual and aggregate views of distributions of data sets is an essential facet of informal inference.

The notion of measures of spread, as a conceptual entity, is lately becoming more of a focus of statistics education research community because of an increasing and early emphasis placed on informal inference. Especially when comparing variability within and between two data sets, spread becomes an important element of reasoning (Pfannkuch, Budgett, Parsonage, and Horring, 2004). The SOLO taxonomy framework, which was proposed here for assessing conceptual understanding of measures of spread qualitatively, could provide teachers with some guidelines in terms of the overall aims and purposes of measures of spread that need to be attended to and developed when teaching this particular topic. Paying particular attention to helping students recognize and develop the interconnections among the various components of measures of spread is a necessary element for understanding statistical informal inference.

As indicated by Gal & Garfield (1997), many studies focused on traditional multiple choice or short answer tests and quizzes to measure statistical understanding in introductory statistics courses. However, alternative methods of assessment are necessary in order to measure students' conceptual understanding of various statistical topics (Garfield, 1995). The implementation of some of these alternative methods of assessments, such as semi-structured interviews and other open-ended questioning techniques in this study revealed that they are much more conducive to learning about and better in assessing students' conceptual understanding of measures of spread. The findings of the study indicated that students tend to actually learn from these types of continuous formative assessment techniques, as opposed to traditional summative assessment methods.

While these types of alternative methods of assessment are a useful way to integrate assessment and teaching by improving and assessing student learning simultaneously, as suggested by GAISE recommendations, creating such assessment items can be a daunting task for and place time demands on many teachers. The ARTIST and CASUE sites proved to be very useful resources for locating such alternative assessment items which can be used to integrate assessment and teaching to improve students' learning. The content, pedagogy and technology related changes suggested by GAISE recommendations, and some of the related changes implemented by this study can also be daunting tasks for teachers, in general. Based on my personal experience, it is particularly important to note that teachers' incomplete content knowledge and insufficient understanding of the subject matter may constitute an initial roadblock in terms of initiating their own transformations as teachers when implementing some of the changes presented in this study.

Recommendations for Future Research

There were several studies, such as Slauson (2008), and Brandsma (2000), which used a control and treatment design to study whether reform inspired approaches on certain statistical topics would facilitate deeper understanding, and allow students to develop a better conceptual understanding of those statistical topics. One might pose a question of whether a reformed curriculum approach and a textbook, such as Workshop Statistics - Discovery with Data and the Graphing Calculator by Rossman, Chance, and Von Oehsen (2008), would facilitate deeper understanding of measures of spread, and allow students to develop a better conceptual understanding of measures of spread, as opposed to using a traditional approach.

Even though research on students' understanding of variability and measures of spread is slowly growing at the community college and university level, there have been limited research efforts with respect to upper level undergraduate, graduate, and doctoral level students in statistics courses. Research studies focusing on graduate students might inform us whether there is a similar pattern of difficulties and conceptual gaps in the understanding of measures of spread that continues to propagate through that level in a persistent manner.

Since several researchers indicate that there is a strong connection among the conceptual understanding of distributions, measures of spread and informal inference, future research studies should examine whether an improved student understanding of measures of spread leads to better understanding of sampling distribution and statistical informal inference. There is a lot more to learn about the nature of how students come to understand the concepts of measures of spread. Continued scholarly activity probing into the nature of students' understanding of introductory statistics topics is needed in order for research to generate various theories to guide our practices, in the spirit of praxis.

Developing a framework is considered to be a continuous work-in-progress. Likewise, the SOLO taxonomy framework proposed here for assessing students' conceptual understanding of measures of spread qualitatively is far from being a finished product. There is a need to further investigate the levels of conceptual understanding of measures of spread among community college students.

More empirical studies should be conducted in order to collect more evidence to validate, add on to or revise the SOLO taxonomy framework proposed here. Even though the framework proposed in this study may be incomplete, it could provide teachers with some basic guidelines for helping students recognize and develop the interconnections among the various components of measures of spread, in particular, and for curriculum development in introductory statistics, in general.

Limitations of the Study

Even though the aim of this research study was not to generalize the findings to a larger population or transfer to other contexts, it might be noted that students self-selected to enroll into the two sections of the introductory statistics course, which were taught by the researcher. Since there were eight more sections of the course during the semester, the students enrolled in the two sections did not comprise the entire introductory statistics student population at the community college during the semester in which the research was conducted.

Although I was aware of my role as teacher/researcher, it may not be easy to determine the possible influence or bias my dual role might have carried over into the study. However, conducting of the MCIs required a conscious effort on my part to resist the lure of taking the MCIs as a teaching opportunity or treating them as a “teachable moment” or leading students to appropriate responses. I caught myself being lured into doing that on a couple of occasions during a part of an MCI, and thus decided to exclude that particular piece of data from the study.

Finally, the use of qualitative analysis methodology introduces a certain limitation. A qualitative analysis of the same data by different researchers or by the same researcher at different times might potentially lead to different results or conclusions. Having been the only researcher in this study, I tried to ensure and increase validity of the interpretation of the classroom data by triangulating a variety of data sources. Secondly, I asked an expert colleague to assess student responses and then checked the results for inter-rater reliability. In cases of a couple of disagreements, a discussion ensued until a consensus was reached.

REFERENCES

American Statistical Association (2007). *Using statistics effectively in mathematics education research.* [Online: www.amstat.org/research_grants/pdfs/SMERReport.pdf]

Ball, D. L. (2000). Working on the Inside: Using One's Own Practice as a Site for Studying Teaching and Learning. In A. E. Kelly, & R. A. Lesh (Eds.), *Handbook of Research Design in Mathematics and Science Education*, (pp. 365-401). Lawrence Erlbaum Publishers, Mahwah, NJ.

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, *3*(2), 64-83. [Online: www.stat.auckland.ac.nz/serj/]

Bakker, A., Biehler, R., and Konold, C. (2005). Should young students learn about box plots? In G. Burrill, and M. Camden (Eds.), *Curricular Development in Statistics Education: International Association for Statistics Education Roundtable*, Lund, Sweden, (pp. 163-173). Voorburg, The Netherlands: International Statistical Institute. [Online: www.stat.auckland.ac.nz/~iase/publications.php]

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi, and J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, reasoning, and Thinking*, (pp. 147-168). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Batanero, C., Garfield, J. B., and Ottaviani, M. G. (2001). Building a research agenda for statistics education: A response to reactions in SERN 2(1). *Statistics Education Research Newsletter*, *2*(2), 9-14.

Begg, A. (2004). Statistics curriculum and development: New ways of working. In G. Burrill, and M. Camden (Eds.), *Curricular Development in Statistics Education: International Association for Statistics Education.* Voorburg, The Netherlands: International Association for Statistical Education. [Online: www.stat.auckland.ac.nz/~iase/index/php]

Ben-Zvi, D. (1999). Alternative assessment in statistics education. In *Proceedings of the 52nd Session of the International Statistical Institute, 3*, 175-176. Helsinki, Finland: Edita Ltd., on behalf of the International Statistical Institute.

Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, reasoning, and Thinking* (pp. 3-15). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Biehler, R. (2004). Variation, co-variation, and statistical group comparison: Some results from epistemological and empirical research on technology supported statistics education. Paper presented at the Tenth International Congress on Mathematics Education, Copenhagen, Denmark. 4-11 July, 2004.

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

Bill, A., & Watson, J. (2007). Three student tasks in a study of distribution in a "best practice" statistics classroom. In J. Watson & K. Beswick (Eds.), *Proceedings of the 30th annual conference of the Mathematics Education Research Group of Australasia*, (Vol. 1, pp. 123-132). Tasmania, Australia. 2-6 July, 2007.

Brandsma, J. A. (2000). Data collection and analysis: Examining community college students' understanding of elementary statistics through laboratory activities. Unpublished doctoral dissertation. North Carolina State University.

Briggs, J., & Peat, F. D. (1999). *Seven Life Lessons of Chaos: Spiritual Wisdom from the Science of Change.* New York: New York: Harper and Row Publishers.

Broers, N. J. (2006). Learning goals: The primacy of statistical knowledge. Paper presented at the International Conference on Teaching Statistics, Salvador, Bahai, Brazil.

Budé, L. (2006). Assessing students' understanding of statistics. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics: Working cooperatively in statistics education*, Salvador, Brazil. Voorburg, The Netherlands: International Statistical Institute.

Budé, L. (2007). On the improvement of students' conceptual understanding in statistics education. Unpublished doctoral dissertation. Universiteit Maastricht.

Chance, B. L. (1997). Experiences with authentic assessment techniques in an introductory statistics course. *Journal of Statistics Education* [Online], *5*(3). http://www.amstat.org/publications/jse/v5n3chance.html

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

Chance, B., delMas, R., and Garfield, J. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education* [Online], *7*(3). http://www.amstat.org/publications/jse/v7n3/chance.html

Chance, B. L., & Garfield, J. B. (2002). New approaches to gathering data on student learning for research in statistics education. *Statistics Education Research Journal*, *1*(2), 38-41. [Online: www.stat.auckland.ac.nz/serj/]

Chance, B., Garfield, J., and delMas, R. (2000). Developing simulation activities to improve students' statistical reasoning. *Proceedings of the International Conference on Technology in Mathematics Education,* Auckland, New Zealand. 11-14 December, 2000.

Chance, B., delMas, R., and Garfield, J. (2004). Reasoning about sampling distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking,* (pp. 295-323). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Chance, B. L., & Rossman, A. J. (2001). Sequencing topics in introductory statistics: A debate on what to teach when. *The American Statistician, 55*(2), 140-144.

Chance B. L., & Rossman, A. J. (2006). Using simulation to teach and learn statistics. Paper presented at the 7th International Conference on Teaching Statistics (ICOTS-7), Salvador, Bahai, Brazil. [Online: http://www.stat.auckland.ac.nz/~iase/serj/publications/17/7E1_CHAN.pdf]

Cobb, G. (1992). Teaching statistics. In L. Steen (Ed.), *Heeding the Call for Change*, (pp. 3-34). MAA Notes No. 22, Washington: Mathematical Association of America.

Cobb, G. (1993). Reconsidering statistics education: A national science foundation conference. *Journal of Statistics Education* [Online], *1*(1). http://www.amstat.org/publications/jse/

Collins, L. B., & Mittag, K.C. (2005). Effect of calculator technology on student achievement in an introductory statistics course. *Statistics Education Research Journal*, *4*(1), 7-15. [Online: http://www.stat.auckland.ac.nz/serj/]

Cooper, L. L., &, Shore, F. S. (2008). Students' misconceptions in interpreting center and variability of data represented via histograms and stem-and-leaf plots. *Journal of Statistics Education* [Online], *16*(2). http://www.amstat.org/publications/jse/v16n2/cooper.html

Cross, K. P., & Steadman, M. H. (1996). *Classroom research: Implementing the scholarship of teaching.* San Francisco: Jossey-Bass Publishers.

Davis, B., & Sumara, D. (2007). Complexity Science and Education: Reconceptualizing the Teachers' Role in Learning. *Interchange, 38*(1), 53-67.

Dantonio, M., & Beisenherz, P. (2001). *Learning to question, questioning to learn: Developing effective teacher questioning practices.* Needham Heights, MA: Allyn and Bacon.

delMas, R. C. (2001). What makes the standard deviation larger or smaller? [Online: http://www.causeweb.org/repository/StarLibrary/activities/delmas2001]

delMas, R. C. (2002). Statistical literacy, reasoning, and learning. *Journal of Statistics Education* [Online], *10*(3). http://www.amstat.org/publications/jse/secure/v10n3/delMas_intro.html

delMas, R. C., Garfield, J. B., and Chance, B. L. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education* [Online], *7*(3). http://www.amstat.org/publications/jse/secure/v7n3/delMas.cfm delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal, 4*(1), 55-82.

delMas, R., Ooms, A., Garfield, J., and Chance, B. (2006). Assessing students' statistical reasoning. Paper presented at the 7th International Conference on Teaching Statistics (ICOTS-7), Salvador, Bahai, Brazil.

delMas, R., Garfield, J., Ooms, A., and Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6*(2), 28-58. [Online: http://www.stat.auckland.ac.nz/serj]

Devlin, K. (2007). What is conceptual understanding? Mathematical Association of America (MAA) Columns, Devlin's Angle. [Online: http://www.maa.org/devlin/devlin_09_07.html]

Dewey, J. (1933). *How We Think. A restatement of the relation of reflective thinking to the educative process*. Boston: D. C. Heath.

Dewey, J. (1938). *Experience and education.* New York: Teachers College Press.

Everson, M., Zieffler, A., and Garfield, J. (2008). Implementing new reform guidelines in teaching introductory college statistics courses. *Teaching Statistics. 30*(3), 66-70.

Feldman, A., & Minstrell, J. (2000), Action Research as a Research Methodology for the Study and Teaching of Science. In A. E. Kelly, & R. A. Lesh (Eds.), *Handbook of Research Design in Mathematics and Science Education*, (pp. 429-456). Lawrence Erlbaum Publishers, Mahwah, NJ.

Forster, P. A. (2006). Assessing technology-based approaches for teaching and learning mathematics. *International Journal of Mathematical Education in Science and Technology, 37*(2), 145-164.

Freedman, D., Pisani, R., and Purves, R. (1978). *Statistics*. New York: W. W. Norton & Co.

Freire, P. (1970). *Pedagogy of the oppressed.* New York: The Continuum International Publishing Group Inc.

Gal, I. (2003). Teaching for statistical literacy and services of statistics agencies. *The American Statistician, 57*(2), 80-84.

Gal, I., & Garfield, J. (1997). Curricular goals and assessment challenges in statistics education. In I. Gal and J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education.* (pp. 1-13). [Online: www.stat.auckland.ac.nz/~iase/publications/assessbkref]

Gal, I., & Short, T. (2006). Editorial. *Statistics Education Research Journal, 5*(2), 2-3.

Garfield, J. B. (1993). Teaching statistics using small-group cooperative learning. *Journal of Statistics Education* [Online], *1*(1). http://www.amstat.org/publications/jse

Garfield, J. B. (1994). Beyond testing and grading: Using assessment to improve student learning. *Journal of Statistics Education* [Online], *2*(1). http://www.amstat.org/publications/jse

Garfield, J. B. (1995). How students learn statistics. *International Statistics Review, 63*(1), 25-34.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education*, *19*(1), 44-63.

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

Garfield, J., & Ben-Zvi, D. (2008). Helping students develop statistical reasoning: Implementing a statistical reasoning learning environment. *Teaching Statistics. 31*(3), 72-77.

Garfield, J., & Ben-Zvi, D. (2009). *Developing students' statistical reasoning: connecting research and teaching practice*. Dordrecht, The Netherlands: Springer.

Garfield, J., & Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematical Thinking and Learning, 2*(1&2), 99-125.

Garfield, J. B., & Gal, I. (1999). Assessment and statistics education: Current challenges and directions. *International Statistical Review, 67*(1), 1-12.

Garfield, J., delMas, R., and Chance, B. (1999). *The role of assessment in research on teaching and learning statistics.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Garfield, J., Hogg, B., Schau, C., and Whittinghill, D. (2002). First course in statistical science: The status of educational reform efforts. *Journal of Statistics Education* [Online], *10*(2). http://www.amstat.org/publications/jse/v10n2/garfield.html

Giaud, G. (1997). Cooperative learning and statistics instruction. *Journal of Statistics Education* [Online], *5*(3). http://www.amstat.org/publications/jse/v5n3/giraud.html

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

Glencross, M. J. (1988). A practical approach to the central limit theorem. *Proceedings of the Second International Conference on Teaching Statistics*, Victoria, B.C.: The

Organizing Committee for the Second International Conference on Teaching Statistics, 91-95.

Gordon, S. (1995). A theoretical approach to understanding learners of statistics. *Journal of Statistics Education* [Online], *3*(3). http://www.amstat.org/publications/jse/v3n3/gordon.html

Greenwood, D. J., & Levin, M. (1998). *Introduction to action research.* Thousand Oak, CA: SAGE Publications, Inc.

Groth, R. E. (2003). Development of a high school statistical thinking framework. Unpublished doctoral dissertation. Illinois State University.

Hassad, R. (2008). Reform-oriented teaching of introductory statistics in the health, social and behavioral sciences – Historical context and rationale. *World Academy of Science, Engineering and technology*, *40*(1), 398-403.

Hawkins, A. (1996). Myth-Conceptions. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: 1996 proceedings of the 1996 IASE round table conference* (pp. 1-14). International Statistical Institute.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws, (Ed.), Handbook of research in mathematics teaching and learning: A project of the National Council of Teachers of Mathematics, (pp. 65-97). New York: MacMillan Publishing Company.

Heibert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics,* (pp. 1-27). New Jersey: Lawrence Erlbaum Associates, Inc.

Houser, N. (2006). Worldviews and learning theories: A transactional reconsideration of teaching and learning. *Curriculum & Teaching*, *21*(1), 5-32.

Inzuna, S. (2006). Some conceptions and difficulties of university students about variability. In S. Alatorre, J. L. Cortina, M. Saiz & A. Mendez (Eds.), *Proceedings of the 28th annual meeting of the International Group for the Psychology of Mathematics Education,* (Vol. 2, pp. 244-250). Merida, Mexico: Universidad Pedagogica Nacional.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26.

Jones, G., Thornton, C., Langrall, C., Mooney, E., Perry, B., and Putt, I. (2000). A framework for characterizing children's statistical thinking. *Mathematical Thinking and Learning*, *2*(4), 269-307.

Kang, N. (2007). Elementary teachers' teaching for conceptual understanding: Learning from action research. *Journal of Science Teacher Education, 18*(4), 469-495.

Keeler, C. M., & Steinhorst, R.K. (1995). Using small groups to promote active learning in the introductory statistics course: A report from the field. *Journal of Statistics Education* [Online], *3*(2). http://www.amstat.org/publications/jse

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education* [Online], *3*(1). http://www.amstat.org/publications/jse/v3n1/konold.html

Lane, D. M., & Peres, S. C. (2006). Interactive simulations in the teaching of statistics: Promises and pitfalls. Paper presented at the 7th International Conference on Teaching Statistics (ICOTS-7), Salvador, Bahai, Brazil.

Leavy, A. (2006). Using data compression to support a focus on distribution: Examining preservice teachers' understandings of distribution when engaged in statistical inquiry. *Statistics Education Research Journal*, *5*(2), 89-114. [Online: http://www.stat.auckland.ac.nz/serj]

Lee, C., Zeleke, A., and Wachtel, H. (2002). Where do students get lost: The concept of variation? Paper presented at the Sixth International Conference on Teaching Statistics: Developing a statistically literate society, Cape Town, South Africa. July, 2002.

Liu, Y., & delMas, R. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, *4*(1), 55-82.

Lovett, M. C., & Greenhouse, J. B. (2000). Applying cognitive theory to statistics instruction. *The American Statistician, 54*(3), 196-206.

Lunsford, M. L., Rowell, G. H., and Goodson-Espy, T. (2006). Classroom research: Assessment of student understanding of sampling distributions of means and the central limit theorem in post-calculus probability and statistics classes. *Journal of Statistics Education*, *14*(3). http://www.amstat.org/publications/jse/v14n3/lunsford.html

Magel, R. C. (1998). Using cooperative learning in a large introductory statistics class. *Journal of Statistics Education* [Online], *6*(3). http://www.amstat.org/publications/jse/v6n3/magel.html

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

Meier, J., & Rishel, T. (1998). Writing in the Teaching and Learning of Mathematics, Published by MAA [Online: http://www.math.virginia.edu/~der/public/Writing in the Teaching and Learning of Math.pdf]

Meletiou, M., & Lee, C. (2002). Students understanding of histogram: A stumbling stone to the development of intuitions about variation. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a statistically literate society*, (Vol. 2, pp. 1-4). Cape Town, South Africa.

Meletiou-Mavrotheris, M., Lee, C., and Fouladi, R. T. (2007). Introductory statistics, college student attitudes and knowledge – qualitative analysis of the impact of technology-based instruction. *International Journal of Mathematical Education in Science and Technology, 38*(1), 65-83.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis.* Thousand Oaks, CA: SAGE Publications, Inc.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(1), 123-165.

Moore, S. D. (2007). *The basic practice of statistics.* New York: W. H. Freeman and Company.

Murphy, S. A. (1989). Multiple triangulation: Applications in a program of nursing research. *Nursing Research. 38*(5), 294-297.

Mvududu, N. (2005). Constructivism in the statistics classroom: From theory to practice. Teaching Statistics, 2(2), 49-54.

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). *Knowing what students know.* Washington, D.C.: National Academy Press.

Noffke, S., & Stevenson, R. (1995). *Educational Action Research*, NY: Teachers College Press.

Onwuegbuzie A. J., & Leech, N. L. (2003). Assessment in statistics in courses: More than a tool for evaluation. *Assessment & Evaluation in Higher Education. 23*(2), 115-127.

Onwuegbuzie, A. J., & Leech, N. L. (2004). Enhancing the interpretation of significant findings: The role of mixed methods research. Paper presented at the Annual Meeting of the Eastern Educational Research Association, Clearwater, FL.

Palmer, P. J. (2007). *The courage to teach: Exploring the inner landscape of a teacher's life.* San Francisco, CA: Jossey-Bass A Wiley Imprint.

Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal, 5*(2), 27-45.

Pfannkuch, M., Budgett, S., Parsonage, R., and Horring, J. (2004). Comparison of data plots: Building a pedagogical framework. Paper presented at the Tenth International Congress on Mathematics Education (ICME-10), Copenhagen, Denmark. 4-11 July, 2004.

Posner, G. J., Strike, K. A., Hewson, P. W., and Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211-227.

Prichard, G. R. (1995). The NCTM standards and community colleges: Opportunities and challenges. *Community College Review, 23*(1), 23-32.

Priselac, N. J. (1995). Distance learning and introductory statistics. *American Statistical Association 1995 Proceedings of the Section on Statistical Education* (pp. 262-267). Alexandria, VA: American Statistical Association.

Rasmussen, C., Zandieh, M., King, K., and Teppo, A. (2005). Advancing mathematical activity: A practice-oriented view of advanced mathematical thinking. *Mathematical Thinking, 7*(1), 51-73.

Reading, C., & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: From a variation perspective. *Statistics Education Research Journal, 5*(2), 46-68.

Reading, C., & Reid, J. (2007). A critical step in students' reasoning about distribution: Moving from understanding to using for inference. International Statistical Institute, 56th Session.

Reyes, M. L. (2002). Computational and conceptual understanding of the connections among standard deviations, z-scores, and normal distributions. *International Journal of Mathematical Education in Science and Technology, 37*(2), 145-164.

Rittle-Johnson, B., Siegler, R., and Alibali, M. W. (2001) Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*(2), 346-362.

Roseth, C. J., Garfield, J. B., and Ben-Zvi, D. (2008). Collaboration in learning and teaching statistics. *Journal of Statistics Education* [Online], *16*(1). http://www.amstat.org/publications/jse/v16n1/roseth.html

Rossman, A. J. (1996). *Workshop statistics: Discovery with data.* New York: Springer.

Rossman, A. J., Chance, B. L., and Von Oehsen, J. B. (2008). *Workshop Statistics - Discovery with Data and the Graphing Calculator.* Key College Publishing.

Sanches, E., & Mercado, M. (2006). Exploring students' notions of variability in chance settings. In J. Novotna, H. Moraova, M. Kratka & N. Stehlikova (Eds.), *Proceedings of the 30th conference for the International Group for the Psychology of Mathematics Education,* (Vol. 5, pp. 25-32). Prague, Czech Republic.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action.* New York: Basic Books.

Shaughnessy, J. M., & Ciancetta, M. (2002). Students' understanding of variability in a probability environment. In B. Phillips (Ed.), *Proceedings of the Sixth*

International Conference on Teaching Statistics: Developing a statistically literate society, (Vol. 1, pp. 1-6). Cape Town, South Africa.

Shaughnessy, J. M., Watson, J., Moritz, J., and Reading, C. (1999, April). School mathematics students' acknowledgment of statistical variation. Paper presented at the Presession Research Symposium-There's more to life than centers, *77th Annual National Council of Teachers of Mathematics Conference,* San Francisco, California.

Slauson, L. (2008). Students' conceptual understanding of variability. Unpublished doctoral dissertation. The Ohio State University.

Smith, G. (1998). Learning statistics by doing statistics. *Journal of Statistics Education* [Online], *6*(3). http://www.amstat.org/publications/jse

Steinhorst, R. K., & Keeler, C. M. (1995). Developing material for introductory statistics courses from conceptual, active learning viewpoint. *Journal of Statistics Education* [Online], *3*(3). http://www.amstat.org/publications/jse

Turegun, M. (2009). Designing a curriculum matrix for introductory statistics: A more holistic approach. Paper presented at the Third International Conference on Design Principles and Practices, Berlin, Germany. February, 2009.

Turegun, M., & Reeder, S. (2008). An exploration of students' understanding of spread. Paper presented at the Inaugural World Universities Forum, Davos, Switzerland. January, 2008.

von Glasersfeld, E. (1995). *Radical constructivism a way of knowing and learning*. New York: RoutledgeFalmer.

Watson, J. M., Kelly, B. A., Callingham, R. A., and Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology, 34*(1), 1-29.

Wheatley, G. (1991). Constructivist perspectives on science and mathematics learning. *Science Education, 75*(1), 9-21.

Whitehead, A. N. (1929). The Aims of Education. New York: The Free Press.

Werner, J. (1997). Are the students in elementary statistics classes using statistical reasoning? Paper presented at the Annual Pennsylvania State System of Higher Education Conference, Mansfield, Pennsylvania. [Online: http://www.passhema.org/proceedings/1997/1997werner.pdf]

Wilson, S. (1977). The use of ethnographic techniques in educational research. *Review of Educational Research, 47*(1), 245-265.

Wu, H. (1999). Basic skills versus conceptual understanding: A bogus dichotomy in mathematics education. *American Educator, 23*(3), 1-7. [Online: htpp://aft.org.pdfs/americaneducator/fall1999/wu.pdf]

Zieffler, A., & Garfield, J. (2007). Studying the role of simulation in developing students' statistical reasoning. International Statistical Institute, 56th Session.

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., and Chang, B. (2008). What does research suggests about the teaching and learning of Introductory Statistics at the college level? A review of the literature. *Journal of Statistics Education, 16*(2). http://www.amstat.org/publications/jse/v16n2/zieffler.html

APPENDICES

APPENDIX A

Course Information

COURSE INFORMATION, Spring 2010

Course No: MATH 2013 Section: 002	Professor: Turegun
Course: Introduction to Statistics	E-mail: mturegun@occc.edu
Time & Room: MW 1:00 – 2:20 PM & 1B5 SEM Center	Office: 2BC7 SEM Center

Office Hours: MTWTh 9:00 – 10:50 AM
TTh 3:00 – 3:50 PM

Course Expectations

This course is activity based and problem centered in nature. Since learning involves a degree of cognitive dissonance, you should be prepared to be taken out of your comfort zone at times and struggle with ideas. Hence, you should be prepared to work and to think. Throughout the course you will be expected to think deeply, creatively, critically, and support your conclusions with varied forms of evidence, and support our classroom community of learners.

Consistently demonstrated patterns of commitment, preparation, and quality (of thought, work, and participation) are required to receive an exceptional grade in this class. Emphasis is also placed on your contributions to the development of a healthy and active classroom community. Thus, you should contribute your energy to class activities, demonstrate respect for all participants and help ensure that everyone is permitted and encouraged to share equally in class opportunities and responsibilities. Every member of our class is considered to be both a learner and a teacher. You will need to assume responsibility for both roles. A major responsibility, both as a learner and a teacher, is to attend class regularly, arrive on time, stay for the duration of the class, and prepare by studying all materials and completing all assignments carefully and on time. It is ***extremely*** unlikely that you will earn an “A” for this course if you miss more than five classes for any reason, or a “B” if you miss more than ten classes. Please note that even perfect attendance does not guarantee an “A” or “B.”

Course Assessment

Your success in this course will be based on your understanding of how to think and learn creatively and critically. It is *your* responsibility to provide evidence of this understanding through the following criteria:

Preparation & Class Participation	10%
Weekly Reflective Journals	15%
Portfolio	25%
Problem Solution Presentations	25%
Closed Book Journals (CBJ)	15%
Midterm Conference	10%

Preparation & Quality Class Participation: Class participation is a qualitative judgment of student preparation, intellectual curiosity and articulation of mathematical/statistical ideas and concepts as well as a quantitative assessment of class attendance. Throughout the course you will be expected to think deeply, creatively, support your conclusions with varied forms of evidence, and support our classroom community. Your contribution to the development of a healthy and active classroom community is essential. Every member of our class is considered both a learner and a teacher, and you will need to assume the responsibility for both roles. Thus, you should engage in class activities and conversations, contribute to collaborative learning activities in a meaningful, professional way, demonstrate respect for all participants, and help ensure that everyone is permitted and encouraged to share equally in class opportunities and responsibilities.

Weekly Reflective Journals: The purpose of the weekly journals is to help you reflect at the end of each class period on your experience in the class. After each class you are to record your thinking related to the following items:

- significant mathematical/statistical ideas you have developed during the class
- the environment in which you are engaged in mathematical/statistical thinking
- any other thoughts or questions you have about mathematics/statistics learning
- any other questions/topics which I might post on the ANGEL course site

The weekly journals will be submitted to the instructor at the end of each week during the course for evaluation.

Portfolio: During this course you will be expected to choose problems for which you have developed solutions, write these solutions up formally, and include them in your portfolio, as evidence of your mathematical/statistical activity/thinking. The number of problems to be included can be negotiated with the instructor, taking into consideration such things as the nature of problem chosen for inclusion and the depth of the analysis involved. On average, there will be a couple of portfolio items for each module of class (5-6 total problems). You are encouraged to begin working on your portfolio very early in the semester and receive feedback from me. The completed portfolio will be submitted on the last week of class.

Problem Solution Presentations: You and a partner will choose and present 5 problem solutions. Include in your presentations the thought process you have used to approach the solving of each problem. You will turn in some form of your presentation (CD, power point slide show, etc.) the night before each presentation is due.

Closed Book Journal (CBJ): You will be expected to write concise remarks to describe and/or summarize your understanding of a main concept, procedure or topic of the course during the first 10-15 minutes of the class. There will be about five CBJs during the semester.

Midterm Conference: Each of you will need to schedule a 15-20 minute meeting with me during the middle of the semester. During this meeting the focus of the conversation will be on your conceptual understanding and abilities related to the course topics and goals. The meeting will be audio recorded. I will indicate strength and weaknesses and provide a non-binding estimate of where I believe you currently stand during your midterm conference. At this time you might also discuss your progress and clarify any questions you might have about the class.

Graphing Calculator Guide

A statistics web site has been set up to provide you with necessary and important information regarding the usage of your TI-83/83 Plus, TI-84/84 Plus calculators. The site contains the required TI-83/83 Plus, TI-84/84 Plus calculator solutions to many of the examples taken out of the 2nd edition of your textbook. You will find helpful demonstrations of the statistical functions and features of the TI-83/83 Plus, TI-84/84 Plus calculators, which are required for the course. The statistics web site address is: **http://www.occc.edu/statistics**

Academic Misconduct and Dishonesty

Cooperative effort is allowed on assignments and/or collaborative activities, even encouraged if one student acts in a tutorial capacity. However, copying of assignments will result in no credit for the first offense. Repeated instances will result in a harsher penalty which may result in a score of zero for all parties involved and/or may result in failure for the semester. Any incident of academic misconduct or dishonesty will be a basis for referral to the Dean with a request for withdrawal from class. Please refer to the college catalog for details.

TENTATIVE COURSE TIME LINE (subject to change as the semester progresses)

Module One: Collecting Data and Drawing Conclusions, and Summarizing Data
Units 1 and 2 (Topics 1 – 10)
Jan. 19 - Feb. 19 (5 weeks)

Module Two: Randomness in Data and Inference from Data: Principles
Units 3 and 4 (Topics 11 – 20)
Feb. 22 - Apr. 2 (5 weeks)

Module Three: Inference from Data: Comparisons and Categorical Data, and Relationships in Data
Units 5, 6, and 7 (Topics 21 – 29)
Apr. 5 - May 7 (5 weeks)

APPENDIX B

Interview Protocol

I am going to ask you a few questions. The point of the interview is for me to try to understand your thought processes as you work through a statistical question.

Please try to explain as much of your reasoning out loud as you can. Do not worry about whether you are right or wrong. You are welcome to use whatever tools you need to work these out (paper, calculator, etc), but do your best to verbalize as much of your reasoning as possible out loud. I may ask you follow up questions as you answer the question.

1) What is meant by the range, interquartile range (IQR), and standard deviation in statistics?

2) What is meant by the phrase measures of spread in statistics?

3) Why or how might you use the standard deviation?

4) Why or how might you use the IQR?

5) Why would you use the standard deviation over the IQR or the range? Can you think of when you might use one over the other one?

6) Suppose your best friend was studying for a stat test and has to know everything there is to know about the standard deviation. What would you try to tell your best friend?

APPENDIX C

Open-ended Response CBJ Item

Name:____

What are the measures of spread? How are they similar? How are they different? Discuss their similarities and differences.