Preservice Elementary Teachers' Conceptual and Procedural Knowledge of Mean, Median, and Mode

Randall E. Groth and Jennifer A. Bergner

Salisbury University

Corresponding Author:

Email: regroth@salisbury.edu

Preservice Elementary Teachers' Conceptual and Procedural Knowledge of Mean, Median, and Mode

Abstract

The paper describes aspects of the statistical content knowledge of 46 preservice elementary school teachers. The preservice teachers responded to a written item designed to assess their knowledge of mean, median, and mode. The data produced in response to the written item were examined in light of the Structure of the Observed Learning Outcome (SOLO) Taxonomy (Biggs & Collis, 1982, 1991) and Ma's (1999) conception of Profound Understanding of Fundamental Mathematics (PUFM). Four levels of thinking in regard to comparing and contrasting mean, median, and mode are described. Several different categories of written definitions for each measure of central tendency are also described. Connections to previous statistical thinking literature are discussed, implications for teacher education are given, and directions for further research are suggested.

Preservice Elementary Teachers' Conceptual and Procedural Knowledge of Mean, Median, and Mode

Shulman (1987) used the term "content knowledge" to describe knowledge of the structure and methods of discourse within any given discipline. He stated that the many aspects of content knowledge "are properly understood as a central feature of the knowledge base of teaching" (p. 9). Research illustrates that attaining content knowledge is an important part of a mathematics teacher's development (Putnam, Heaton, Prawat, & Remillard, 1992). At the same time, research illustrates that teachers sometimes lack the content knowledge they need in order to teach elementary school mathematics (Ma, 1999). Ball, Lubienski, and Mewborn (2001) argued that

> Without such (mathematical) knowledge, teachers lack resources necessary for solving central problems in their work – for instance, using curriculum materials judiciously, choosing and using representations and tools, skillfully interpreting and responding to their students' work, and designing useful homework assignments (p. 433).

The statistical measures of mean, median, and mode comprise a fundamental portion of the content knowledge elementary teachers need to possess. Curriculum documents around the world recommend that students develop some understanding of mean, median, and mode before entering secondary school (Australian Education Council, 1994; School Curriculum and Assessment Authority & Curriculum and Assessment Authority for Wales, 1996; National Council of Teachers of Mathematics, 2000). The National Council of Teachers of Mathematics (2000), for example, recommended that students should develop the abilities to "select and use appropriate statistical methods to analyze data" (p. 178) throughout Grades PreK-8. This included attaining proficiency in determining and applying the mean, median, and mode by the end of Grade 8.

In order to help teachers develop the content knowledge they need to posses about mean, median, and mode, it is important to have an understanding of their thinking about these measures. Just as successful instruction of children uses the children's thinking as a starting point, successful instruction of teachers uses the teachers' thinking as a starting point (Mewborn, 2003). The present study aims to describe the anatomy of preservice teachers' content knowledge in regard to mean, median, and mode in order to help inform teacher education efforts and further research into preservice teachers' thinking.

Types of Mathematics Content Knowledge

Mathematics content knowledge can be thought of as consisting of two complementary aspects: conceptual and procedural. Hiebert and Lefevre (1986) offered definitions for both. This section discusses their definitions and how they apply to knowledge of mean, median, and mode. The definitions are elaborated upon here because the present study was concerned with investigating aspects of both conceptual and procedural knowledge.

Conceptual Knowledge

In discussing conceptual knowledge of mathematics, Hiebert and Lefevre (1986) stated:

> Conceptual knowledge is characterized most clearly as knowledge that is rich in relationships. It can be thought of as a connected web of knowledge, a network in which the linking relationships are as prominent as the discrete pieces of information. Relationships pervade the individual facts and propositions so that all pieces of information are linked to some network (Hiebert & Lefevre, 1986, p. 3).

From this description, the essence of conceptual knowledge is that it involves forming cognitive connections between bits of information that might otherwise be perceived as unrelated.

Conceptual knowledge relating to mean, median, and mode can be formed in many ways. For instance, one might recognize that all three have the common characteristic that they are sometimes used to find central and/or typical values of data sets. One might also recognize that in some instances one measure is more suitable than another for determining central and/or typical values. These examples do not comprise an exhaustive list of conceptual knowledge of mean, median, and mode. Rather, they illustrate that conceptual knowledge grows as cognitive networks consisting of connections between discrete bits of information about the measures are formed. As conceptual understanding grows, such networks are constantly realigned and reconfigured (Hiebert & Carpenter, 1992).

Procedural Knowledge

Hiebert and Lefevre (1986) described procedural knowledge of mathematics in the following manner:

> Procedural knowledge…is made up of two distinct parts. One part is composed of the formal language, or symbol representation system, of mathematics. The other part consists of the algorithms, or rules, for completing mathematical tasks (Hiebert & Lefevre, 1986, p. 5).

The first type of procedural knowledge "includes a familiarity with the symbols used to represent mathematical ideas and an awareness of the syntactic rules for writing symbols in an acceptable form" (p. 6). The second type "consists of rules, algorithms, or procedures used to solve mathematical tasks" (p. 6). Using an example from algebra to clarify the distinction between the two types of procedural knowledge, recognition that the string $6 + x = 5$ is syntactically acceptable, and that $6 + \ = x5$ is not, falls under the first type of procedural knowledge. Knowing an algorithm for determining the value of x in $6 + x = 5$ falls under the second type of procedural knowledge.

Both types of procedural knowledge discussed by Hiebert and Lefevre (1986) can be mapped to the case of the mean. A student who recognizes that it is conventional to use the symbol $\bar{X}$ to denote the mean of a data set is demonstrating the first type of procedural knowledge, which is concerned with syntax. Stating definitions correctly can also be considered a syntactic issue (Weber & Alcock, 2004). In the case of the mean, the acceptable syntax varies, as some authors define the mean to be *the* average of a data set, while others define it as just one type of average (Upton & Cook, 2002). A student who can execute the "add-and-divide" algorithm to calculate the mean demonstrates the second type of procedural knowledge, which encompasses knowledge of rules, algorithms, and procedures.

It should also be noted that syntactic and algorithmic knowledge are not always displayed in isolation. For instance, in the discussion above, it was noted that stating a correct definition for the mean can be considered a matter of syntax. It is not uncommon, however, for authors to include the "add-and-divide" algorithm for the mean in a stated definition. For example, Moore (1997) defined the mean in the following manner: "The mean of a set of observations is their arithmetic average. It is the sum of the observations divided by the number of observations" (p. 237). It appears that both syntactic and algorithmic knowledge are on display in such a definition. Simply providing an accepted definition for mean reflects an understanding of syntax. Including the "add-and-divide" procedure reflects algorithmic knowledge. Hence, when one states a definition for the mean, there is the potential to display both syntactic and algorithmic knowledge. The same is true for the median and mode. One's definition for each measure might include a statement that they are types of averages along with the algorithms used for determining each.

Research Questions

This report focuses on two research questions:

1) What levels of thinking do preservice teachers exhibit in comparing and contrasting the statistical concepts of mean, median, and mode?
2) In what ways do preservice teachers define the mean, median, and mode?

Research question one focuses on preservice teachers’ knowledge of the relationships among the three statistical measures, and hence *primarily* addresses conceptual knowledge, since the concern is with investigating perceived connections among the measures. Research question two focuses on their definitions for each measure, and therefore *primarily* addresses syntactic and algorithmic procedural knowledge. It is important to note, however, that each of the research questions do not deal exclusively with one type of knowledge. For example, in regard to research question 1, an individual could compare and contrast mean, median, and mode by simply reciting algorithms for calculating each one, and hence display only procedural knowledge. In regard to research question 2, an individual’s personal definitions for mean, median, and mode may reveal aspects of conceptual knowledge if connections to other bits of knowledge are inherent in the definitions. Therefore, each research question has the potential to deal with both procedural and conceptual knowledge.

Theoretical Perspective

Two complementary theoretical constructs were incorporated in the present study. The first was the Biggs and Collis (1982, 1991) Structure of the Observed Learning Outcome (SOLO) taxonomy. The second was Ma’s (1999) description of Profound Understanding of Fundamental

Mathematics (PUFM). PUFM provided a vision of the ideal structure of elementary teachers' conceptual and procedural content knowledge, while SOLO provided a framework for beginning to describe levels of thinking below that ideal structure.

A Perspective on Levels of Thinking: The SOLO Taxonomy

Biggs and Collis (1991) theorized that levels of thinking are situated within several different modes of representation. The concrete symbolic and formal modes were pertinent to the present study. Chick (1998) described the difference between the two modes in the following manner:

> While the concrete-symbolic mode is characterized by the use of symbols which represent concrete objects, the formal modes take abstraction a degree further…knowledge is more theoretical and general, removed from the specific and concrete of the earlier modes (p. 5).

School learning up to the university level usually requires functioning in the concrete symbolic mode (Biggs & Collis, 1982) while university study usually requires some functioning in the formal mode (Collis & Biggs, 1983). Individuals typically acquire the formal mode of thinking at approximately 14 years of age, although some do so later or earlier, and some adults do not develop formal mode thinking at all (Biggs & Collis, 1991). Given that the participants for the present study were university students, it was expected that some would exhibit formal mode thinking, while others might exhibit concrete symbolic thinking.

The SOLO taxonomy describes different levels of thinking situated within the modes of representation. Biggs (1999) described five levels in the taxonomy. The lowest SOLO level is called prestructural. At this level, students' responses to tasks show that they have little understanding of the task demands. Nothing of relevance is included in the response. At the second level, unistructural, students show that they have some understanding of the task. While unistructural responses are far from complete, they do include one aspect of relevance to the given task. Responses falling within the third level, multistructural, contain more than one aspect of relevance. However, multistructural responses amount to little more than "fact-telling," since there is no apparent theme unifying the relevant aspects discussed. A unifying theme is apparent at the fourth level of the taxonomy, called the relational level. The fifth level of the taxonomy is called extended abstract. Extended abstract responses include all of the characteristics of the relational level, and they also include aspects of relevance that go beyond the immediate task requirements.

Previous research has built upon and refined the SOLO Taxonomy in order to make it a more powerful tool for describing levels of thinking. Campbell, Watson, and Collis (1992) discussed how the unistructural, multistructural, and relational levels cycle within modes. Several studies have pointed out that the entire SOLO Taxonomy can be discussed entirely in terms of unistructural-multistructural-relational (UMR) cycles (Watson, Collis, & Moritz, 1997; Pegg & Davey, 1998; Watson & Moritz, 1999). The middle three levels in SOLO comprise one complete UMR cycle, while the lowest level (prestructural) comprises the relational level in a less sophisticated UMR cycle, and the highest level (extended abstract) comprises the unistructural level in a more sophisticated UMR cycle. Any given mode of representation might contain a number of these UMR cycles.

A Perspective on Teachers' Content Knowledge: PUFM

Ma (1999) described Profound Understanding of Fundamental Mathematics (PUFM) as "an understanding of the terrain of fundamental mathematics that is deep, broad, and thorough" (p. 120). Teachers who have PUFM know the field of mathematics in much the same manner as a taxi driver familiar with a town knows its roads. Teachers with PUFM do not simply have knowledge of a number of disconnected procedures and definitions in elementary mathematics. Rather, they have a connected knowledge of elementary mathematics that encompasses both concepts and procedures.

In terms of modes of representation, PUFM links most strongly to what Chick (1998) described as the formal-1 mode. Upon acquiring the formal-1 mode, individuals are able to learn and operate at a higher level of abstraction than the concrete symbolic. This is distinct from the formal-2 mode, which involves creating original theories and extending the bounds of professional knowledge. Elementary teachers with PUFM have knowledge that lies beyond the concrete symbolic. Concrete-symbolic thinking is generally the required mode of operation for elementary students (Biggs & Collis, 1982), but university courses designed to foster PUFM aim beyond that mode. Such courses are not just designed to help prospective teachers perform the concrete mathematics their students are expected to learn, but also to help them understand how the various concepts of elementary mathematics fit together into a network (Conference Board of the Mathematical Sciences, 2001). While teachers with PUFM have acquired these abstract concept networks, they do not necessarily create original theories about concepts relating to elementary mathematics. Hence, while PUFM is beyond the concrete symbolic mode, its essential characteristics link more closely with the formal-1 mode than with the formal-2 mode.

Illustrating the connection between SOLO and PUFM

In order to solidify the relationship described above between the SOLO taxonomy and PUFM, we illustrate in this section how teachers' response patterns to an interview task used by Ma (1999) can be described using the SOLO taxonomy. A slight digression from measures of center will be necessary, since the Ma (1999) study did not include any questions about the topic. The task discussed below required teachers to construct a word problem to model $1\frac{3}{4} \div \frac{1}{2}$. The most sophisticated response pattern, which was given by teachers who had been characterized as having attained PUFM, resembles the thinking of an individual who has acquired the formal-1 mode. The response patterns before it do not seem to advance past concrete-symbolic mode thinking.

The first pattern of response to the division task was that in which teachers were not able to construct any kind of word problem for the given situation. Six of the teachers interviewed said they were not able to create a problem. This type of response mirrors the prestructural level in SOLO, since the teachers did not even engage the problem. No relevant progress was made toward answering the task.

In a second response pattern, teachers did generate word problems. However, the word problems each reflected a misunderstanding of the structure of the task. Some wrote word problems that actually modeled $1\frac{3}{4} \div 2$. Others produced word problems that modeled $1\frac{3}{4} \times \frac{1}{2}$. Still others produced word problems that showed confusion stemming from a combination of the previous two ideas. In each case, teachers did not recognize the problematic nature of their representations. However, the responses each contained one aspect relevant to solving the problem: each performed an operation on the quantity $1\frac{3}{4}$. This marked an advance from prestructural responses, which did not contain any relevant aspect. Since one relevant aspect was included in these responses, they can be characterized as unistructural. In this response pattern, some understanding of symbol systems was exhibited, so it appears to lie in the concrete symbolic mode. The next two patterns of response were also characteristic of that mode.

The next pattern of response can be described as multistructural. Like the unistructural responses, the multistructural responses recognized the need to construct a problem that would involve an operation on the quantity $1\frac{3}{4}$. Unlike the unistructural responses, multistructural ones also acknowledged that division by $\frac{1}{2}$ was different from division by 2. Hence, teachers giving multistructural responses expressed recognition of more than one relevant piece of knowledge for solving the problem. However, at the same time, they were not able to bring closure to the problem by integrating those relevant pieces of knowledge.

Several of the Chinese teachers in the study were able to produce one correct word problem. One teacher, for example, offered,

> Cut an apple into four pieces evenly. Get three pieces and put them together with a whole apple. Given that $\frac{1}{2}$ apple will be a serving, how many servings can we get from the $1\frac{3}{4}$ apples? (Ma, 1999, p. 73).

The ability to construct a word problem marked an advance from the previous level of thinking described. At the previous level, multistructural, responses just recognized that an operation on the quantity $1\frac{3}{4}$ needed to take place and that division by $\frac{1}{2}$ or division by 2 was not adequate. The production of a correct word problem can be described as relational, since it implicitly incorporated recognition of the pieces of knowledge grasped at the multistructural level, but also went beyond the multistructural responses to use that knowledge in the production of reasonable closure to the task.

One last pattern of thinking can be identified in teachers' responses to the word problem construction task. Some of the Chinese teachers produced multiple correct word problems for the given operation, and they discussed the difference between the abstract concepts of partitive and measurement division. Therefore, instead of simply supplying either a partitive problem or a measurement problem, they produced both and discussed the differences between them. This type of response is reflective of PUFM and formal-1 thinking, since it shows knowledge of the network of concepts that elementary school students study pertaining to division. It also has the characteristics of an extended abstract level of thinking because it goes beyond the immediate task requirements by including additional relevant aspects not explicitly requested in the problem.

Summary

The SOLO taxonomy and PUFM coupled together provide a framework for describing teachers' content knowledge. PUFM serves to describe the ideal structure of elementary teachers' content knowledge. The levels of the SOLO taxonomy provide a model for describing different levels within teachers' content knowledge. Since the SOLO taxonomy has been gainfully employed in other studies of statistical content knowledge pertaining to measures of center (Jones et al., 2000; Mooney, 2002; Groth, 2002), and PUFM provides impetus to consider important procedural knowledge issues that may fall outside the domain of a SOLO analysis, the two theoretical lenses coupled together provided perspective for the current study.

Previous Research Concerning Measures of Central Tendency

There are several cognitive challenges that individuals encounter in developing their thinking about measures of central tendency. Three types of challenges documented by previous research are considered in this section: (1) Learning and executing the procedures for calculating measures of center; (2) Moving beyond a purely procedural conception of measures of center to understand what information they represent about any given set of data; (3) Having multiple measures of center at one's disposal and recognizing when one might be more appropriate than another for describing a given data set. Since these challenges appeared in previous research, it was postulated, a priori, that they might arise in the present study.

Learning and executing procedures for measures of center

Several studies document students' difficulties with learning and executing procedures for finding the arithmetic mean. Pollatsek, Lima, and Well (1981) found that many undergraduates were unable to calculate weighted means. Callingham (1997) found that the same was true for preservice and in-service teachers. Mevarech's (1983) study of undergraduates showed that they experienced difficulty in identifying situations where means had been calculated incorrectly. Several of the Grade 6 students studied by Cai (2000) were not able to execute the algorithm for the arithmetic mean in reverse.

Difficulties in determining the medians of data sets have also been documented by research. Friel and Bright (1998) noted that elementary school teachers have difficulty in determining the medians of data sets presented graphically. Zawojewski and Shaughnessy (2000) noted that only about one-third of Grade 12 students in the U.S. were able to determine the median when given a set of unordered data. Becoming computationally fluent in calculating medians and arithmetic means is a non-trivial matter for many students.

Unlike the median and the arithmetic mean, students' procedural knowledge of the mode has not drawn a great deal of attention in the research literature. This might be due to a perception that the procedure for calculating the mode is less complex than the procedures for the mean and median. It might also be due to the mode occupying a less prominent position in school statistics curricula.

Moving beyond procedural understanding

Computational fluency is just one aspect of the knowledge base concerning measures of center. One must also realize that the mean, median, and mode are not merely mathematical objects to be computed, but that they are in some way representative of the data sets from which they arise. Mokros and Russell (1995) provided several examples of children who had mastered the algorithm for the mean, but who did not realize that it was representative of the center of a data set. Watson and Moritz (2000) also observed that several children exposed to the mean algorithm were not able to apply it in a complex problem-solving situation. Both studies cautioned teachers not to take a purely algorithmic approach to measures of central tendency.

There are several levels of thinking below the level at which students view measures of central tendency as useful for describing typical values in data sets. Jones et al. (2000) documented three levels of thinking below the one in which elementary school students used common measures such as the median or the mean to describe "typicality" of data. At the lowest of the three levels, students were not able to describe data in terms of representativeness. At the second level, some attempt was made to come to terms with representativeness, but descriptions given were hesitant and incomplete. Students thinking at the third level reasoned using valid measures of center that were close to conventional measures. Mooney (2002) described an almost parallel set of levels of thinking for middle school students.

Konold and Pollatsek (2002) illustrated the importance of conceptual understanding of averages beyond the idea that they can represent typical values in fixed distributions. They recommended that students should come "to see statistics as the study of noisy processes – processes that have a signature, or signal, which we can detect if we look at sufficient output" (p. 260). Their view entails that students study applications of averages such as approximating the correct weight from a set of repeated measurements and approximating average measures for populations in flux. They argued that applications such as these are at the core of the practice of statistics, but not currently at the core of statistics curricula.

Choosing among multiple measures of center

Another challenge in developing understanding of measures of center is choosing the best measure of center from several alternatives. National Assessment of Educational Progress (NAEP) data confirm that students frequently make poor choices in selecting measures of center to describe data sets (Zawojewski & Shaugnessy, 2000). When asked to choose between the mean and the median for describing sets of data, students predominantly chose the mean, even when outliers within the sets of data made the mean less indicative of center than the median. On a test item containing a data set structured so that the median would be a better indicator of center than the mean, only four percent of Grade 12 students correctly chose the median to summarize the data set and satisfactorily explained why it was the appropriate measure of center to use. Callingham (1997) administered a similarly structured item to a group of preservice and inservice teachers and found that most of them also calculated the mean when the median would have been a better measure of center. Groth (2002) showed that choosing an appropriate measure of center also presented challenges for students enrolled in a high school Advanced Placement Statistics course.

Summary

The research in this section documents important developmental milestones in understanding and using measures of central tendency. A low level of understanding often involves little more than the knowledge of the procedures and/or definitions for different measures of center. A more sophisticated level of understanding seems to be that at which students come to realize that measures of central tendency tell what is “typical” or “average” about a set of data. A still more complex level of thinking is that at which students flexibly draw upon knowledge of previously learned measures of center and use only those that are most suitable for the set of data they are analyzing. In conducting the present study, special attention was paid to trying to detect whether or not the participants had obtained these important components of thinking about measures of central tendency.

Several aspects of the present study are unique to the research literature. Unlike previous studies, the focus was upon studying the interrelationships among students’ understanding of mean, median, and mode. This was done by explicitly attending to both conceptual and procedural knowledge of the topics. As noted above, the mode has not been given as much research attention as the mean and median, so the present study helps to fill that gap in the literature. The study also aimed to help teacher educators design effective instruction by providing a description of preservice teachers’ content knowledge in the area of measures of central tendency. The description given may prove to be helpful to other educators as well, since some of the patterns of thinking described most likely can also be found among student populations other than preservice teachers.

Methodology

Study design

Since the objectives of the present study concerned describing preservice teachers’ levels of thinking, a qualitative study design was chosen. Qualitative designs are helpful for exploring complex phenomena and describing them (Miles & Huberman, 1994). Within the qualitative design, written responses to open-ended test items were the data sources. The open-ended items allowed the researchers to categorize responses into various levels of thinking. Analysis of the data was carried out by the two authors, with the SOLO Taxonomy providing the theoretical basis for coding responses according to levels of thinking.

Participants

Data were gathered from preservice elementary and middle school teachers in three different sections of an elementary mathematics teaching methods course taught by the first author. There were 48 preservice teachers in the three sections. All but two signed and returned the consent form necessary to have their data included in the study. Hence, this study reports on data from a total of 46 preservice teachers. Seven of the preservice teachers were male, and 39 were female. Each participant had taken two elementary mathematics content courses and one introductory statistics course at the college level before taking the methods course that provided the setting for the present study. Some had taken the pre-requisite courses at the same university where the methods course took place, while others had transferred the courses from other institutions.

Procedure

At the conclusion of the first three weeks of the methods course, a six-question written content knowledge assessment was administered to all students enrolled in the course. The present study focuses on responses to the first item on the assessment, which pertained to measures of center. It read: “How are the statistical concepts of mean, median, and mode different? How are they similar?” Students were given 23 blank lines between the statement of the first item and the second item. The first item and the rationale behind its construction are described in the next section. The remainder of the items on the assessment dealt with calculating probabilities, reflecting on a lesson they had taught to a group of peers, designing a statistical study, and interpreting the results of a study. One hour of class time was allowed for the assessment, and students were allowed to have extra time if they needed it. None of the students used more than the hour of class time that was provided.

Instrument

The assessment question pertaining to measures of center, as discussed above, asked the preservice teachers to compare and contrast the statistical concepts of mean, median, and mode. The question was designed to allow the preservice teachers to exhibit PUFM, since it provided them an opportunity to make connections between the concepts given. In addition to serving as an instrument to gather data for this report, the assessment task was designed to help inform future instruction in the methods course by giving a picture of the preservice teachers' abilities to make connections among the three measures.

Data Analysis

Two rounds of analysis were carried out on the data collected. The first round focused on answering research question 1: identifying levels of thinking exhibited in comparing and contrasting measures of center. The second round focused on answering research question 2: investigating the structure of the written definitions given for each of the measures of mean, median, and mode.

Data analysis to answer research question 1. The first author grouped the responses to the assessment item asking for similarities and differences between mean, median, and mode into categories. The grouping was done with the SOLO Taxonomy (Biggs & Collis, 1982, 1991) in mind. Hence, the responses were grouped according to how many statistically relevant aspects were incorporated in the response and how well they were integrated. Statistically relevant aspects considered included those that have arisen in the literature discussed earlier: computing measures of center, having conceptual understanding of the measures, and being able to discuss when it is more advantageous to use one measure than another. Each level of thinking identified was given a brief qualitative description.

The data analysis process for research question 1 then followed some of the aspects of a check-coding procedure described by Miles and Huberman (1994). After the first author had grouped responses into levels of thinking, the second author independently analyzed all of the responses. The second author read the responses from each of the preservice teachers, and attempted to place each one within one of the descriptions of levels of thinking formulated by the first author. In cases where the second author placed a response in a category different from where the first author had placed it, the appropriate placement of the response was discussed. Cases where category descriptions needed to be revised were also discussed until a mutually agreed upon set of descriptions was produced.

In placing responses in categories, it was agreed that some of the responses were too vague to be placed in a given category with certainty because some students used the phrase “measures of central tendency” in their responses, but it was uncertain what significance they attached to the phrase. A written follow-up item was administered to all 46 participants approximately four weeks after the initial assessment item in order to further explore the nature of the preservice teachers’ conceptions in regard to the phrase. The follow-up item read: “Suppose that you used the phrase ‘measures of central tendency’ while you were teaching, and then a student asks you what the phrase means. How would you respond?” If they did not indicate on the follow up question that a measure of central tendency is a measure of what is average or typical for a data set, their responses to the initial question were categorized at a lower SOLO level than those who did indicate such an understanding of the phrase.

Data analysis to answer research question 2. In answering the initial assessment item, all of the participants, except one, attempted to provide a definition for mean, median, and mode. Therefore, a second layer of analysis was possible in which these definitions were examined. It was considered worthwhile to examine the definitions because they often reveal aspects of procedural knowledge, and such knowledge forms part of the overall "knowledge package" (Ma, 1999) that teachers need to teach elementary mathematics.

Elements of the constant comparative method (Maykut & Morehouse, 1994) were used to help analyze definitions for mean, median, and mode. First, each preservice teacher's definition for the arithmetic mean was analyzed. Definitions having the same structural characteristics were grouped together. Each group was then given a qualitative description. This was repeated for each preservice teacher's definition of median and mode. The end result was a set of qualitative descriptions of different categories of definitions for the mean, median, and mode.

Results

The results of the study are reported in two main sections. The first section discusses findings about levels of thinking exhibited in regard to comparing and contrasting mean, median, and mode. The second discusses findings about the types of definitions preservice teachers gave for each of the measures.

Levels of Thinking in Regard to Comparing and Contrasting Measures of Center

Four distinct levels of thinking were found in regard to preservice teachers' abilities to compare and contrast the concepts of mean, median, and mode. The first three levels mirrored the type of concrete symbolic thinking one might expect from elementary school children learning the concepts. The final level mirrored formal-1 mode thinking, since some knowledge beyond the concrete symbolic level was demonstrated. The qualitative characteristics of each of the levels of thinking are discussed below, along with sample responses from each of the levels of thinking identified. The number of responses categorized at each level of thinking and the characteristics of each level are provided in Table 1.

<INSERT TABLE 1 HERE>

Unistructural level responses. Eight of the preservice teachers exhibited unistructural thinking in comparing and contrasting the mean, median, and mode. Unistructural responses did not contain any strategy other than definition-telling in response to the request to compare and contrast the three measures. The following response was typical of those at the unistructural level:

> Mean, median, and mode are similar in that they all are derived from a set of numbers. The mean is obtained by averaging all the numbers in a set. The median is the number in a sequential set of numbers directly in the center; it may or may not be an actual number in the set. The mode is the number appearing the most often in a set.

In unistructural responses, the task's request to discuss how the measures are different was addressed by listing the different definitions. The request to discuss how they are similar did not proceed past noting that each is the end result of a process or definition applied to a set of data.

Multistructural level responses. Multistructural responses included definition-telling along with a vague notion that the mean, median, and the mode are all tools that can be used to analyze a set of data. However, they did not reflect an understanding that each of the tools is intended to measure what is central or typical to data sets. The following sample response captures the characteristics of the multistructural level:

> The mean, median, and mode are different in many ways. The mean tells us the average of the data set, where the median tells us what the middle number in the set of data is.

> The mode is different from the mean and median because it is the number from the data set that occurs most often. The mean, median, and mode are similar in that they can all be used to summarize a data set. The numbers each have different meanings though.

In responses at this level, it was evident that students had acquired the idea that each of the measures captures some of the characteristics of data sets.

In general, in multistructural responses it was not clear that preservice teachers viewed mean, median, and mode as measures that capture, to varying degrees, what is typical or central in a set of data. Several phrases in multistructural responses, however, hinted that preservice teachers viewed the measures as tools to analyze data. For example, some of the thoughts in multistructural responses included: (1) "They are all similar because they are ways of representing statistical data and portraying the information," (2) "These concepts are all trying to figure out a certain number that is relevant to the data," (3) "They represent different aspects of the data." Hence, while multistructural responses still included the aspect of definition-telling, it was somewhat evident that the preservice teachers giving them viewed mean, median, and mode as mathematical objects for data analysis rather than just mathematical procedures to be executed. The responses still lacked closure, however, in that they did not state what these data analysis tools might tell one about any given set of data.

Relational level responses. Relational level responses differed from multistructural responses in that they included recognition of the fact that the mean, median, and mode all measure the center of the data or what is typical about it in some manner. This element was considered the "unifying theme" (Biggs, 1999) that distinguished relational level responses from multistructural ones. The following is a sampling of the phrases that were included in relational level responses that were not present in multistructural responses: (1) "The three of them find a sort of middle”; (2) “All three of them measure the average of a group of data in some way”; (3) “They are similar because they all show somewhat of a typical thing for that particular set of data”; (4) “Each of the concepts of mean, median, and mode are measurements of the center of a set of data.”

Some preservice teachers noted that mean, median, and mode were all “measures of central tendency.” The first author initially categorized these types of responses as relational level responses. In the process of data analysis, the two authors agreed it was highly possible that the phrase “measures of central tendency” might not be anything more than a memorized phrase from previous instruction. It was agreed that a response should be categorized as multistructural if the preservice teachers who used the phrase did not associate it with notions such as the middle, average, center, or typical value in a data set. In order to investigate whether or not the phrase held such meanings for the teachers who used it, all of the preservice teachers studied were asked to give a written response to the follow-up item discussed above approximately four weeks after the initial assessment item had been administered.

As a result of the responses to follow-up item, four responses mentioning “measures of central tendency” and initially categorized as relational were re-categorized as multistructural. When asked to explain the phrase “measures of central tendency,” three of the four in the re-categorized group gave no response that went beyond saying that the mean, median, and mode are measures of central tendency. There was no indication that they thought about these as measures of what is average or typical for a set of data. The fourth re-categorized response said, “It’s how often a piece of data occurs in a data set.” In each of the four cases, the phrase “measures of central tendency” was not associated with words like “average” or “typical.”

Extended abstract level responses. Extended abstract level responses included all of the characteristics of those classified at the relational level. They went beyond relational level responses in that they included discussions of when one measure of center might be more useful than another. One extended abstract response noted that the mean is highly affected by outliers, while the median and mode are not. The response went on to show a situation where the mean was greater than the median and mode of a data set because of high numerical values in the data set. A second extended abstract response also noted that the mean is more affected than the median by extremely high or low values in a data set. The third response categorized as extended abstract noted that the mode is not always as indicative of center as median and mean because "a low value could appear the most but that does not mean that it is the average value of the data set." In all extended abstract responses, pre-service teachers moved beyond recognizing that mean, median, and mode were all measures of center to identifying a situation where one would be a more appropriate measure of center than another.

A Non-Categorized Response. While most of the responses of preservice teachers fit within one of the SOLO levels discussed above, the authors agreed that one response did not fit well into the categorization scheme:

> The statistical concepts are different in the methods they use. Each uses the data given in a different way to find the "middle" of the data. They are much different in the answers they produce. The concepts are the same in that they all use the same data and all of the data to find a similar concept in a different way.

The above response was the only one that did not engage in some form of definition-telling. The response resembles those at the relational level, which included recognition that all three concepts approximate the middle of a set of data. It might also be argued that the student giving the response is capable of extended abstract thinking, since it is acknowledged, "they are much different in the answers they produce." The response is presented here for the reader to help round out the picture of the types of thinking observed in comparing and contrasting measures of center.

Summary. The unistructural, multistructural, relational, and extended abstract levels helped describe much of the thinking observed in preservice teachers' comparisons of the mean, median, and mode. The overall picture the data paint is a set of levels of thinking that progress from little more than discussion of procedures for finding each measure to the point of recognizing that each one measures the center of a set of data in a different manner.

Preservice Teachers' Definitions and Procedures for Mean, Median, and Mode

While the SOLO levels of response tell part of the story of the data collected, part of the story remains untold. As noted in the discussion of SOLO levels, all of the preservice teachers, except one, engaged in definition-telling in giving a response. The definitions they described for mean, median, and mode had qualitative differences that were not discussed in the description of SOLO levels. The qualitative differences are summarized in Table 2 and discussed in detail below.

<INSERT TABLE 2 HERE>

It should be noted that definitions for the mean, median, and mode were not explicitly requested of the participants. The findings below should be read in that light. It is possible that if definitions were explicitly requested, the structures of the responses would have been different. We initially just asked participants to compare and contrast the measures of center. Drawing out a number of personal definitions was an unexpected benefit. Given these circumstances, we recognize that there may be more to say about the participants' knowledge of definitions than what we were able to document. However, we believe that it is still valuable to document the knowledge they revealed in this setting, because there are likely to be several situations in teaching practice when they share aspects of their knowledge of definitions with students without being explicitly prompted to do so.

The mean. The most common explanation of the mean included language that equated the words “mean” and “average.” Thirty-six preservice teachers stated that the mean is the average of a set of data. Among responses of this type, 16 did not say anything about the mean beyond calling it the average of a data set. Sixteen other responses included the add-and-divide algorithm as part of the explanation. Two responses stated that the mean is “obtained by averaging.” The last two responses in this category discussed the add-and-divide algorithm along with the fact that outliers affect the mean. In each case, the terms “mean” and “average” were presented as synonymous.

The second most frequent type of response was that which stated that the mean was just one type of average that could be computed from a data set. Five responses fell into this category. Three of the five students in this category supplemented their discussion of the mean by giving the add-and-divide algorithm for computing it. One student stated that the mean was most useful when the data are “equally distributed.” Another in this category said that the mean was an average that served as a “balance point” for a set of data, but did not provide any evidence of conceptual knowledge of a balance point interpretation of mean. In each instance of responses in the second most common category, the median and mode were called types of averages along with the mean.

The least frequent category of response for the mean was that which did not make any type of statement about the mean being an average. Instead, only procedures for finding the mean were discussed. Four responses fell into this last category. Three of the four students simply noted that the add-and-divide algorithm finds the mean. The last student in the category discussed a vague procedure for “leveling off” a set of data in order to identify the mean.

The median. The most common explanation for the median was that it is the “middle number” or “center number” when the data are placed in order. Twenty-five preservice teachers gave explanations of median that fell in this category. While the explanation captures some of the relevant aspects of the concept of median, it does not provide any information about what one is to do in the case of an even number of values in a data set.

The second most common explanation for the median was more structurally impoverished than the first. A total of 12 preservice teachers did not go beyond saying that the median is “the middle number” in a data set. The idea of ordering the data was not touched upon at all. The course of action to take when one is confronted with an even number of values in a data set was also missing.

The third most common category of explanations for the median contained five responses. The five preservice teachers in this category stated that the median is the “middle number” in a data set, and they also included the idea that it may not be an actual value in the data set. This category of response is similar to the second in that the idea of ordering the data in the set was not touched upon at all. The category is unique from the others discussed so far in that it did not imply that the median must be one of the numbers present in a set of data.

The least common category of response was that in which preservice teachers touched upon ordering data before finding the median, and they also discussed the idea that the median may not appear within the set of data. Only three preservice teachers gave responses that were structured in this manner. It was the only category of response in which the relevant aspects of ordering the data and determining the median when it was not among the values in the data set were both discussed.

The mode. The mode was most often described as the “most frequent number” or “most common number” within a data set. Thirty-four of the responses analyzed fell into this category. One of the notable missing aspects of this category of response was that being a “number” was stated as one of the defining characteristics of the mode. It was apparently not recognized that the mode could be a category as well as a number.

The second most frequent category of response in regard to the mode contained nine responses. In this category, the mode was still described exclusively as a “number.” However, explanations also noted that there is not necessarily one mode in any given data set. Three explanations stated that there might not be any mode in a given set of data. Another three stated that there might be more than one mode in a data set. The last three in this category stated that there might be one or more than one mode in a data set. The nine responses in the second category differed from those in the first category in allowing for multi-modality and/or no mode in a data set.

The least common type of response was that which left room for the mode to be a category rather than just a number. Only two preservice teachers gave responses structured in this manner. One of them stated that the mode “tells which data point is used the most.” The second said the mode is “the most frequent value in a data set.” While each response in this category did not explicitly state that the mode could be a category rather than a number, they also did not restrict the mode to being a numerical value. The fact that room was left for the mode to be a category could possibly have been the result of a fortunate choice of words rather than a thorough knowledge of the measure.

Summary. The preservice teachers' personal definitions revealed several frequent types of explanations about mean, median, and mode. The most frequent explanation for the mean equated it with *the* average of a data set. The most common explanation for the median described how one calculates the measure only in cases having an odd number of values in a data set. The most common explanation for mode implied that the measure is only of use for numerical data sets.

Discussion

The preceding description of preservice teachers' content knowledge of mean, median, and mode raises a number of issues for consideration. To conclude the paper, four issues relating to the results of the study are discussed: (i) How the thinking exhibited in the present study compares to thinking exhibited in previous studies; (ii) The relationship between PUFM and the thinking exhibited in the present study; (iii) Implications for preservice teacher education; and (iv) Implications for future research.

How the Thinking Exhibited Compares to that of Previous Studies

The thinking exhibited by preservice teachers in the present study strongly resembles the thinking of elementary and middle school children in the research literature summarized earlier (Mokros & Russell, 1995; Jones et al., 2000; Cai, 2000; Watson & Moritz, 2000; Mooney, 2002). The difficulties encountered by students in those previous studies – namely, mastering and applying the procedures for measures of center; recognizing mean, median, and mode as measures of what is central or typical in a set of data; and being able to work flexibly with a number of different measures of center – all turned up over the course of the SOLO analysis and the analysis of preservice teachers' definitions for mean, median, and mode. Only a few preservice teachers exhibited more abstract thinking than that of typical elementary students by identifying hypothetical situations in which one would be a more appropriate measure of center than another.

Relationship between PUFM and the Thinking Exhibited

The highest level of thinking observed in the present study, while having the characteristics of the formal-1 mode, could not be accurately described as being congruent with PUFM. The highest SOLO level exhibited was that in which preservice teachers began to consider when one measure might be more valuable than another, and that marked a significant leap in abstraction the lower levels of thinking. However, one might expect a teacher with profound understanding of elementary data analysis to identify several different situations in which one measure is more valuable than another, since PUFM is marked by its connectedness (Ma, 1999). A teacher with profound understanding of elementary data analysis would recognize that the median is often more effective than the mean in describing the center of a skewed data set – as some of the participants in the present study did – but would likely go beyond this to identify situations where the mode and mean are more desirable measures of center. Profound understanding of elementary data analysis, therefore, seems to be linked with a higher formal-1 SOLO level than that observed in the present study. Some participants in the present study showed some capacity for abstract thought typical of formal-1 by identifying one hypothetical situation in which a given measure of center is more effective than another, but none gave examples of several such situations or the structural connections among them.

One would also expect a teacher with a profound understanding of elementary data analysis to give well-developed definitions for each of the measures. Even though restating previously-learned definitions is largely a procedural exercise, sound procedural knowledge is still a necessary component of PUFM (Ma, 1999). Well-developed definitions were not given by all of the participants who gave the highest SOLO level of response observed. For instance, one participant at the highest SOLO level defined the median as “the middle number in the data set,” which is incomplete in the sense that it does not touch upon the procedural aspects of (i) ordering the data before determining the middle, and (ii) dealing with data sets containing an even number of values. The same participant defined the mode as “the most occurring number in the data,” which is incomplete in the sense that the definition does not encompass calculating the mode for non-numerical data sets. A minimal understanding of each individual measure can make it difficult for one to build a rich conceptual knowledge network that incorporates them (Nesher, 1986). The lack of more instances of thinking beyond the concrete-symbolic mode might therefore be explained by incomplete understandings of each individual measure. Presumably, it would be quite difficult to identify situations when one measure might be more appropriate than another if the individual measures themselves are inadequately understood. Concrete symbolic understanding of mean, median, and mode would seem to be prerequisite to attaining formal-1/PUFM thinking about them.

Implications for Preservice Teacher Education

Given the preceding discussion, the question of how to build profound understanding of elementary data analysis arises. One immediate implication is that a college course in introductory statistics is not sufficient, since all study participants had taken such a course, yet apparently had not attained extensive knowledge of measures of center. Cobb and Moore’s (1997) description of college statistics courses provides some insight as to why this might be the case. They noted that in many introductory courses, the focus is on selecting the appropriate abstract statistical method to apply to any given situation. Such a course is driven by abstract theory rather than concrete data. A student who has acquired formal-1 thinking might experience some success in such a course, but it seems unlikely that a student who functions predominantly in the concrete symbolic mode would. The present study illustrated that many preservice teachers seem to function predominantly in the concrete symbolic mode in thinking about the mean, median, and mode. Hence, it does not seem realistic to expect them to acquire profound understanding of the measures by engaging in a theory-driven course. Courses driven by concrete data are likely to be more meaningful.

While courses driven by concrete data appear to be important, it also important to keep in mind that content courses for elementary teachers should be designed to help them acquire formal-1 thinking, since PUFM is related to that mode. It is not sufficient for preservice teachers to master the concrete data analysis tasks they will use with their elementary school students. They also need to understand how the concepts embedded in those tasks relate to one another so they are able to create and select tasks that adequately engage students in learning the material they desire to teach (Ball, Lubienski, & Mewborn, 2001). Part of a mathematics content course for preservice teachers might consist of engaging in tasks from elementary school mathematics, but such content courses also need to help preservice teachers understand how the concepts in the tasks they are performing relate to one another. PUFM is obtained when a rich cognitive network of relationships among concepts has been built.

While carefully constructed coursework holds some potential for helping preservice teachers acquire PUFM, it seems to be the case that university coursework alone is not sufficient. Ma (1999) noted four different ways that the teachers she interviewed attained PUFM: studying teaching materials intensely, learning mathematics from colleagues, learning mathematics from students, and learning mathematics by doing it. Each of these happened within the context of teaching in an elementary school rather than taking a university course. While some of these ways of attaining PUFM can be built into university courses, it may be the case that many individuals do not attain PUFM until they have the need for it in within the context of a teaching situation. Situated cognition literature supports this hypothesis, since it suggests that sociocultural contexts have profound influences on learning (Bransford et al., 2000; Peressini, Borko, Romagnano, Knuth, & Willis, 2004). The need for PUFM may be more apparent for some individuals in the context of teaching children mathematics than it is in the context of taking a course at the university. Hence, interventions within the context of teaching may ultimately be of equal or greater importance than those within the context of undergraduate courses at the university.

Implications for Future Research

One important area for future research is investigating the question of how teacher educators can help in the development of PUFM for data analysis. Data-driven courses seem to have some role to play, so future studies could focus on working out plausible structures for such courses. Learning within the context of teaching is another important aspect of developing PUFM, so researchers might also wish to investigate approaches for integrating content knowledge learning within the context of teaching into undergraduate programs. While the present study highlights the need to develop approaches to help elementary teachers attain PUFM, the question of how it is best attained remains largely a task for further study.

One very specific question for researchers to investigate is how to help preservice teachers develop profound understanding of mean, median, and mode in light of the fact that many seem to come to the university with mainly isolated procedural knowledge of the measures. In the present study, the majority of the preservice teachers exhibited thinking at the unistructural and multistructural levels, each of which involved little more than displays of syntactic and algorithmic procedural knowledge. Some take the perspective that learning procedures before concepts interferes with attaining conceptual understanding (Skemp, 1987; Mack, 1990; Wearne & Hiebert, 1988; Byrnes & Wasik, 1991; Pesek & Kirshner, 2000). From this perspective, finding methods to help procedurally-oriented preservice teachers attain conceptual understanding is a monumental task. Since the university instructor generally has little control over the manner in which preservice teachers initially encounter the mean, median, and mode in elementary school, there is a great need for research to help university instructors better understand how to overcome interference with conceptual learning that may be caused by previously-built procedural knowledge of the three measures.

While the present study does not explicitly demonstrate how to help teachers develop conceptual understanding, it does provide some guidance for investigations concerned with that question. One way to measure the success of any given intervention is to document its impact on the SOLO levels students exhibit in response to tasks. Jones et al. (2001) used SOLO levels for this purpose in the context of a teaching experiment in data analysis. The impact of the teaching experiment was described in terms of the SOLO levels exhibited by the students before and after the intervention. Similar use could be made of the SOLO levels documented by the present study. The SOLO levels described by the present study may also inform the construction of hypothetical learning trajectories (Simon, 1995), which are an important component of teaching experiments (Steffe & Thompson, 2000). There is some reason to believe that SOLO levels can perform this function, especially since in some cases they seem to outline a developmental sequence (Chick, 1998). The categories of personal definitions for mean, median, and mode may also help in the construction of hypothetical learning trajectories for use with pre-service teachers, since the categories indicate possible difficulties with teachers' knowledge that may arise over the course of a teaching experiment.

Conclusion

The present study serves to shed light on the structure of preservice teachers' knowledge of mean, median, and mode. It illustrates that attaining sophisticated understanding of the content needed to teach measures of center at the elementary level is a non-trivial matter. There are complex conceptual and procedural ideas that need to be developed. Preservice teacher educators and researchers can potentially use the results of the present study to help design and implement instruction that results in the development of the rich procedural and conceptual understanding of mean, median, and mode that is characteristic of teachers with a profound understanding of the measures.

References

Australian Education Council. (1994). *Mathematics: A curriculum profile for Australian Schools*. Carlton, VIC: Curriculum Corporation.

Ball, D.L., Lubienski, S.T., & Mewborn, D.S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (4th edition) (pp. 433-456). New York: Macmillan.

Biggs, J. B. (1999). *Teaching for quality learning at university*. Philadelphia: Open University Press.

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

Biggs, J.B. & Collis, K.F. (1991). Multimodal learning and quality of intelligent behavior. In H.A.H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 57-66). Hillsdale, NJ: Erlbaum.

Bransford, J., Zech, L., Schwarts, D., Barron, B., Vye, N., & The Cognition and Technology Group at Vanderbilt. (2000). Designs for environments that invite and sustain mathematical thinking. In Cobb, P., E. Yackel, & K. McClain (Eds.), *Symbolizing and communicating in mathematics classrooms* (pp. 275-324). Mahwah, NJ: Erlbaum.

Byrnes, J.P., & Wasik, B.A. (1991). Role of conceptual knowledge in mathematical procedural learning. *Developmental Psychology, 27*, 777-786.

Cai, J. (2000). Understanding and representing the arithmetic averaging algorithm: An analysis and comparison of US and Chinese students' responses. *International Journal of Mathematical Education in Science and Technology, 31*, 839-855.

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

Campbell, K., Watson, J., & Collis, K. (1992). Volume measurement and intellectual development. *Journal of Structural Learning and Intelligent Systems*, *11*, 279-298.

Chick, H. (1998). Cognition in the formal modes: Research mathematics and the SOLO Taxonomy. *Mathematics Education Research Journal, 10*, 4-26.

Cobb, G., & Moore, D. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly, 104*, 801-823.

Collis, K.F., & Biggs, J.B. (1983). Matriculation, degree requirements, and cognitive demand in universities and CAEs. *Australian Journal of Education, 27*, 151-163.

Conference Board of the Mathematical Sciences. (2001). *The mathematical education of teachers*. Providence, Rhode Island: American Mathematical Society.

Friel, S.N., & Bright, G.W. (1998). Teach-Stat: A model for professional development in data analysis and statistics for teachers K-6. In S.P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in Grades K-12* (pp. 89-117). Mahwah, NJ: Lawrence Erlbaum.

Groth, R.E. (2002). Characterizing secondary students' understanding of measures of central tendency and variation. In D.S. Mewborn, P. Sztajn, D.Y. White, H.G. Wiegel, R.L. Bryant, & K. Nooney (Eds.), *Proceedings of the twenty-fourth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education: Volume 1* (pp. 247-257). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Hiebert, J., & Carpenter, T.P. (1992). Learning and teaching with understanding. In D.A.

Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65-97). Reston, VA: NCTM.

Hiebert, 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. 1-28). Hillsdale, NJ: Erlbaum.

Jones, G.A., Langrall, C.W., Thornton, C.A., Mooney, E.S., Wares, A., Jones, M.R., Perry, B.P., Putt, I.J., & Nisbet, S. (2001). Using students' statistical thinking to inform instruction. *Journal of Mathematical Behavior, 20*, 109-144.

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

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*, 259-289.

Ma, L. (1999). *Knowing and teaching elementary mathematics*. Mahwah, NJ: Lawrence Erlbaum Associates.

Mack, N.K. (1990). Learning fractions with understanding: Building on informal knowledge. *Journal for Research in Mathematics Education, 21*, 16-32.

Maykut, P., & Morehouse, R. (1994). *Beginning qualitative research: A philosophic and practical guide.* London: The Falmer Press.

Mevarech, Z.R. (1983). A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics, 14*, 415-429.

Mewborn, D.S. (2003). Teaching, teachers' knowledge, and their professional development. In Kilpatrick, J., W.G. Martin, & D. Schifter (Eds.), *A research companion to Principles* *and Standards for School Mathematics* (pp. 45-52). Reston, VA: National Council of Teachers of Mathematics.

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

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

Mooney, E.S. (2002). A framework for characterizing middle school students' statistical thinking. *Mathematical Thinking and Learning, 4*, 23-64.

Moore, D.S. (1997). *Statistics: Concepts and controversies* (4th edition). New York: W.H. Freeman.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Nesher, P. (1986). Are mathematical understanding and algorithmic performance related? *For the Learning of Mathematics, 6* (3), 2-9.

Pegg, J., & Davey, G. (1998). Interpreting student understanding of geometry: A synthesis of two models. In R. Lehrer & D. Chazan (Eds.), *Designing learning environments for developing understanding of geometry and space* (pp. 109-135). Mahwah, NJ: Lawrence Erlbaum Associates.

Peressini, Borko, Romagnano, Knuth, & Willis (2004). A conceptual framework for learning to teach secondary mathematics: A situative perspective. *Educational Studies in Mathematics, 56*, 67-96.

Pesek, D.D., & Kirshner, D. (2000). Interference of instrumental instruction in the subsequent relational learning. *Journal for Research in Mathematics Education, 31*, 524-540.

Pollatsek, A., Lima, S., & Well, A.D. (1981). Concept or computation: Students' understanding of the mean. *Educational Studies in Mathematics, 12,* 191-204.

Putnam, R.T., Heaton, R.M., Prawat, R.S., & Remillard, J. (1992). Teaching mathematics for understanding: Discussing case studies of four fifth-grade teachers. *Elementary School Journal, 93,* 213-228.

School Curriculum and Assessment Authority & Curriculum and Assessment Authority for Wales. (1996). *A guide to the national curriculum*. London: School Curriculum and Assessment Authority.

Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57,* (1), 1-22.

Simon, M.A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*, 114-145.

Skemp, R.R. (1987). *The psychology of learning mathematics*. Hillsdale, NJ: Erlbaum.

Steffe, L.P., & Thompson, P.W. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A.E. Kelly & R.A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 267-306). Mahwah, NJ: Erlbaum.

Upton, G., & Cook, I. (2002). *Oxford dictionary of statistics*. Oxford: Oxford University Press.

Watson, J.M., Collis, K.F., & Moritz, J.B. (1997). The development of chance measurement. *Mathematics Education Research Journal, 9,* 62-80.

Watson, J.M., & Moritz, J.B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37,* 145-168.

Watson, J.M., & Moritz, J.B. (2000). The longitudinal development of understanding of average. *Mathematical Thinking and Learning*, *2*(1&2), 11-50.

Wearne, D., & Hiebert, J. (1988). A cognitive approach to meaningful mathematics instruction: Testing a local theory using decimal numbers. *Journal for Research in Mathematics Education*, *19*, 371-384.

Weber, K., & Alcock, L. (2004). Semantic and syntactic proof productions. *Educational Studies in Mathematics*, *56*, 209-234.

Zawojewski, J.S., & Shaugnessy, J.S. (2000). Data and chance. In E.A. Silver & P.A. Kenney (Eds.), *Results from the seventh mathematics assessment of the national assessment of educational progress*. (pp. 235-268). Reston, VA: National Council of Teachers of Mathematics.

Table 1

Levels of Thinking in Regard to Comparing and Contrasting Measures of Center

Level/Mode	Qualitative Characteristics	Frequency
Unistructural/ Concrete symbolic	Response relies solely upon discussing the processes for finding each measure. (i.e., "process-telling").	8
Multistructural/ Concrete symbolic	Response includes process-telling along with a vague notion that mean, median, and mode can all be used as data analysis tools.	21
Relational/ Concrete symbolic	Response includes process-telling along with the idea that mean, median, and mode all measure the center of a data set and/or what is typical about the set.	13
Extended Abstract = Unistructural/ Formal-1	Response includes process-telling but goes beyond it to discuss situations when one of the three measures (mean, median, mode) might be a better measure of center and/or typicality than another.	3

Table 2

Definitions/procedures for mean, median, and mode

Measure	Definition/procedure description	Frequency
Mean	Process only: Add-and-divide algorithm given	3
	Leveling off procedure discussed vaguely	1
	The mean is *the* average	18
	The mean is *the* average with the add-and-divide algorithm given	18
	The mean is *an* average that is the balance point	1
	The mean is *an* average when data are "equally distributed"	1
	The mean is *an* average and the add-and-divide algorithm is given	3
Median	The median is the middle number	12
	The median is the middle number when the data are in order	25
	The median is the middle number, but may not appear in the set	5
	The median is the middle number when data are in order, but may not appear in the set	3
Mode	The mode is the most frequent *number*	34
	The mode is the most frequent *number* and there may be no mode	3
	The mode is the most frequent *number* and there may be several	3
	The mode is the most frequent *value*	2
	The mode is the most frequent *number* and there may be several or none	3