STATICICAL PREPARATION OF TEACHERS: PRESERVICE ELEMENTARY TEACHERS (PSTS) CONCEPTIONS OF DISTRIBUTIONS OF DATA – THINKING ABOUT MEASURE OF CENTER AND VARIAIBLITY

Eva Thanheiser
Portland State University
evat@pdx.edu

Jennifer Noll
Portland State University
noll@pdx.edu

Aisling Leavy
Mary Immaculate College - University of Limerick
Aisling.leavy@mic.ul.ie

Prior research investigating PSTs' statistical thinking is sparse, yet teacher educators need to know how best to prepare future teachers for their work and given the increasing importance of STEM education we need teachers who are capable of preparing students at an early age to think about data. In this paper we share the results from a study investigating pre-service teachers' (PSTs') initial thinking about distributions of data in four different contexts. In particular we investigated how PSTs reasoned about different distributions of data - including how they consider measures of center and measures of variability. Data was collected through surveys and interviews conducted prior to their beginning a statistical unit in their elementary teachers mathematics content course.

Without basic statistical literacy adults are unlikely to have the knowledge base to make informed personal decisions and, on a professional level, the door to many higher paying jobs closes. (Cobb & Moore, 1997; Garfield & Ben-Zvi, 2008).

Why Focus on PSTs' Conceptions?

Recent research on teacher knowledge has shown that the knowledge needed for teaching mathematics/statistics is complex and multifaceted (i.e. Hill, Ball, & Schilling, 2008; Ma, 1999; Shulman, 1986) and goes beyond content knowledge including such aspects as knowledge of how students think about the mathematical/statistical content. Hill, Ball, and Schilling (2008) introduced a framework for *mathematical knowledge for teaching* listing six different types of knowledge needed to teach mathematics, three in the realm of subject matter knowledge and three in the realm of pedagogical content knowledge (see Figure 1). In addition, teachers' mathematical knowledge for teaching has been empirically linked to instruction (Borko & et al., 1992; Fennema & Franke, 1992); and student achievement gains (Hill, Rowan, & Ball, 2005).

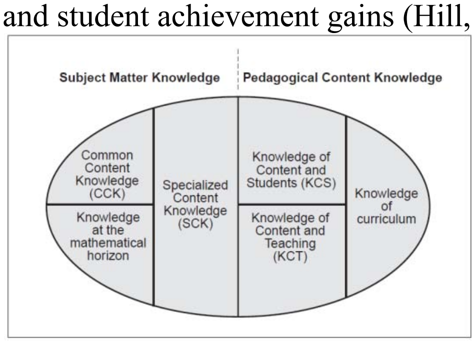


Figure 1: Domain map of mathematical knowledge for teaching (Hill et al., 2008).

New mathematical ideas are built on currently held conceptions. In order to help PSTs

Wiest, L. R., & Lamberg, T. (Eds.). (2011). Proceedings of the 33rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Reno, NV: University of Nevada, Reno.

develop such multifaceted knowledge of mathematics/statistics for teaching, teacher educators need to build on PSTs incoming conceptions. The conceptions we examine in this paper fall beyond common content knowledge (i.e. how to calculate the mean for example) and are situated in the realm of specialized content knowledge (i.e. knowledge of center and distribution specifically used for teaching). Mathematics teacher educators need to understand PSTs' initial conceptions when they enter our classrooms (Bransford, Brown, & Cocking, 1999) as these types of understandings underpin and inform the pedagogical decisions made when teaching.

Why Focus on PSTs' Conceptions of Statistics?

The importance of statistics for all citizens living in a democratic society is reflected by the Guidelines and Assessment in Statistics Education (GAISE) report (Franklin et al., 2007). GAISE highlights the important role of statistics education in preparing students to navigate our current society (reading newspapers, understanding political implications etc.). In particular, understanding distributions of data is a unifying theme in the study of statistics and one of the foci of the elementary math curriculum as laid out by NCTM and Common Core Standards (National Governors Association & Council of Chief State School Officers, 2010). Bakker and Gravemeijer (2004) suggest that distribution is an "organizing structure or conceptual entity" for looking at data (p. 149) and they argue that without a notion of distribution it is not possible to reasonably summarize a data set and make appropriate choices between different measures of center. While substantial efforts have been made to better understand students' statistical reasoning and how they think about distributions of data, there is a paucity of empirical research investigating teachers' statistical reasoning (Shaughnessy, 2007b). What little empirical research has been conducted (Groth & Bergner, 2006; Jaccobe, under review; Leavy & O'Loughlin, 2006) indicates that PSTs do not have the knowledge needed to teach statistics at the level endorsed by GAISE. It is not sufficient to know that teachers do not have adequate statistical knowledge for teaching statistics, teacher educators need to better understand teachers' statistical knowledge when they enter our classrooms so we can build on those conceptions when working with teachers.

In this study we examined PSTs' conceptions of distributions of data in four different contexts. In particular, we investigated: (a) how PSTs might compare two different distributions of data that had the same mean and median, but differed in variability; (b) the ways in which PSTs' construct a data set for a given mean (with a particular focus on the range of the data sets they created); (c) whether PSTs recognize the mean as an appropriate measure to compare two data sets of different size; and, (d) the ways in which PSTs' think about the median through a task that focuses on possible student interpretations of median. The results of this study provide a glimpse into U.S. PSTs' initial knowledge of distributions of data. We selected these particular tasks for two reasons. First, we wanted to build off the small body of prior research (see Leavy & O'Loughlin, 2006) and look for possible comparisons between U.S. PSTs and PSTs in Ireland. Second, the tasks used in this study align with Level A in the GAISE document and highlight components of statistical knowledge that PSTs need to be successful in their work. Thus, the strategies used by PSTs' when working on these tasks provides valuable insights into their specialized content knowledge; these insights inform the decisions that teacher educators make regarding the types of experiences we provide in our pedagogy of mathematics courses.

Methods

Twenty-seven PSTs enrolled in the 2nd of 3 mathematics-for-elementary-school-teachers courses completed a written survey followed by a brief interview. The survey items were drawn from various previous research studies (as cited below) to give an initial insight into PSTs' knowledge of distributions of data. We introduce the tasks in the results section. Data analysis was conducted by the first two authors and focused on the survey results. The authors independently categorized student responses to each item and then met to discuss the categories. Once common categories were established the data was coded using those common categories. Disagreements were resolved through discussion.

Results

What Measures Do PSTs Use When Comparing Distributions of Data?

To examine whether PSTs pay attention to distribution in a given data set we asked them to respond to the Movie Wait Time Task (Shaughnessy, 2007a). On this task PSTs are presented with a scenario involving the comparison of two distributions of data. In response to the task, 19 PSTs disagreed with Eddie, 4 PSTs agreed with Eddie; and 4 did not express agreement or disagreement. 25 of the PSTs mentioned the range or the variability of the wait times in their response. The 4 PSTs who agreed with Eddie agreed that the average is 10 but all 4 mentioned the variability in wait times. The two PSTs who did not mention variability performed calculations to find the mean. Thus at least in this task almost all PSTs recognized variability as a measure to take into account when comparing distributions of data.

Movie theaters show commercials along with previews before the movie begins. The *wait-time* for a movie is the difference between the ADVERTISED start time (like in the paper) and the ACTUAL start time for the movie. A class of 21 students investigated the wait-times at two popular movie theater chains: Maximum Movie Theaters and Royal Movie Theaters. Each student attended two movies, a different movie in each theater. The class's results are shown in the chart below. (Times were rounded to the nearest half-minute.) The students in the class found the median wait-time for both of the theaters to be 10 minutes. The students also calculated the mean wait-time for each theater to be 10 minutes.

Here's what Eddy, a 4th grader, concluded: There is no difference in wait-times for the theaters because they both were about 10 minutes.

Do you agree or disagree with Eddy? Explain all of your reasoning:

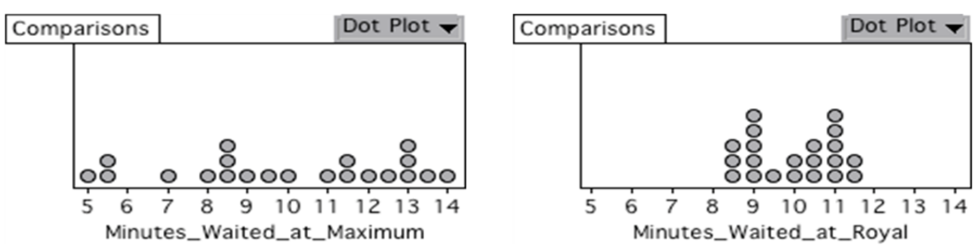


Figure 2: Movie Wait Time Task (Shaughnessy, 2007a)

Can PSTs Construct a Distribution of Data to Reflect a Given Mean? How Does Variability Factor in their Decisions?

To examine whether PSTs' can construct a distribution to reflect a given "average" the PSTs were asked to label prices on individual potato chip bags given that the average price is 27 cents. PSTs were asked to come up with a second labeling excluding 27 as a value (see Figure 3). In our analysis we focused on the PSTs' strategies and the range of the data set they provided.

Wiest, L. R., & Lamberg, T. (Eds.). (2011). Proceedings of the 33rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Reno, NV: University of Nevada, Reno.

Potato Chip Task A: The average price of a bag of chips is 27cents. We have seven bags of chips each of which has an empty price tag. Place prices on each of the bags so that the average price is 27 cents.	*Potato Chip Task B:* The average price of a bag of chips is 27cents. We have seven bags of chips each of which has an empty price tag. Place prices on each of the bags so that the average price is 27 cents. *However, none of the bags of chips can cost 27 cents.*

Figure 3: Potato Chip Tasks (see Leavy & O'Loughlin, 2006; Mokros & Russell, 1995)

For Part (A) of the Potato Chip Task our results showed that 14 PSTs (about 50% of the PSTs as opposed to 25% of PSTs in Leavy & O'Loughlin's study) put a 27 into each of the boxes on the first task (all listed results can be seen in Table 1). Seven PSTs (about 25% of the PSTs, similar to Leavy and O'Loughlin's study) started with 27 in the middle and then varied the numbers by +1/-1 as they moved outward resulting in the following prices: 24, 25, 26, 27, 28, 29, 30. Three PSTs used the mean algorithm, one selected numbers across a wide range summing to 189, two iteratively subtracted numbers from 189. One PST entered some 27s and then used balancing (i.e., adding and subtracting the same number from 27 in order to "balance" out to 27) to find the other prices in the data set. Two PSTs used what we call reasonable responses – using some type of reasoning that the bags combined have to be $1.89 the average price would have to be 27 thus the values would be *around* 27.

As expected, it seemed most PSTs understood the "average" as the "mean" rather than another measure of center. One PST used the mode in their reasoning for all 27s. Twenty-three PSTs chose numbers for the prices that stayed within a close range [24,30]. Four PSTs chose wide ranges for their data such as (100, 20, 20, 10, 10, 10, 19). These PSTs either gave no justification or used the mean algorithm (summing up to $1.89) as their explanation.

For Part (B) of the Potato Chip task all PSTs interpreted "average" as mean. Twenty-one PSTs gave values in between 20 and 35, while six PSTs gave wide ranges in value. Most PSTs used the mean algorithm some with guess and check, some with some basic balancing. Only four PSTs used a conceptual approach using balancing (pairs and one set of three) or a reasonable approach. Initially many of the responses in Part (A) looked like balancing approaches (i.e., balanced around the 27 like this response: 24, 25, 26, 27, 28, 29, 30), however, in Part B these PSTs seemed to focus more on the algorithm for finding mean. This indicated to us that just because their work "looks like" a balance approach does not mean it is. In fact, PSTs may be drawing on procedural rather than conceptual knowledge and using the mean algorithm to find pairs that average to 27.

Table 1: PSTs' responses to the Potato Chip Tasks

Responses to Part (A)	Reasoning for Part (A)	Counts
All 27s	No Justification	9
	Mean Algorithm	6
	Mode	1
27 & Values close to 27 (e.g., 26, 26, 27, 27, 27, 28, 28)	No Justification	3
	Mean Algorithm	1
	Balance	1
24, 25, 26, 27, 28, 29, 30	No Justification	1
	Mean Algorithm	1
	Mean Algorithm & Basic Balance	3
	Balance	2
Wide Ranges (e.g., 100, 20, 20, 10, 10, 10, 19)	No Justification	1
	Mean Algorithm	3

Responses to part (B)	Reasoning to Part (B) & Interview follow-up	Counts*
Range 20-35	Mean Algorithm / Guess & Check	6
	Mean Algorithm	3
	Mean Algorithm & Basic Balance (few over/few under)	8
	Algorithm & Balancing (pairs or triples)	1
	Reasonable	2
	Balance	1
Wide Ranges	Mean Algorithm	3
	Mean Algorithm / Guess & Check	3

Responses to Part (A)	Reasoning for Part (A)	Counts	Reasoning Part (B)
24, 25, 26, 27, 28, 29, 30	No Justification	1	Balance
	Mean Algorithm	1	Mean Algorithm
	Mean Algorithm & Basic Balance	3	Mean Algorithm (2) Mean Algorithm /Guess and check (1)
	Balance	2	Algorithm

All of the wide range responses appeared to use only the mean algorithm in order to justify their solutions – no evidence of average as balance or reasonable. Those PSTs who displayed balance type reasoning did so via general sense (some over, some under) or pair-balance. Only one PST displayed explicit knowledge of balancing three values; we argue that this may be an important first step in developing a more conceptual understanding of mean as point of balance.

Combining the results of the Movie Wait Time Task and the Potato Chip Task it seems that PSTs can recognize variability in the data but when asked to create a set of data may stay close to the mean value. Although our sample size is small and we only have evidence of this from two tasks, this finding is consistent with work done by Rubin, Bruce and Tenney (1991) with middle and high school students who seemed to recognize variability in some data contexts, but not in others. In particular, it seems that in tasks where students are asked to provide the data they may gravitate toward the expected value and not consider variability in the data; yet, in other contexts where students see different sets of data they may be more likely to focus on variability within the data or between data sets. However context is always an important consideration when dealing with data. In this case the actual context of the potato chip task (i.e. pricing bags of chips) may impact the variability of the distributions constructed by PSTs. Even though PSTs may recognize the 'possible/theoretical' variability in the distribution of prices, the context may limit the variability of the values they choose to represent the distribution of data. This raises an interesting issue for further research, would PSTs construct data sets with a wider variability if the context was not restrictive?

Do PSTs Recognize the Mean as a Way to Compare to Data Sets of Different Size?

A third task (see Figure 4) was given to examine whether PSTs would recognize the mean as a suitable measure to compare the data sets (Leavy & O'Loughlin, 2006). This recognition of the mean as a comparison measure falls into the realm of specialized content knowledge (SCK) and is an important component of the knowledge needed for teaching.

Basketball Task: Coach Andrews is selecting students to play on the All Star Team. He has decided to look at the scoring of each player during the last three weeks of the season. Below are the points scored by Bob and Deon. If Coach Andres can only select one of the two players, who would you recommend and why? [Note: The coach has not received scores for Deon's last two games played]

Bob	**21**	**16**	**23**	**21**	**20**	**17**	**16**	**22**
Deon	**24**	**18**	**21**	**25**	**22**	**28**		

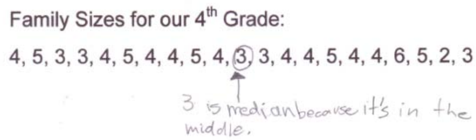
Figure 4: Basketball Task (developed by McGatha et al., 2002 and also used by Leavy & O'Loughlin, 2006)

In this task 15 PSTs (a little more than half of the PSTs in the US vs. 70% of the PSTs in Ireland) recognized the mean as an appropriate measure to compare scores. Thirteen PSTs compared group means while 2 PSTs made equal sized groups (putting in 0s or 23s for Deon's last scores – 23 is the average of Deon's first 6 scores). Eleven PSTs compared the scores without comparing the means. Seven PSTs compared pairs of games, one PST, for example, stated: "I would choose Deon because he has done better than Bob 5 out of the 6 games. Deon's avg. is higher than Bobs." One PST compared total points scored thus far as well as the differences in the first 6 games, one PST looked at consistently scoring above 18, one PST gave no response, and two PSTs wanted to await the last two scores before making a decision. This task illuminates that only about half of the PSTs recognized the appropriateness of the mean to compare data sets of different sizes.

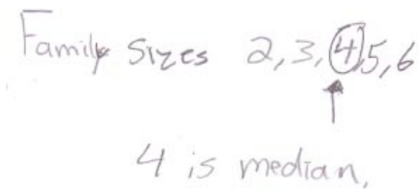
How Do PSTs' Think about the Median in Response to Children's Incorrect Interpretation of the Median?

A fourth task was constructed to examine whether PSTs would recognize children's misconceptions when responding to a task about the median (see Figure 5).

Family Size Problem: In a 4th grade class the students collected data on their family sizes (how many people are in their immediate family). The teacher put the data on the board and asked the students to find the median of the data. The students copied down the numbers and wrote their answers (See Figures 3a and 3b for Anna and Lisa's solutions). Please write a reaction to each of the solutions. In your reaction please include (a) whether you think the answer is correct/incorrect, and (b) what you might do next with that child if you were the teacher.



Anna's solution to identifying the median in the family size problem.



Lisa's solution to identifying the median in the family size problem

Figure 5: Two children's solution to a Family Size Problem**.**

The problem was designed to highlight two potential conceptual difficulties students may have understanding the median, the order of the data and the number of times each data point appears. Thus, the task was designed so that one student chose the middle number of the unordered data set, another the middle number of the various family sizes. The first of these solutions results in an incorrect answer the second in a correct answer, but based on an incorrect argument. Agreement with Anna's solution indicates a conception of median that may not include the notion of the importance of ordered data. Agreement with Lisa's solution, indicates that the frequencies of each family size do not matter. At Level A of the GAISE document students "should understand that the median describes the center of a numerical set in terms of how many data points are above it and below it" (p. 29). All but one PST identified the misconception in Anna's solution (i.e. that the data set needs to be ordered), however only 15 of 27 PSTs recognized that Lisa's argument was incorrect (see Table 2). This may be due to the fact that it resulted in the correct answer or due to the fact that the PSTs did not pay attention to the fact that each data point needs to be accounted for when calculating the median. One PST argued that Lisa's solution is correct because "each number only needs to be represented once to find the median" and another that Ana's is incorrect "Anna is … confused what the middle is in this situation… she should be considering the range of the numbers given (2-6) not the order in which they were written." Thus while most of the PSTs recognized that order mattered when all the data was presented about 44% of the PSTs did not recognize the need for all data points when they were not presented in the solution.

Table 2: PSTs responses to Family Size Problem:

	Identified misconception	Did not identify misconception
Anna's solution	26	1
Lisa's solution	15	12

Conclusions

The PSTs responses to these four tasks give a glimpse into U.S. PSTs' initial thinking about distributions of data. All PSTs, except one, displayed knowledge of the mean algorithm (Potato Chip Task), however, it can often be difficult to decipher procedural averaging of pairs from conceptual averaging through balancing (Potato Chip Task). What may look like balancing may be grounded in the algorithm. All but one of the PSTs interpreted average as "mean" – this is not surprising as the terms are often used interchangeably and PSTs often encounter the notion of average referring to all three measures of center for the first time in their mathematics/statistics for preservice teachers course. When given tasks like comparing two data sets of unequal size (Basketball task) only about half the PSTs recognized the mean as an appropriate measure of center and when confronted with a child's solution to the Family Size Problem about half the PSTs did not recognize the need for all data points to calculate the median. This shows that these PSTs conceptions of mean and median need to be further developed to be at the level recommended in GAISE. When we relate these findings back to the knowledge needed to teach mathematics, we can conclude that most PSTs possess common content knowledge (CCK) of the measures that index distribution, however performance on the tasks reveal problems with specialized content knowledge (SCK) of center and variability and understandings that extend beyond the ability to carry out procedures and manipulation of quantities. It is this specialized content knowledge of statistics that supports thinking and reasoning about distributions and that underpins the instructional decisions PSTs may make in classrooms when providing explanations for procedures, when selecting tasks and in recognizing sources of students errors.

PSTs seem to recognize variability in the data when presented with a a (Movie Wait Time Task) or already constructed data sets, but seem to provide responses close to the mean value when asked to construct data from a given mean (Potato Chip Task). Given that distributional reasoning entails the ability to simultaneously consider measures of center and measures of variability in data, more work needs to be done to understand how PSTs consider distributions of data and how they coordinate measures of center and variability when reasoning about data. Furthermore, we cannot ignore the potential influence of context on the decisions PSTs make when choosing values to represent a distribution of data. This area needs to be explored in future research.

To help PSTs develop their ability to reason coherently about distributions of data, we may need to explicate the relationship between the various measures of center and the data points from which they are created, as well as the role variability plays in statistics. In particular to help PSTs develop a conceptual understanding of the mean we may need to begin with balancing two numbers (pair) or fair share, then move to balancing 3 numbers or fair share among three, and then move to balancing more numbers or fair share. In addition we need to focus PSTs' attention to the range of values in a data set. The fact that one mean can originate from many different data sets is something that PSTs need to come to understand and they must begin to negotiate how context may play a role in the type of variability they expect to see in data.

References

Bakker, A., & Gravemeijer, K. P. E. (2004). Learning to reason about distribution. In J. B.-Z. Garfield, D. (Ed.), *The challenge of developing statistical literacy, reasoning and thinking.* (pp. 147-168). Dordrecht, The Netherlands: Kluwer.

Borko, H., & et al. (1992). Learning to Teach Hard Mathematics: Do Novice Teachers and Their Instructors Give Up Too Easily? *Journal for Research in Mathematics Education, 23*(3), 194-222.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How People Learn.*

Washington, D. C.: National Academy Press.
Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104*(9), 801-823.
Fennema, E., & Franke, M. L. (1992). Teachers' knowledge and its impact. In D. A. Groews (Ed.), *Handbook of reserach on mathematics teaching and learning* (pp. 147 - 164). New York: Macmillan.
Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report*. American Statistical Association. Retrieved November, 2010, from the World Wide Web:
Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*: Springer.
Groth, R. E., & Bergner, J. A. (2006). Preservice Elementary Teachers' Conceptual and Procedural Knowledge of Mean, Median, and Mode. *Mathematical Thinking & Learning: An International Journal, 8*(1), 37-63.
Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking Pedagogical Content Knowledge: Conceptualizing and Measuring Teachers' Topic-Specific Knowledge of Students. *Journal for Research in Mathematics Education, 39*(4), 372-400.
Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371-406.
Jaccobe, T. (under review). Elementary school teachers' understanding of the mean and median - Implications for teacher preparation programs. *Journal of Mathematics Teacher Education*.
Leavy, A., & O'Loughlin, N. (2006). Preservice Teachers Understanding of the Mean: Moving beyond the Arithmetic Average. *Journal of Mathematics Teacher Education, 9*(1), 53-90.
Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Erlbaum.
McGatha, M., Cobb, P., & McClain, K. (2002). An analysis of students' initial statistical understandings: developing a conjectured learning trajectory *Journal of Mathematical Behavior, 21*, 339 - 355.
Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education, 26*, 20-39.
National Governors Association, & Council of Chief State School Officers. (2010). *Common Core Standards Intitiative*. Retrieved Februray, from the World Wide Web: http://www.corestandards.org/
Rubin, A., Bruce, B., & Tenney, Y. (1991). *Learning about sampling: Trouble at the core of statistics.* Paper presented at the Proceedings of the Third International Conference on Teaching Statistics, Voorburg, The Netherlands.
Shaughnessy, J. M. (2007a). An investigation of secondary students and teachers' conceptions of variation.
Shaughnessy, J. M. (2007b). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 957-1009). Charlotte: National Council of Teachers of Mathematics.
Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.