Statistical Knowledge for Teaching

Alejandra Sorto, Michigan State University and Montgomery College
and Alexander White, American University

More than ever before, data analysis and statistics are an important part of the school mathematics curriculum in the United States. Because new curricula are challenging teachers not just with new teaching approaches but also with new content, the preparation of teachers and teachers' knowledge in this field have also become a special interest. The purpose of this study is to contribute to the current discussion and analysis of *content knowledge for teaching* by identifying important aspects of content knowledge for teaching data analysis and statistics and to describe the knowledge of prospective middle school teachers with respect to these aspects.

There are two central research questions under consideration here. First, what are the important aspects of statistical knowledge needed for teaching at the middle school level? In particular, what aspects of content and pedagogical content knowledge do middle school teachers need in order to teach data analysis and statistics? Once these various aspects have been identified a second question arises, what do prospective teachers know about the various aspects of statistical knowledge for teaching? In particular, what do they know about the content and what is their pedagogical content knowledge of data analysis and statistics?

Methodology

In order to answer the first research question analysis of existing documents was conducted. More specifically, to identify the important aspects of content (i.e. "the big ideas"), documents at the student and teacher level were considered. At the student level, ten sets of standards from those states that have middle grades certification, *Principles and Standards for School Mathematics* (NCTM, 2000), and a standardized middle school assessment (PRAXIS II) were reviewed and analyzed. At the teacher level, the report, *The Mathematical Education of Teachers* (CBMS, 2001) was also analyzed. To identify the other important aspects of knowledge for teaching such as pedagogical knowledge, knowledge of students as learners and knowledge of assessment others document were reviewed. These include, the National Board for Professional Teaching Standards, the National Council for Accreditation of Teacher Education (NCATE) Standards, professional standards from several states (Florida, Georgia, Missouri, and North Carolina), *Professional Teaching Standards* (NCTM, 1991), *Knowing and Learning Mathematics for Teaching* (NRC, 2001). The Unit Data About Us from the Connected Mathematics Project (Lappan, et. al, 2002) was reviewed to examine how all these aspects relate to each other.

A *content matrix* originally developed by Porter(2002) to measure agreement between standards and assessment in mathematics was adapted to the context of knowledge for teaching. The matrix crosses content areas and level of performance. The levels of performance were inspired by frameworks in Friel et. al (1998) and Garfield (2002). The matrix was then used to quantify the frequency of occurrence of each aspect and to measure the agreement between documents. Based upon this analysis the most important and ubiquitous aspects were identified.

Results from the content analysis show that there is a great variability in what aspects are considered important. By putting them all together in a single map (see Figure 1), we can identify clearly that they put more emphasis on graphical representation and measures of center at the level of construction/computation and appropriate use.

Topics:

1. Bar graphs, Tables, Pictograms
2. Histograms, Stem and Leaf
3. Scatter plots
4. Boxplots
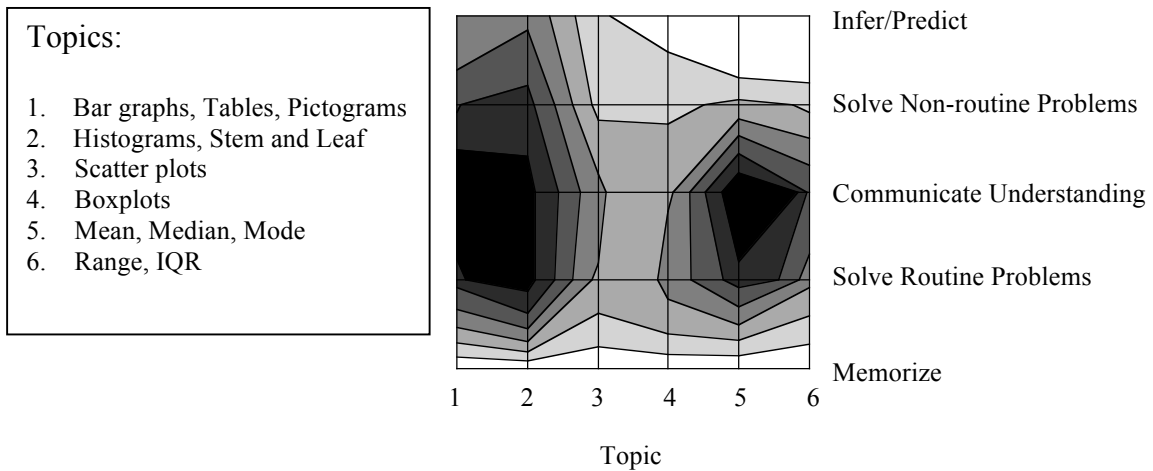5. Mean, Median, Mode
6. Range, IQR

Figure 1. Content Map for Middle School Teacher Knowledge In Statistics

Similar analysis was done for the documents related to teaching. Here more topics were identified and only three levels of performance. As in the content analysis, the documents related to teaching show great variability. In figure 2, we can identify more emphasis in teaching methods/strategies at the level of selection of appropriate and powerful method/strategy.

Topics:
Knowledge of students as learners
1. Conceptions and misconceptions
2. Ways of learning (developmental)
3. Student work
4. Classroom discourse
Knowledge of pedagogy
5. Teaching strategies/methods
6. Explanations
7. Teaching resources
Knowledge of assessment
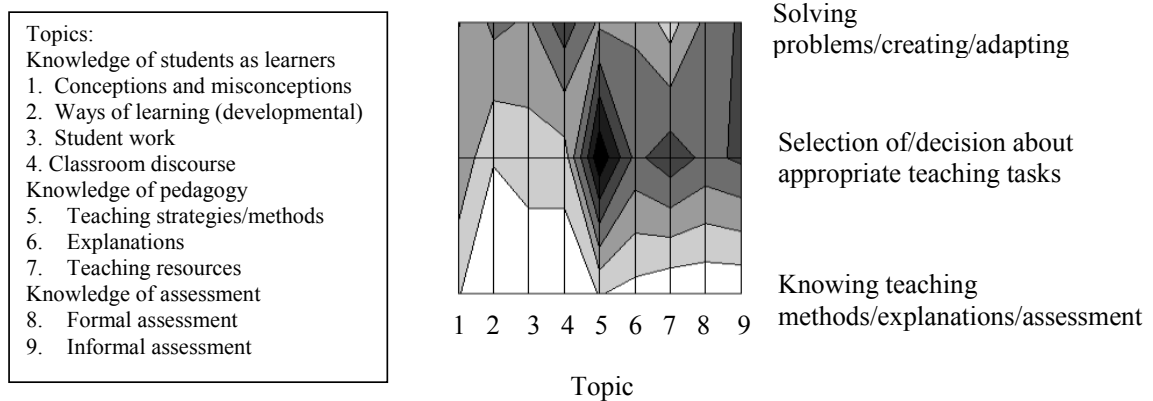8. Formal assessment
9. Informal assessment

Figure 2. Content Map for Middle School Teacher Knowledge for Teaching

The second research question - what do prospective teachers know about the various aspects of statistical knowledge for teaching? -was addressed via the development and application of a written instrument, complemented by face-to-face interviews A total of 42 prospective middle school teachers from the middle Atlantic region participated in the study. Most of the participants were female seniors in their twenties with a very strong mathematics and education background. 38 out of the 42 prospective teachers surveyed had taken at least one course beyond calculus. The average number of mathematics classes starting at calculus is 5.2. Subjects had an average of 2 mathematics education classes. All but four had taken a basic or introductory statistics class at the college level. Seven of the subjects were interviewed.

Instrument
The instrument we developed to assess the aspects identified in the analysis of documents measures two major domains of knowledge: pure statistical knowledge, and statistical knowledge applied to teaching. Reliability of the instrument as a whole was tested using a Cronbach-Alpha. The results show $\alpha = .80$ overall ($\alpha = .74$ for pure statistical knowledge, $\alpha = .5$ statistical knowledge for teaching).

The statistical knowledge domain includes the following aspects:
- Reading, interpreting, and inferring data using graphical displays such as histograms, line plots, stem-and-leaf plots and tables.
- Recognition, description and use of shapes of data distributions.
- Development and use of measures of center and spread.

For each aspect, the instrument measures different levels of performance. For graphical displays, three levels are considered: extracting information from the data, finding relationships in the data, and moving beyond the data (Friel et al., 2001). For distribution, measures of center, and spread, the three levels of performance or cognitive outcomes are statistical literacy, reasoning, and thinking (delMas, 2002; Garfield, 2002). Statistical literacy refers to recognition, identification, computation or basic understanding of concepts. Statistical reasoning refers to the way students reason with statistical ideas when asked why or how results are produced. For example, knowing why the type of data leads to a particular graph or statistical measure, knowing what factors influence the shape of a distribution, selecting the appropriate measure of center or, or interpreting what these measures reveal about the data. The third level, statistical thinking, refers to the application of students' understanding to real world problems, "to critique and evaluate the design and conclusions of studies, or to generalize knowledge obtained from classroom examples to new and somewhat novel situations" (delMas, 2002, p. 6). For measures and distribution this might mean using them to make predictions and inferences about the group to which the data pertain.

As for statistical knowledge applied to teaching, the instrument focuses on the knowledge of students as learners. Although there are many other domains, such as pedagogical and assessment knowledge, they are not considered here and are left for future investigation. The aspects considered for the instrument within this domain are:
- Interpretation of students' oral and written responses in relation to the content.
- Examination of students' strategies and solutions to exercises to make inferences about their understanding.

In each item the prospective teacher was confronted with either a student response or solution to an exercise. In contrast to the items related to pure statistical knowledge, these items are not characterized by level of performance. In some items the prospective teacher is asked to judge whether a response is correct and then to explain what thought process the student might have used to arrive at that response. In others the prospective teacher is asked to describe the method or solution used by the student and then to make inferences about his/her understanding.

Data Analysis
The responses were characterized not only by identifying levels of correctness but also by observing response patterns and solution strategies. Therefore, the analysis of the written instrument was done at two levels, item analysis and global analysis. Item analysis was conducted by coding each item's responses utilizing a rubric for levels of correctness (0 – 4) adapted from Garfield (1993) and Thompson & Senk (1998). To assure reliability of the rubric, two graders were trained to code the responses independently. The researcher adjudicated when there was disagreement between graders.

Performance

Prospective teachers' overall performance in the written instrument was obtained by scoring each question (some items had more than one question) on a scale of 0 to 4. The total number of questions was 18, making a total of 72 possible points. Totals for each participant were converted to a percentage. The average mean percentage in the assessment for the 42 participants is

58.86%, the median is 60.42%, the standard deviation 16.21% and the distribution is slightly skewed to the left. None of the participants showed mastery or near mastery level (score 4 or 3 from rubric) on all the questions of the instrument.

The domain of pure statistical knowledge was measured with 12 (out of 18) questions from the assessment, making a total of 48 possible points for that part. The domain of knowledge for teaching was measured with 6 (out of 18) questions and 24 possible points. Prospective teachers performed better in the domain of pure statistical knowledge (65.72%) than the domain where they had to apply this knowledge to teaching (45.14%). Again, none of the prospective teachers showed mastery or near mastery level of correctness for all items in either domain of knowledge considered here.

For pure statistical knowledge, the assessment instrument also measures three different levels of performance. The level of statistical literacy was measured with 5 questions; the level of statistical reasoning with 5 questions; and the level of statistical thinking with 2 questions. As expected, prospective teachers perform better at the lowest level of performance – Statistical Literacy. This involves mainly extracting information from a graph and recognition, identification, or computation. At the higher levels, statistical reasoning and thinking, prospective teachers do progressively worse.

Item Analysis

Often more interesting and detailed information can be gleaned by careful examination of the responses item by item. Due to space considerations, we will just present the two most telling items. The first item (shown below) measures the ability of prospective teachers to identify errors in students' responses. The data and suggestions for student responses were taken from materials developed by the *Connected Mathematics Project* (Lappan et al., 2002).

Pet	Frequency
bird	2
cat	4
cow	2
dog	7
duck	1
fish	2
goat	1
horse	3
rabbit	3

Item 1

One middle school class generated data about their pets shown below. Students were talking about the data and one said:

" The mode is dogs, the median is duck, and the range is 1 to 7."

If you think the student is right, explain why.

If you think the student is wrong, identify the mistake(s).

For this item no one received a top score, which corresponded to correctly identifying that the student is right when he talks about the mode, but for categorical data the definition of measures of center and spread do not apply in the same way as they do for numerical data. Only one person got close to a correct response (score of 3) by saying:

> The data is showing how many people have each type of pet. When we refer to measures of central tendency, we are using numerical data. … In this case dog is the most frequent choice, but it has no numerical value….We cannot order non-numerical data. …There are no data values to subtract, you do not use frequency to find the range, and if you did the range would be 7-1 = 6.

Two others (score = 2) showed some awareness of the type of data and mentioned something about not making sense of measures of center in this case. For example, one prospective teacher wrote:

> The student is right on the mode & the range but wrong on the median part. Median can't be duck. It has to be a number. Median is the average # of animals in each type of pet that a one middle school class should have.

The rest of the responses split into those that recognized that the student is correct in saying that the mode is "dog", but incorrectly judged the rest of the student's statement receiving a score of one (21 out of 42) and those that incorrectly judged the entire statement, receiving a score of zero (18 out of 42). All of those with a score of one said that the median is not "duck", and proposed an alternative value for the median. Three principle strategies were used to compute the median. The first group treated the frequencies as data observations and found the median of the frequencies yielding a median of 2. The second strategy, ordering the pet names by frequency and choose the middle name, yielded a median of "fish". When one of these prospective teachers was asked in the interview "what does this tell you about the pets when you said that the median is 'fish'? She responded, "that the average student in the class have fish [sic]". The final strategy involved listing out all 25 observations and choosing the 13th ordered value. The observations were either ordered by frequency (median = "rabbit") or alphabetically (median = "dog"). As one respondent put it *"...the student incorrectly stated that the median is duck, because he or she failed to place the data in order of increasing frequency first. Once the data is arrange [sic] in the correct order, it is clear that the median is rabbit."* One who opted for alphabetical ordering recognized the difficulty in placing an ordering on the categorical variable pet, writing *"...because when finding the median, you're supposed to put it in order from smallest to largest and when it comes to animals how to judge which is larger or smaller. There are dogs smaller than ducks, ducks smaller than cat and vice versa."*

The second item, which uses graphical displays also developed by the *Connected Mathematics Project* (Lappan et al., 2002), measures the ability to compute the mean from a line plot. Secondarily this item also assesses graph comprehension and the concept of the mean beyond the algorithm. Parts b and c measure properties of the mean and the ability to create distinct distributions with the same mean.

Item 2 The following line plot shows the number of people in households in a neighborhood.

Number of People in Households

a. Find the mean. Show how you find it.

b. Is it possible to have other sets of data with the same mean? Explain why or why not.

c. Is it possible to have a data set of six households with mean $3\frac{1}{2}$ people? If yes, give an example. If not, explain why.

```
                    X           X
              X     X     X     X
        _____
        1     2     3     4     5  6  7
```

For part a, most prospective teachers scored very high. About two thirds of the prospective teachers correctly found the mean and wrote a valid mathematical equation to justify their answer. Another 16% successfully identified the data values and the number of data points but made a serious notational error of equating the value of the sum to the ratio and therefore to the mean. All of the prospective teachers who found the mean used the mathematical algorithm. Only one person attempted to use the idea of center and unfortunately he used it incorrectly, saying, "using the idea of 'cluster' point associated with the mean, the mean is 3, for the data set above." In follow up interviews we tried to ascertain if they could estimate the mean for a line plot of a larger data set without reaching for the algorithm. One prospective teacher had a method to estimate the median from a line plot but struggled to find a way to estimate the mean for a larger data set. In the attempt to estimate the mean, she "moves Xs around" and tried to relate the concept of balance but ended up finding the average of the frequencies instead.

Parts b and c were designed to measure the ability of the prospective teacher to argue or explain why it is possible to have different distributions with the same mean and to create distributions with a particular mean. About half of the prospective teachers gave a successful response by providing a mathematical argument, or giving an example of different distributions with equal means in the context of the problem. However, only a third of those provided the ideal response of connecting the algorithm with the concept of the mean as a balance point of a distribution. A typical argument was, "*Yes because the average of numbers can be the same depending on what # you are dividing by and what #'s you are adding.*" A smaller portion of prospective teachers (about 12%) gave arguments in relation to the context of the problem – people in households – to explain the possibility of having another data set with the same mean. A third category of responses is one that talks about distribution, sample size, or variation. However, these responses were too vague to be accepted as valid arguments. One example was, "*Yes, the distribution can be proportionately spread out.*"

As for creating a data set with a specific mean, about 70% of the prospective teachers were able to create a correct data set with the specified mean and number of data points. About half of those chose a symmetric distribution about the point 3.5, and the other half chose a non-symmetric distribution but with clear indication that they created it by finding the sum of the data values first. A small percentage of prospective teachers (about 15%) argued that it was not possible to have six households with a mean of 3.5 people because when it comes to people, it does not make sense to have a non-integer value. In interviews, respondents were asked in addition what it means to have a mean of 3.5 people per household and how they would explain it to a child. In some cases this revealed a lack of depth of understanding. One particular exchange was:

> *Int.:* What does and average of 3.5 people per household mean? How do we interpret this number?
> C: I guess when the average…many many houses average about 3.5. I don't know what that means, I don't know how to explain what that means though. Because the kids are gonna go "what is 3.5 of a person?!"

Conclusion

A review of NCTM and state standards revealed a variety of aspects covered but with a general emphasis on graphical representation and measures of center. Standards for teacher preparation and certification emphasize teaching strategies. Led by these results and conscious of the limitations of a written instrument, we chose to assess pure statistical knowledge, knowledge for teaching, and knowledge of students as learners. In terms of pure statistical knowledge, we see that prospective teachers are able to compute and estimate the mean of a small data set represented in a line plot. When given a bigger data set and asked to estimate the mean, they either estimate the median without observing the distribution or in their attempt to find the average; they find the average of the frequencies instead. The latter seems to be a reflection of the methodology taught in elementary and middle school of leveling off stacks of cubes. Prospective teachers know that it is possible to have many sets of data with the same mean. However, only a small percentage can justify it with an argument that relies on both the algorithm and the concept of the mean as a balance point. A smaller percentage of prospective teachers are able to create a distribution with a specific mean that is not a whole number. The majority of these prospective teachers are bothered by having a non-whole number as an average. The observed difficulties in the aspect of statistical knowledge for teaching are more pronounced. They can not recognize students' mistakes about median and range for categorical data. In this particular case, prospective teachers were very creative inventing their own methods to have a median for categorical data. Prospective teachers do not think of what the measures of center and of spread tell you about the data when trying to find them. Instead they reach for a procedural method to get an answer. They could neither estimate the mean of a data set without reaching for the algorithm, nor convince a child that there could be several data sets with the same mean, nor explain what an average of 3.5 people means.

REFERENCES

Bright, G. W., & Friel, S. N. (1998). Graphical representations: Helping students interpret data. In S.P. Lajoie (Ed.), *Reflections on Statistics* (pp. 63-117). Mahwah, New Jersey: Lawrence Erlbaum.

delMas, R. (2002). Statistical Literacy, Reasoning, and Learning: A Commentary. *Journal of Statistics Education* [Online], 10(3). (www.amstat.org/publications/jse/v10n3/delmas_discussion.html).

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications. *Journal for Research in Mathematics Education*, 32(2), 124-158.

Friel, S. N., Bright, G.W., Frierson, D., & Kader, G. D. (1997). A Framework for Assessing Knowledge and Learning in Statistics (K-8). In I. Gal & J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 55 – 63). IOS Press.

Garfield J. B. (2002). The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education* [Online], 10(3). (www.amstat.org/publications/jse/v10n3/delmas_discussion.html).

Garfield J. B. (1993). An Authentic Assessment of Students' Statistical Knowledge. In N. L. Webb & A. F. Coxford (Eds.) *Assessment in the Mathematics Classroom*, 1993 Yearbook of the National Council of Teacher of Mathematics (NCTM), (pp. 187-96). Reston, VA: NCMT, 1993.

Lui, H. J. (1998). *A cross-cultural study of sex differences in statistical reasoning for college students in Taiwan and the United States*. Doctoral dissertation, University of Minnesota, Minneapolis.

Middle School Mathematics Praxis II Test, ftp://ftp.ets.org/pub/tandl/0069.pdf

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

National Research Council (2001). *Knowing and Learning Mathematics for Teaching*. *Proceedings of a Workshop*. Washington, D. C.: National Academy Press.

Porter, Andrew C. (2002). Measuring the Content of Instruction: Uses in Research and Practice. *Educational Researcher*, 31(7), 3 – 14.

Senk, S., Steven Viktora, Zalman Usiskin, Nils Ahbel, Virginia Highstone, David Witonsky, Rheta Rubenstein, James Schultz, Margaret Hackworth, John McConell, Dora Aksoy, James Flanders, & Barry Kissane (1998). *Function, Statistics, and Trigonometry*. The University of Chicago School Mathematics Proyect. Second Edition. Scott Foresman Addison Wesley.

The Mathematical Education of Teachers (2001). Published by the American Mathematical Society.

TIMSS 1999 Mathematics Items, http://timss.bc.edu/timss1999i/pdf/t99math_items.pdf

Thompson D. & Senk S. (1998). Using rubrics in high school. *Mathematics Teacher*. December.

Thompson D. & Senk S. (1993). Assessing Reasoning and Proof in High School. In N. L. Webb & A. F. Coxford (Eds.) *Assessment in the Mathematics Classroom*, 1993 Yearbook of the National Council of Teacher of Mathematics (NCTM), (pp. 167-76). Reston, VA: NCMT, 1993.

Watson, J. M. (1997). Assessing Statistical Thinking Using the Media. In I. Gal & J.B. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 107-121). IOS Press.