Using Work Samples from the National Assessment of Educational Progress (NAEP) to Design Tasks that Assess Statistical Knowledge for Teaching

Randall E. Groth
Salisbury University

Key Words: Teacher preparation; Pedagogical content knowledge; Subject matter knowledge; Classroom example.

Abstract

Statistics courses for prospective teachers are being developed in response to recent K-12 curriculum recommendations. As these courses are developed, it is important to design accompanying assessments. This manuscript describes a strategy for assessing aspects of statistical knowledge for teaching. The strategy involves analyzing and responding to children's work samples from the National Assessment of Educational Progress. An example is provided to illustrate how such an assessment task can serve to reveal different aspects of respondents' subject matter knowledge as well as their pedagogical content knowledge. Thoughts are provided on how to use the assessment tasks to identify levels of statistical knowledge for teaching.

1. Introduction

There is currently a pressing need to prepare practicing K-12 teachers and undergraduates in teacher education programs to teach statistics. Over the past century, statistics has steadily risen from obscurity to prominence in K-12 curricula (Jones and Tarr 2010). Its rise to prominence has been accelerated in recent decades by curriculum documents such as *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics (NCTM) 1989), *Principles and Standards for School Mathematics* (NCTM 2000), and the *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) *Report* (Franklin et al. 2007). These documents were influential in decisions to include statistics in state standards in the U.S. For example, the *Common Core State Standards* (Common Core State Standards Initiative 2010)

include a "measurement and data" strand in elementary school and a "statistics and probability" strand in middle and high school. Similar emphasis on probability and statistics in early grade levels has become an international trend (Jones, Langrall, and Mooney 2007).

Although increased focus on statistics in K-12 is a potentially exciting development, it comes with steep challenges for teacher education. Many prospective teachers have not had opportunities to study statistics in the manner envisioned by reform documents during their pre-college experiences (Conference Board of the Mathematical Sciences (CBMS) 2001). Hence, in many cases, they exhibit the same types of difficulties in statistical reasoning as the students they are to teach (Groth 2007). Additionally, teachers must not only understand statistics content for themselves, but they must develop the ability to make it comprehensible to children. CBMS (2012) recommended that courses intentionally designed to foster this ability be included in teacher preparation programs. Recent policy documents on teacher preparation in statistics support and amplify this recommendation (CUPM in preparation; Franklin et al. in preparation).

Courses designed to prepare teachers according to the vision of NCTM, CBMS, and GAISE have been developed. For example, Metz (2010) used the GAISE guidelines and the NCTM standards to design a statistics course for prospective teachers. Course activities focused in-depth on the statistics content recommended in these curriculum documents. Metz reported that participants generally found it to be more valuable than a conventional introductory statistics course. In another statistics course designed to meet prospective teachers' learning needs, Groth (2012) found that participants experienced substantive increases in both conceptual knowledge and the knowledge needed to teach statistics to children. These gains came about within the context of a course that emphasized analyzing professional articles about children's statistical thinking and teaching strategies for helping it develop. Similarly, a course designed by Green and Blankenship (2013) supported prospective teachers' development of statistical content knowledge and pedagogical content knowledge. During the course, Green and Blankenship not only addressed statistical concepts, but also discussed and modeled how to teach and assess children's understanding of the concepts. Such reports suggest that designing and testing statistics courses for teachers is a promising avenue for continued research.

As a new and emerging area of study, the development of statistics courses for teachers faces a question similar to that faced by reform-oriented statistics programs in the 1990s (Gal and Garfield 1997): How can learning outcomes connected with the courses be assessed? In this article, my goal is to begin to address this question by offering a sample assessment structure that uses children's responses to tasks from the National Assessment of Educational Progress (NAEP). To set the stage, I begin with a brief discussion of the types of statistical knowledge teachers need to develop. I then explain how work samples from NAEP can be used as part of a strategy for assessing the development of statistical knowledge for teaching.

2. The Nature of Statistical Knowledge for Teaching

Statistical knowledge for teaching can be characterized as consisting of both subject matter knowledge and pedagogical content knowledge (Groth 2013). It is imperative for teachers to develop both knowledge aspects, as they are not only responsible for learning subject matter, but also for making it comprehensible to others (Shulman 1987).

2.1 Subject Matter Knowledge

Subject matter knowledge can be thought of as consisting of common content knowledge, specialized content knowledge, and horizon knowledge (Groth 2013; Hill, Ball, and Schilling 2008). Common content knowledge pertains to what one would expect to learn in a conventional introductory statistics course that serves several different majors simultaneously, since it is knowledge that is needed for other professions as well as teaching. It would include, for example, knowledge of how to compute descriptive statistics and design statistical studies (Groth 2007). Specialized knowledge, on the other hand, is specifically required of teachers. It encompasses knowledge of representations that make content understandable to students and the ability to appraise non-conventional student strategies. For example, knowing how to represent the mean as a typical value, a fair share, and a signal amid noise could be indicative of specialized knowledge (Wassong and Biehler 2010). The ability to analyze a child's statistically naïve interpretation of survey results also exemplifies specialized knowledge (Burgess 2011). Horizon knowledge relates to content beyond that which the teacher is responsible for teaching. It enables teachers to take a broad view of the disciplines they teach. Having a broad view can provide teachers a sense of where the curriculum ultimately leads and thus help them scaffold students' learning in productive directions. For instance, it is useful for teachers who are leading children in their initial study of nominal categorical data to realize that in later years the children will encounter ordinal categorical data. Having this type of horizon knowledge may help teachers avoid leading children toward overgeneralizations such as, "categorical data can never be ordered" (Groth and Bergner 2013).

2.2 Pedagogical Content Knowledge

Three aspects of pedagogical content knowledge are knowledge of content and students, knowledge of content and teaching, and curriculum knowledge (Groth 2013; Hill, Ball, and Schilling 2008). Knowledge of content and students allows teachers to anticipate patterns of thinking students are likely to exhibit in regard to content. One well-documented pattern of thinking is that children may view the mean merely as an algorithm without understanding it can be used to signify the typical value of a data set (Mokros and Russell 1995). Teachers who know of such patterns of thinking are positioned to address them during instruction. Knowledge of content and teaching provides a repertoire of strategies for teaching specific content. For example, knowledge of content and teaching can help teachers choose activities and tasks likely to help students understand mean and median as one-number summaries of central clusters of data rather than just algorithms to be computed (Konold and Higgins 2003). Curriculum knowledge can include familiarity with appropriate sequences for the introduction of statistical ideas (Godino, Ortiz, Roa, and Wilhelmi 2011), such as when different types of statistical representations are best introduced (Groth 2013).

The need to develop pedagogical content knowledge is one of the things setting statistics preparation for teachers apart from the statistical preparation needed for other professions. The statistics reform movement of the 1990s helped create courses and assessments focused on conceptual understanding rather than the mainly procedural approach that had been prevalent (Cobb 2013). The learning materials (Scheaffer, Gnanadesikan, Watkins, and Witmer 1996;

Garfield and Ben-Zvi 2008) and assessments (delMas, Garfield, Ooms, and Chance 2007) that have come about as a result have value for building and assessing teachers' conceptual understanding. However, conceptual understanding of statistics is within the domain of subject matter knowledge and not equivalent to pedagogical content knowledge. Having conceptual understanding does not guarantee that one will also know how children think about statistics or effective strategies for helping children's statistical thinking develop. Hence, there is a need to assess the pedagogical content knowledge that teachers develop during courses designed to build statistical knowledge for teaching. In some cases, pedagogical content knowledge and subject matter knowledge may be assessed within the context of the same item. An example of a task assessing both aspects of statistical knowledge for teaching is provided in section 4.

3. Assessment Task Design Principles and Procedure

One approach researchers have used to assess content knowledge for teaching is to have prospective and practicing teachers analyze and respond to children's thinking. For example, some assessments present common misconceptions and patterns of thinking from children and ask respondents to comment on children's work and describe ways to help their thinking develop (e.g., Hill, Ball, and Schilling 2008; McCrory, Floden, Ferrini-Mundy, Reckase, and Senk 2012; Saderholm, Ronau, Brown, and Collins 2010). A well-known example of this sort of assessment item comes from the work of Liping Ma (1999). Ma presented a classroom scenario in which a child told a teacher "that she has discovered that as the perimeter of a closed figure increases, the area also increases" (p. 84). The child then presented a 4×4 rectangle and a 8×4 rectangle as support for the conjecture. Some of the teachers in Ma's study were uncertain how to respond to the child because they did not know for themselves if the conjecture was true or not. Others were able to blend their knowledge of content and pedagogy to ask probing questions to lead the student to discover a counterexample to the claim (e.g., a 12×1 rectangle has a greater perimeter, but less area). Hence, this type of item has the potential to reveal qualitative variability in individuals' content knowledge for teaching.

In order to design items that assess statistical knowledge for teaching, it is helpful to have access to children's work samples. One readily accessible source of such work samples is the National Assessment of Educational Progress (NAEP) Questions Tool website (http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=mathematics). Figure 1 is a screenshot from the website. It shows a search for constructed response items on NAEP tests taken by eighth-graders. Under the "Select Content Classifications" tab, "Data Analysis and Probability" is checked. Under the "Select Years" tab, the years 2009, 2011, and 2013 are selected. These filters produced a collection of 7 items. Clicking on the "Add All Questions" button on the screen put the items in the "My Workspace" area. From "My Workspace," one can request the questions, answers, student responses, and performance summary data by clicking the appropriate boxes (Figure 2). One can then download all of these materials in HTML or Word by clicking on the desired format.

Figure 1. NAEP Questions Tool Search Results Screen

Figure 2. NAEP Questions Tool Workspace Download Screen

After downloading NAEP tasks and accompanying work samples, the process of designing a statistical knowledge for teaching task involves initial considerations such as: (i) Which statistical content tends to be difficult to teach in grades K-12?; and (ii) Which questions have content that matches objectives of the course in which prospective teachers to be assessed are enrolled? Consideration (i) can be addressed by scanning the "Difficulty" column in "My Workspace" (Figure 2). It shows, at a glance, which items were "easy," "medium," and "hard"

for a representative sample of U.S. students from the selected grade levels. If one includes "Performance Summary Data" in the HTML or Word material download, more information about the difficulty of the items is provided. Users are given the percentage of students scoring at the different levels specified in each task rubric. Interesting information about statistical knowledge for teaching may be produced in connection with items containing content that is generally difficult for students to learn, since this content may also be challenging to teach. Consideration (ii) must be addressed on a course-by-course basis. However, it is worth noting that looking at the statistical content that is generally difficult to teach may impact the goals one has set for the course. An instructor might decide, for example, to shift course goals so that difficult-to-teach statistics content receives greater emphasis. The opportunity to reflect on course goals in such a manner is a collateral benefit of engaging in this task design process.

Appendices A, B, and C contain examples of assessment tasks I designed using NAEP items and work samples. I assigned tasks like these as written homework in an undergraduate course I taught for prospective K-8 teachers that focused on the development of statistical knowledge for teaching. For each assessment task, I showed a NAEP item and a sample of children's responses. I then asked prospective teachers to respond to the following prompts:

- Organize the responses from "most developed" to "least developed." You can use as many categories to organize the responses as you wish. Explain why you categorized the students' responses in the manner you chose.
- Describe the steps you would take to further support each student's learning.

The first of the two prompts above mainly assessed respondents' subject matter knowledge, since they had to judge the statistical quality of each given response. The second of the two prompts was intended to assess knowledge of content and teaching, since respondents would have to suggest content-specific teaching strategies to help children progress beyond the types of thinking exhibited in the work samples. In the next section, I describe what I learned about prospective teachers' statistical knowledge for teaching from a NAEP-based task involving mean and median (the entire NAEP-based task is shown in Appendix A).

4. Assessment Task Example: Analyzing and Responding to Children's Thinking about Mean and Median

The task shown in Figure 3 was administered to eighth-graders on the 2007 NAEP. The NAEP rubric for the task stated that for a response to be considered correct, children must choose the median and state that the mean is affected by the data from day 4, when only 10 customers were present. In the U.S., 67% of children gave incorrect responses, 21% gave partial responses, and 6% omitted the item. Only 6% of children gave a correct response.

Figure 3. NAEP Item on Mean and Median (Source: Eighth-grade NAEP test item, block: 2007-8M9, number 8 (NAEP 2007); National Assessment of Educational Progress)

The table below shows the number of customers at Malcolm's Bike Shop for 5 days, as well as the mean (average) and the median number of customers for these 5 days.

Number of Customers at Malcolm's Bike Shop	
Day 1	100
Day 2	87
Day 3	90
Day 4	10
Day 5	91
Mean (average)	75.6
Median	90

Which statistic, the mean or the median, best represents the typical number of customers at Malcolm's Bike Shop for these 5 days?

Explain your reasoning.

Given children's difficulties with the NAEP item, in my course for prospective teachers, we explicitly focused on thinking processes involved in choosing between the mean and median when describing the typical value of a data set. We used *Fathom* and *TinkerPlots* to observe how the mean can change dramatically because of the outliers in some data sets while the median stays the same. We also discussed a similar item from the 1996 NAEP in which children were asked to choose between mean and median to describe the typical daily attendance at a movie theatre. We considered how having children analyze different kinds of data sets containing outliers may help them choose between mean and median in such situations.

4.1 Prompt 1: Analyzing Children's Thinking

When I asked prospective teachers to organize children's responses (Figure 4) from "most developed" to "least developed," most were able to provide a reasonable rationale for ranking the responses in order of sophistication. Of the 20 who responded to the item, 17 categorized the response of student 2 as being the most developed because of the attention given to the extreme value. These 17 all categorized student 1's response as next in order of sophistication, since the student indicated the median but did not provide a detailed explanation. Student 3's response came next, since it used the mean and provided little justification for doing so.

Figure 4. Children's Work Samples from NAEP Item on Mean and Median (Source: Eighth-grade NAEP test item, block: 2007-8M9, number 8 (NAEP 2007);
National Assessment of Educational Progress)

Student 1

> The Median because on a topical day the Bike Shop had normally 90 customers

Student 2

> Median
> The median better represents the statistic in this question because on day 4, Malcom only had 10 customers, which is an outlier and greatly lowered the average

Student 3

> The mean Because its the average of all the days.

The 3 prospective teachers who did not follow the categorization scheme of the other 17 had questionable rationale for doing do. Becky (a pseudonym, as are the rest of the prospective teachers' names in this article) believed that student 3 clearly understood the function of the mean and how to compute it (even though the item did not require the computation of the mean). As a result, she considered student 3's response to be more developed than student 2's. Lisa put student 3's response ahead of student 2's response simply because student 3 wrote more. This response reflected lack of attention to the statistical content of the item. Monica put student 3 ahead of student 2 because she believed that the mean is actually a more suitable measure than the median when outliers exist.

In sum, reasonable categorizations of children's work samples reflected the ability to coordinate statistical content knowledge with attention to children's thinking. Most of the prospective teachers were able to do so successfully, demonstrating sound subject matter knowledge for the given task. Some attempted to coordinate the two areas, but had difficulties in reasoning about statistical content (as in Monica's response) or about children's thinking (as in Becky's response). Lisa's response did not attend to relevant aspects of children's thinking or to statistical content at all. The various levels of response to this first part of the assessment task are summarized in the color gradient diagram shown in Figure 5.

Figure 5. Levels of Response to NAEP-Based Assessment Task on Mean and Median

Response shows lack of attention to statistical content or to children's thinking	Response shows attention to statistical content and to children's thinking, but analysis of one of the two areas is flawed	Response shows coordination of statistical content knowledge and attention to children's thinking

Although prompt 1 primarily assessed respondents' subject matter knowledge, it is worth noting the potential of the item to help *develop* pedagogical content knowledge as well. Specifically, as respondents analyze the work samples, they have the opportunity to engage with children's reasoning about mean and median. This engagement can make them aware of some children's tendencies to ignore outlying values when choosing a measure to describe the typical value in a data set. Developing this aspect of knowledge of content and students may then lead respondents to address this particular student difficulty when choosing teaching strategies and designing lessons.

4.2 Prompt 2: Supporting Children's Learning

The second assessment prompt was designed to assess pedagogical content knowledge, as it asked how the children's learning of mean and median could be supported. In analyzing responses to the prompt and categorizing them, I kept three principles of creating effective learning environments identified by Fuson, Kalchman, and Bransford (2005) in mind: (i) teachers should elicit and address student' preconceptions, (ii) lessons should aim to build both conceptual and procedural knowledge (as opposed to focusing solely on practicing procedures and terminology), and (iii) instruction should support metacognition (i.e., prompting children to explain and think about their own thinking). Responses reflecting these principles were judged to be of higher quality than those that did not. As a consequence, responses that suggested engaging children in tasks with higher levels of cognitive demand (Henningsen and Stein 1997) were considered higher quality than those following the more common, and less engaging, classroom pattern of learning rules and practicing procedures (Stigler and Hiebert 1999).

The second assessment prompt elicited a wide array of responses. Most prospective teachers suggested individualized instructional strategies to address the learning needs of the children whose work samples are shown in Figure 4. Others suggested a single type of instructional strategy for all three children whose work samples they analyzed.

4.2.1 Types of Individualized Support Suggested for Student 1

Three prospective teachers merely described what they believed Student 1 could do to improve without offering any specific pedagogical strategies. Jamie, for example, stated, "Student 1 needs to mention the outlier, because that greatly explains why the median would be a much better choice than the mean." Although such a response highlights an important part of the data set for children to notice, it does not offer thoughts on how to draw children's attention to it. This suggests only partially developed knowledge of content and teaching, since a relevant aspect of content is described, but no relevant teaching strategies are offered. Because of the nature of the

assessment item, however, the possibility that some respondents only paid partial attention to the requirements of the prompt (but actually did have more developed knowledge of content and teaching) cannot be ruled out.

Seven prospective teachers took the approach of asking Student 1 to explain in more detail. Samantha, for example, stated,

> To further support student 1's learning I would ask them why they think the mean would not be the best answer because they were able to know the median would be the best option but did not explain why the mean would not work

Similarly, Emily wrote, "I would have him/her describe and analyze the set of data more thoroughly. He/she has a clear understanding of median and how to determine it. He/she however, did not provide and explanation like it asked for in the question." These types of responses have both strengths and weaknesses. Asking children to explain themselves in detail can provide teachers with insights helpful for adjusting instruction to suit their learning needs. However, it is unclear how children's learning would be supported if they lack the knowledge to expand upon their original answers. In this situation, it is not apparent how prospective teachers like Samantha and Emily would deal with Student 1 if he or she was not able to provide more statistically relevant details when expanding on the original response.

Three prospective teachers suggested pointing Student 1 directly toward the outlier in the data set. Tony's response is illustrative:

> With student one I would encourage them to expand their reasoning to more than mean vs. median and point them in the direction of outliers and the effect that outliers can have on mean and why in this situation they play such a big role.

Such a teaching strategy would likely focus children's attention on relevant aspects of the data set for the NAEP problem. However, it comes with a risk of removing the challenge of the problem for children by lowering its cognitive demand. When the demand of a task of this nature is lowered, it ceases to be an opportunity to learn through problem-solving.

Two prospective teachers proposed strategies that would help preserve much of the demand of the task while still potentially helping Student 1 notice relevant aspects of the data set. Becky suggested having Student 1 work problems similar in structure to the original NAEP item. After working the additional items, Becky suggested having Student 1 discuss his/her reasoning in detail. In such a teaching scenario, the similarity in structure among the items may help a child notice the effects of the outlier in each case. Scott took a slightly different teaching approach. He suggested asking Student 1 to explain "why he thinks the mean is so low when compared to most of the days." This could potentially draw Student 1's attention toward the central cluster of data and its position relative to the mean and median.

Figure 6 summarizes prospective teachers' ideas for supporting Student 1's learning. Darker shades in the diagram represent higher levels of knowledge of content and teaching. Some prospective teachers (e.g., Jamie) were able to identify a flaw in the child's work but did not suggest any teaching strategies to remedy it. Others (e.g., Samantha) recognized problems with the child's work, but did not suggest teaching strategies beyond asking for further explanation of the original response. Still others (e.g., Tony) would remove much of the challenge of the task by pointing the child directly toward a solution. Higher levels of knowledge of content and teaching

were apparent in responses that were likely to help the child notice relevant aspects of the data set while preserving the challenge of the original task (e.g., Scott). Preserving the challenge of a task can be considered a higher level of knowledge of content and teaching because it involves striking a balance between providing appropriate scaffolding without taking over the thinking for children. Even experienced teachers can have difficulty striking such a balance consistently (Henningsen and Stein 1997).

Figure 6. Levels of Knowledge of Content and Teaching Apparent in Individualized Support Suggested for Student 1

Response identifies a flaw in the child's thinking but suggests no teaching strategies to remedy it	Response suggests probing the child's thinking by asking for further explanation	Response includes content-specific questions or tasks likely to prompt the child to notice key aspects of the data set
	Response suggests pointing the child directly toward a key aspect of the data set	

4.2.2 Types of Individualized Support Suggested for Student 2

Prospective teachers considered Student 2's response to the NAEP item to be the most developed of the three work samples provided. Angela remarked, "There are not necessarily any steps I would take to further support this student's learning because they seem to have no problem answering and explaining the answer to this question." Like Angela, Jamie did not see a need for follow-up instruction. She stated, "For Student 2, I think he pretty much gives a good explanation, but they could mention the other numbers in the data to see how great of an outlier that 10 is." The responses of Angela and Jamie were not typical, however, as other prospective teachers offered teaching strategies to help extend Student 2's thinking.

One type of proposed teaching strategy to address Student 2's needs was offered by Brenda. She wrote, "From this response the student has a good grasp of the concept and some independent work will allow them to practice this ability." Her suggestion to provide practice problems for Student 2 may have some potential to solidify the student's thinking about choosing between mean and median. Nonetheless, Brenda's response could be indicative of relatively underdeveloped knowledge of content and teaching. Independent practice alone may help children retain what they have learned, but such practice does not necessarily lend itself to the extension and deepening of statistical knowledge.

Four prospective teachers acknowledged that independent practice on the same types of problems alone would not be enough for Student 2. Emily, for example, wrote, "For Student 2, I would challenge him/her because he/she has a clear understanding of the differences between mean and median and are able to determine which one better represents a set of data." Similarly, Annie stated, "Perhaps the teacher should give this student more challenging work." Although these responses affirmed the importance of extending Student 2's learning, they did not provide specific content-based strategies or tasks for doing so.

Another four prospective teachers believed it would be valuable to extend Student 2's thinking by probing for additional explanation. Two of these responses incorporated general probes, including, "explain a little more" and "add a little more information." The other two were more specific to the data set. Mary, for example, wrote, "I would ask why they believe 10 was so significant in the data. Then ask why mean would not be the correct answer." Shelly decided to probe Student 2's thinking by asking, "How do you know 90 is a typical day? How far away is this number from the rest of the data?" The degree of specificity of a prompt is relevant to the assessment of knowledge of content and teaching, as the type of children's thinking elicited by vague, general prompts will likely not be the same as that elicited by a prompt that is more specific to the data set.

Some prospective teachers decided to alter the structure of the NAEP item in order to provide Student 2 an additional challenge. One strategy for doing so, used by two of the prospective teachers, was to remove the outlier from the data set to see how Student 2's response to the item would change. Another strategy was to actually add outliers to the data set. Samantha suggested such a strategy, stating,

> To further student 2's learning I would ask them if there were two outliers, one greatly lower than the rest of the numbers and one greatly higher than the rest of the numbers, would they still think the median is the best option?

Samantha's proposed alteration of the data set had the potential to push Student 2 to consider a situation with outliers where the mean may actually be an appropriate choice, thus extending Student 2's range of experiences with data. Scott proposed a different type of alteration to the data set. He wrote,

> The one thing I could do is maybe extend the data to 10 days and have him look at it again and see if his answer changes to show how with the more data you have, the less outliers affect the mean.

As with Samantha's strategy, Scott's strategy was designed to help extend students' intuitions about the relationship between outliers and the mean. Because they attended closely to content and to teaching strategies that could feasibly extend students' thinking, these types of responses reflected more fully developed knowledge of content and teaching.

The prospective teachers' responses about how to support Student 2's learning stand out because of the amount of variation in knowledge of content and teaching they suggested (Figure 7). Some considered the child's response to be fine as it stood and did not suggest additional challenge or support. Others suggested the traditional strategy of providing practice problems in an attempt to solidify what the child already knew. Some went beyond simple practice problems to acknowledge a need for more challenging work or deeper explanation of the original task. A few were able to suggest content-specific questions to prompt the child to think and explain further about the original task. Finally, some prospective teachers demonstrated the ability to modify the original task in ways that would be likely to extend the child's learning and prompt them toward reflective thought. The last two types of responses are indicative of higher levels of knowledge of content and teaching in that they are likely to engage children in metacognitive activity (i.e., thinking about their thinking). Metacognition has consistently been demonstrated to be an important factor in building new knowledge (Bransford, Brown, and Cocking 1999; Schoenfeld 2007).

Figure 7. Levels of Knowledge of Content and Teaching Apparent in Individualized Support Suggested for Student 2

No additional support or challenge for the child is suggested	Additional practice problems to help retain knowledge are suggested	Response acknowledges the need for more challenging work for the child	Content-specific probing questions are suggested to prompt the child to explain further	The response suggests altering the original task given to the child in order to enhance the challenge of the task and extend the child's thinking
		General probing questions are suggested to prompt the child to explain further		

4.2.3 Types of Individualized Support Suggested for Student 3

Student 3 selected the mean in response to the NAEP item, noting that it was the "average" of all of the days (see Figure 4). It should be noted that the structure of the NAEP item itself may have led Student 3 to this decision, since mean was labeled as "average" in the statement of the item. Most of the prospective teachers responding to the item provided specific suggestions for teaching strategies intended to help Student 3. There was just one case in which a teaching strategy was not suggested. In her response, Jamie simply observed, "Student 3 needs to examine the data more and think critically about the numbers in the data." She did not, however, suggest any teaching strategies to help Student 3 to attain the desired outcomes she identified.

In seven cases, prospective teachers decided to help Student 3 by directly explaining how to do the problem. Samantha's response was typical of this type of approach:

> To further student 3's learning I would point out to them that all of the numbers are close to each other except 10, making it an outlier. I would then explain how that outlier would affect the mean, making the median the best option. Then to make sure they understood it I would give them another problem like it to see if they can get it correct.

This type of instructional pattern is very typical in U.S. mathematics classrooms. Stigler and Hiebert (1999) observed many instances of it in their classroom video study and described it as "learning rules and practicing procedures." Such an approach lowers the level of cognitive demand of a task by imposing a teacher-prescribed strategy for solving it (Henningsen and Stein 1997). Even when the information shared with students is accurate, the approach is limited in its ability to encourage authentic problem-solving activity.

Four of the prospective teachers suggested asking questions to draw Student 3's attention to salient features of the data set. One such response was given by Mary:

> Ask the student (Student 3) to look at the data and see if they notice that 10 is very unusual for the shop. Then ask why they believe median is not correct. And point them in the right direction asking how big of an affect that 10 customers has on the data (*sic*). Is it skewed?

Although most prospective teachers' questions were designed to draw Student 3's attention to the outlier, Scott took a slightly different approach. He sought to guide students to compare the values of the mean and median to the main cluster of data, writing,

> To push this student further I would ask them if they thought the mean was close to what most days numbers were. I would then ask which was closer to most days numbers, the mean or the median. Then, after answering these two questions, I would ask them to reevaluate their original answer.

The types of responses given by Mary and Scott represent attempts to leave some of the demand of the task intact by using probing questions to help Student 3 think about the situation rather than just providing a solution strategy.

Another approach used to avoid directly telling students how to solve the task was to have Student 3 solve and analyze tasks that paralleled the original one. Three prospective teachers suggested such a strategy. Becky, for example, wrote,

> For Student 3, I would prepare examples of data sets when the mean would be appropriate and data sets when the median would be appropriate. I would then sit down with this student and discuss any differences that they see between the data sets. I would encourage the student to look for outliers. I would also make sure that this student understands that the median can be a good representation of a "typical" value.

Becky's approach contrasts with others in that it begins with student work on related tasks rather than with a teacher explanation or guiding questions. Nonetheless, the teacher still plays an important role by asking Student 3 to reflect on the work they have done on the related task. Explanations or guiding questions from the teacher might follow, but they are to occur only after students have had time to think for themselves about the given situation.

Figure 8 summarizes the types of support suggested for Student 3. Responses varied from suggesting no teaching interventions to having a sequence of tasks to help the child understand key statistical ideas. The approach of learning rules and practicing procedures can be considered a relatively low level of knowledge of content and teaching, as it has been associated with lower quality of student learning (Stigler and Hiebert 1999). Task-based approaches that help children unpack the key statistical concepts at hand can be considered indicative of higher levels of knowledge of content and teaching, since such approaches engage children in constructing their own knowledge through problem-solving. Children who are given opportunities to do so often have higher levels of achievement than those who do not (Boaler 1998; Carpenter, Fennema, Peterson, Chiang, and Loef 1989; Pesek and Kirshner 2000).

Figure 8. Levels of Knowledge of Content and Teaching Apparent in Individualized Support Suggested for Student 3

No teaching intervention is proposed	Response proposes a direct explanation approach consisting of learning rules and practicing procedures	Content-specific probing questions are suggested to prompt the child toward relevant features of data set for the given task	A sequence of tasks designed to help the child understand underlying statistical ideas is proposed, with teacher prompts to help the child reflect on the structures of the tasks

4.2.4 Teaching Strategies Involving Non-Individualized Support

Five prospective teachers did not make a distinction in how they would deal with Students 1, 2, and 3. They did, however, make use of many of the same teaching strategies as their counterparts who did make such a distinction. For example, Isabella adopted the direct approach of simply telling students how and why they should select the median as the response to the NAEP task. Maureen suggested posing a series of problems similar to the NAEP item and then drawing children's attention toward the outliers and distribution of the data. Patty adopted a similar strategy and added the element of emphasizing to children that the median can be considered a representative measure for some data sets. These types of responses differed from most in that prospective teachers positioned themselves as addressing the entire class simultaneously rather than individualizing instruction to match specific observed learning needs. Although there is some value in the approaches Maureen and Patty suggested, teaching tends to reach its full potential for fostering children's growth when attention to addressing individual learning needs is present (Tomlinson 2005).

5. What can NAEP-Based Tasks Reveal about Individuals' Statistical Knowledge for Teaching?

Looking back across all of the prospective teachers' responses to the NAEP-based assessment item on mean and median, a few key factors stand out to distinguish lower levels of statistical knowledge for teaching from higher levels. Responses at higher levels:

- (i) Reflected the ability to attend to both content and pedagogy simultaneously
- (ii) Maintained or extended the challenge of the original task
- (iii) Provided opportunities for children to construct their own knowledge through problem-solving
- (iv) Encouraged children's metacognitive activity
- (v) Used differentiation of instruction rather than a single teaching approach.

These five factors are not necessarily limited to the item on mean and median. They reflect general principles about teaching that can be used to assess statistical knowledge for teaching across a variety of tasks, such as those shown in Appendices B and C or other NAEP-based tasks

readers may devise by using the principles of task design outlined earlier. The five factors can serve as starting points in the assessment of responses to such tasks, and those designing pedagogical content knowledge tasks may discover additional important distinguishing factors as they analyze the empirical data generated from task responses.

The responses given to the mean and median task, taken into consideration with the five factors, also suggest somewhat of a decision-making tree for teachers as they design instruction to suit individual children's learning needs. A sketch of the decision-making tree is provided in Figure 9. Optimal paths along the tree are emphasized in bold. Although the tree is based on the mean and median task described earlier, its underlying principles are broad enough to serve as a starting point for assessing pedagogical content knowledge across a variety of tasks and assessment formats.

Figure 9. Decision-Making Tree for Supporting Children's Learning
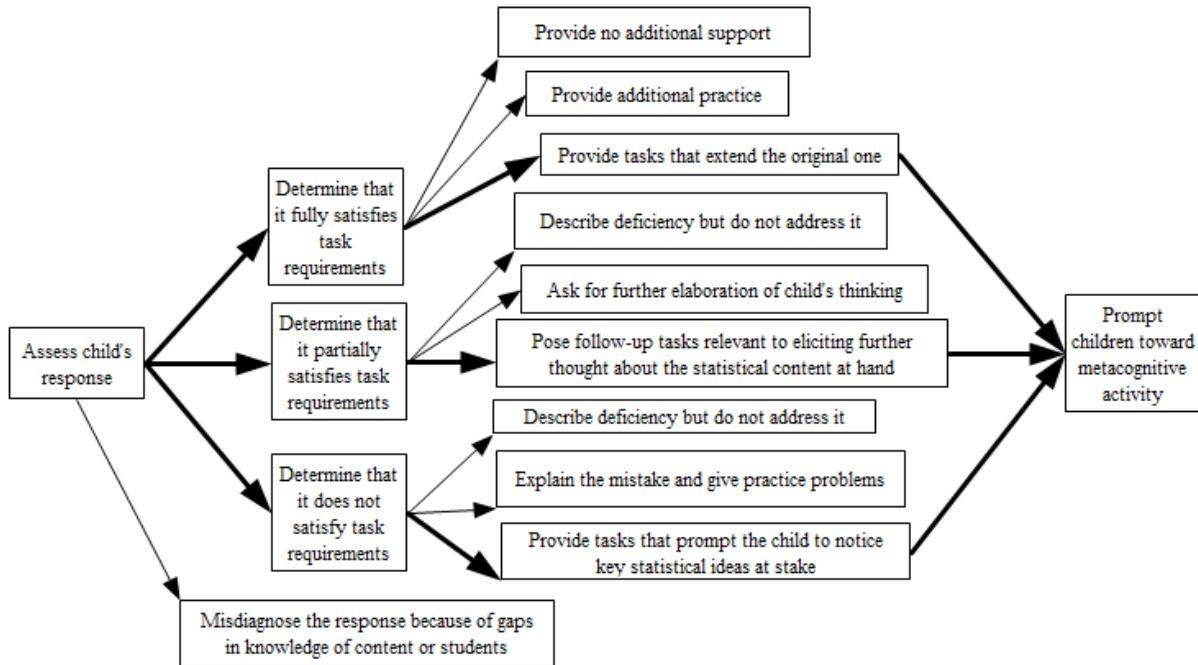
One limitation of using NAEP-based tasks such as the ones I have described is that teachers' knowledge as represented on written assessment items sometimes differs from that which they display in practice (Hill, Sleep, Lewis, and Ball 2007). Nonetheless, those who assess pedagogical content knowledge within a classroom setting rather than with written items may still find the decision tree useful as a starting framework for assessment. In fact, an interesting direction for future research would be to compare the paths prospective teachers travel along the tree when responding to written items vs. those which they travel when face-to-face in a "live" classroom setting with children. Additionally, written items have a pragmatic advantage over assessments of live teaching, in that it is generally not feasible to conduct several assessments of prospective teachers teaching children statistics over the course of a semester as part of an introductory undergraduate statistics course.

The types of written tasks I have described are also useful in that they serve not only as assessment points, but also as instructional tools. Each NAEP-based task shown in Appendices A, B, and C begins with a set of student work samples. Respondents' analyses of the work samples provide an assessment of their subject matter knowledge, but they also provide an opportunity for respondents to *build* knowledge of content and students. As prospective teachers read the work samples, they can strengthen their knowledge of content and students by becoming more aware of some of the common statistical thinking patterns children exhibit. This dual function essentially blurs the line between assessment and instruction. From a pure summative assessment/measurement perspective, this might be cause for concern, but from an instructor's perspective, it can help maximize the limited, precious resource of instructional time.

6. Conclusion

Prospective teachers' statistical learning needs are distinct from those of the general undergraduate population. Assessments must capture not only their subject matter knowledge, but also their pedagogical content knowledge. The NAEP-based tasks I have described present one means for assessing both elements of statistical knowledge for teaching. However, much work remains to be done to fully meet the need to assess statistical knowledge for teaching. I encourage readers to experiment with the tasks presented in Appendices A, B, and C in their own statistics courses for teachers. I also hope that readers will design their own statistical knowledge for teaching assessment tasks. It would be fruitful, for example, to design tasks that directly assess additional aspects of statistical knowledge for teaching, such as specialized content knowledge, horizon knowledge, knowledge of content and students, and curriculum knowledge. As more such tasks are designed and disseminated, we can build an increasingly stronger infrastructure to address the task of assessing statistical knowledge for teaching.

APPENDIX A

MEAN AND MEDIAN TASK
Analyzing Children's Thinking about Mean and Median

The following item appeared on the 2007 National Assessment of Educational Progress (NAEP):

The table below shows the number of customers at Malcolm's Bike Shop for 5 days, as well as the mean (average) and the median number of customers for these 5 days.

Number of Customers at Malcolm's Bike Shop	
Day 1	100
Day 2	87
Day 3	90
Day 4	10
Day 5	91
Mean (average)	75.6
Median	90

Which statistic, the mean or the median, best represents the typical number of customers at Malcolm's Bike Shop for these 5 days?

Explain your reasoning.

Here are some children's responses to the NAEP item:

Student 1

> The Median because on a typical day the Bike Shop had normally 90 customers

Student 2

> Median
> The median better represents the statistic in this question because on day 4, Malcom only had 10 customers, which is an outlier and greatly bured the average

Student 3

> The mean Because its the average of all the days.

Source: Eighth-grade NAEP test item, block: 2007-8M9, number 8 (NAEP 2007); National Assessment of Educational Progress.

After reading each student response presented above, do the following:
1. Organize the responses from "most developed" to "least developed." You can use as many categories to organize the responses as you wish. Explain why you categorized the students' responses in the manner you chose.
2. Describe the steps you would take to further support each student's learning.

APPENDIX B

SAMPLING TASK
Analyzing Children's Thinking about Sampling

The following item appeared on the 2013 National Assessment of Educational Progress (NAEP):

> Richard wants to estimate the average (mean) monthly temperature of the United States last year. He will choose one of the following two methods to do this.
>
> Method I: Richard selects his state and 9 other states that are near his state. Then he finds the average (mean) monthly temperature of each of those 10 states and uses those numbers to compute the average monthly temperature of the United States.
>
> Method II: Richard selects 10 different states by writing the names of all 50 states on cards, with one state's name on each card. Then he places all of the cards in a hat and takes out 10 cards without looking. Finally, he finds the average (mean) monthly temperature of each of those 10 states and uses those numbers to compute the average monthly temperature of the United States.
>
> Which method is better, I or II? Explain why.

Here are some children's responses to the NAEP item:

Student 1

Method II is better because the states he chose were random. In the first method all the states he chose were in the same region so they were have the same climate.

Student 2

Method II is better because instead of states in just one area being chosen, any state can be chosen.

Student 3

method II is better because if he chooses all the states around him then they code will all have relatively the same temp.

Student 4

Which method is better, I or II?

Method I

Explain why.

Because method I uses a sure way to find out the mean while method II is not

Source: Eighth-grade NAEP test item, block: 2013-8M6, number 13 (NAEP 2013); National Assessment of Educational Progress

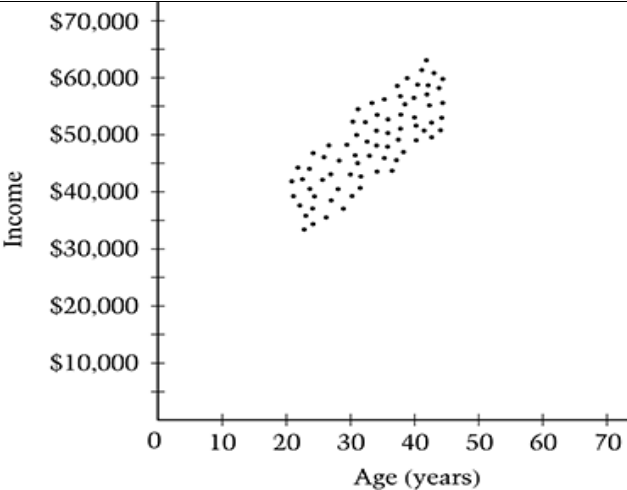After reading each student response presented above, do the following:
1. Organize the responses from "most developed" to "least developed." You can use as many categories to organize the responses as you wish. Explain why you categorized the students' responses in the manner you chose.

2. Describe the steps you would take to further support each student's learning.

APPENDIX C

BIVARIATE DATA TASK
Analyzing Students' Thinking about Bivariate Data

The following item appeared on the 2009 National Assessment of Educational Progress (NAEP):

A random sample of graduates from a particular college program reported their ages and incomes in response to a survey. Each point on the scatterplot above represents the age and income of a different graduate.

Based on the data in the scatterplot, predictions can be made about the income of a 35 year old and the income of a 55 year old. For which age is the prediction more likely to be accurate?

◯ 35 year old   ◯ 55 year old

Justify your answer.

Here are some students' responses to the NAEP item:

Student 1

Based on the data in the scatterplot, predictions can be made about the income of a 35 year old and the income of a 55 year old. For which age is the prediction more likely to be accurate?

● 35 year old   ○ 55 year old

Justify your answer.

There is actual data that can be referenced for 35 year olds

Student 2

Based on the data in the scatterplot, predictions can be made about the income of a 35 year old and the income of a 55 year old. For which age is the prediction more likely to be accurate?

● 35 year old   ○ 55 year old

Justify your answer.

There are no data points for 55 years old, so it would just be a prediction. For 35 years old, there are data points, so it would not just be a prediction.

Student 3

Based on the data in the scatterplot, predictions can be made about the income of a 35 year old and the income of a 55 year old. For which age is the prediction more likely to be accurate?

○ 35 year old   ● 55 year old

Justify your answer.

because the graph does not show a 55 year old person's income

Student 4

Based on the data in the scatterplot, predictions can be made about the income of a 35 year old and the income of a 55 year old. For which age is the prediction more likely to be accurate?

○ 35 year old   ● 55 year old

Justify your answer.

The 35 year old has data and the 55 year old does not.

Student 5

Based on the data in the scatterplot, predictions can be made about the income of a 35 year old and the income of a 55 year old. For which age is the prediction more likely to be accurate?

● 35 year old   ○ 55 year old

Justify your answer.

because they are still working.

Student 6

Based on the data in the scatterplot, predictions can be made about the income of a 35 year old and the income of a 55 year old. For which age is the prediction more likely to be accurate?

○ 35 year old   ● 55 year old

Justify your answer.

The older you get the more work experience you will have

Student 7

Based on the data in the scatterplot, predictions can be made about the income of a 35 year old and the income of a 55 year old. For which age is the prediction more likely to be accurate?

○ 35 year old   ● 55 year old

Justify your answer.

A fifty five year old would be more likely to accumulate that much wealth than a thirty five year old.

Source: Twelfth-grade NAEP test item, block: 2009-12M2, number 4 (NAEP 2009); National Assessment of Educational Progress

After reading each student response presented above, do the following:

1. Organize the responses from "most developed" to "least developed." You can use as many categories to organize the responses as you wish. Explain why you categorized the students' responses in the manner you chose.

2. Describe the steps you would take to further support each student's learning.

References

Boaler, J. (1998), "Open and Closed Mathematics: Student Experiences and Understandings," *Journal for Research in Mathematics Education*, 29, 1, pp. 41–62.

Bransford, J.D., Brown, A.L., and Cocking, R.R. (1999), *How People Learn: Brain, Mind, Experience, and School*. Washington, D.C.: National Academies Press.

Burgess, T.A. (2011), "Teacher Knowledge of and for Statistical Investigations," in *Teaching Statistics in School Mathematics – Challenges for Teaching and Teacher Education*, eds. C. Batanero, G. Burrill, and C. Reading, Dordrecht: Springer, pp. 259-270.

Carpenter, T.P., Fennema, E., Peterson, P.L., Chiang, C., and Loef, M. (1989), "Using Knowledge of Children's Mathematics Thinking in Classroom Teaching: An Experimental Study," *American Educational Research Journal*, 26, 4, pp. 499-531.

Cobb, G.W. (2013), "What Might a Twenty-Year Old Conference Tell us about the Future of our Profession?," *Journal of Statistics Education* [online], 21, 2, Available at www.amstat.org/publications/jse/v21n2/cobb.pdf.

Common Core State Standards Initiative (2010), *Common Core State Standards for Mathematics*, Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Conference Board of the Mathematical Sciences (2001), *The Mathematical Education of Teachers*, Providence, R.I. and Washington, DC: American Mathematical Society and Mathematical Association of America.

Conference Board of the Mathematical Sciences (2012), *The Mathematical Education of Teachers II*, Providence, R.I. and Washington, DC: American Mathematical Society and Mathematical Association of America.

CUPM (in preparation), *Undergraduate Programs and Courses in the Mathematical Sciences: CUPM Curriculum Guide 2015*, The Mathematical Association of America.

delMas, R., Garfield, J., Ooms, A., and Chance, B. (2007), "Assessing Students' Conceptual Understanding After a First Course in Statistics," *Statistics Education Research Journal* [online], 6, 2, pp. 28-58, Available at
http://www.stat.auckland.ac.nz/~iase/serj/SERJ6%282%29_delMas.pdf

Franklin, C., Bargagliotti, A., Case, K., Jacobbe, T., Kader, G., Spangler, D., and Schaefer, R. (in preparation), *The Statistics Education of Teachers Report* commissioned by the American Statistical Association.

Franklin, C., Kader, G., Mewborn, D. S., Moreno, J., Peck, R., Perry, M., and Scheaffer, R. (2007), *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework*, Alexandria, VA: American Statistical Association.

Fuson, K.C., Kalchman, M., and Bransford, J.D. (2005), "Mathematical Understanding: An Introduction," in *How Students Learn: Mathematics in the Classroom*, eds. M.S. Donovan and J.D. Bransford, Washington, D.C.: National Academies Press, pp. 217-256.

Gal., I, & Garfield, J. (1997), *The Assessment Challenge in Statistics Education*, Amsterdam: IOS Press.

Garfield, J.B., & Ben-Zvi, D. (2008), *Developing Students' Statistical Reasoning*: *Connecting Research and Teaching Practice*. New York: Springer.

Godino, J.D., Ortiz, J.J., Roa, R., and Wilhelmi, M.R. (2011), "Models for Statistical Pedagogical Knowledge." in *Teaching Statistics in School Mathematics – Challenges for Teaching and Teacher Education*, eds. C. Batanero, G. Burrill, and C. Reading, Dordrecht: Springer, pp. 271-282.

Green, J.L., & Blankenship, E.E. (2013), "Primarily Statistics: Developing an Introductory Statistics Course for Pre-Service Elementary Teachers," *Journal of Statistics Education* [online], 21, 3, Available at www.amstat.org/publications/jse/v21n3/green.pdf.

Groth, R. E. (2007), "Toward a Conceptualization of Statistical Knowledge for Teaching," *Journal for Research in Mathematics Education*, 38, 5, pp. 427-437.

Groth, R.E. (2012), "The Role of Writing Prompts in a Statistical Knowledge for Teaching Course," *Mathematics Teacher Educator* [online], 1, 1, pp. 23-40. Available at http://www.nctm.org/publications/article.aspx?id=33985.

Groth, R.E. (2013), "Characterizing Key Developmental Understandings and Pedagogically Powerful Ideas within a Statistical Knowledge for Teaching Framework," *Mathematical Thinking and Learning*, 15, 2, pp. 121-145.

Groth, R.E., and Bergner, J.A. (2013), "Mapping the Structure of Knowledge for Teaching Nominal Categorical Data Analysis," *Educational Studies in Mathematics*, 83, 2, pp. 247-265.

Henningsen, M., and Stein, M. K. (1997), "Mathematical Tasks and Student Cognition: Classroom-Based Factors that Support and Inhibit High-Level Mathematical Thinking and Reasoning," *Journal for Research in Mathematics Education*, 28, 5, 524–549.

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, pp. 372-200.

Hill, H.C., Sleep, L., Lewis, J.M., and Ball, D.L. (2007), "Assessing Teachers' Mathematical Knowledge: What Knowledge Matters and What Evidence Counts?," in *Second Handbook of Research on Mathematics Teaching and Learning*, ed. F.K. Lester, Charlotte, NC: Information Age Publishing and National Council of Teachers of Mathematics, pp. 111-155.

Jones, D., and Tarr, J.E. (2010), "Recommendations for Statistics and Probability in School Mathematics Over the Past Century," in *Mathematics Curriculum Issues, Trends, and Future Directions*: *Seventy-Second Yearbook*, eds. B.J. Reys and R.E. Reys, Reston, VA: National Council of Teachers of Mathematics, pp. 65-75.

Jones, G.A., Langrall, C.W., and Mooney, E.S. (2007), "Research in Probability: Responding to Classroom Realities," in *Second Handbook of Research on Mathematics Teaching and Learning*, ed. F.K. Lester, Charlotte, NC: NCTM and Information Age Publishing, pp. 909-955.

Konold, C., and Higgins, T.L. (2003), "Reasoning about Data," in *A Research Companion to Principles and Standards for School Mathematics*, eds. J.W. Kilpatrick, W.G. Martin, and D. Schifter, Reston, VA: NCTM, pp. 193-215.

Ma, L. (1999), *Knowing and Teaching Elementary Mathematics*: *Teachers' Understanding of Fundamental Mathematics in China and the United States*, Mahwah, NJ: Erlbaum.

McCrory, R., Floden, R., Ferrini-Mundy, J., Reckase, M.D., and Senk, S.L. (2012), "Knowledge of Algebra for Teaching: A Framework of Knowledge and Practices," *Journal for Research in Mathematics Education*, 43, 5, pp. 584-615.

Metz, M. L. (2010), "Using GAISE and NCTM Standards as Frameworks for Teaching Probability and Statistics to Pre-service Elementary and Middle School Mathematics Teachers," *Journal of Statistics Education* [online], 18, 3. Available at http://www.amstat.org/publications/jse/v18n3/metz.pdf.

Mokros, J., and Russell, S.J. (1995), "Children's Concepts of Average and Representativeness," *Journal for Research in Mathematics Education*, 26, 1, pp. 20-39.

National Council of Teachers of Mathematics (1989), *Curriculum and Evaluation Standards for School Mathematics*, Reston, VA: Author.

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

Pesek, D.D., & Kirshner, D. (2000), "Interference of Instrumental Instruction in Subsequent Relational Learning," *Journal for Research in Mathematics Education*, 31, 5, pp. 524-540.

Saderholm, J., Ronau, R., Brown, E.T., and Collins, G. (2010), "Validation of the Diagnostic Teacher Assessment of Mathematics and Science (DTAMS) Instrument," *School Science and Mathematics*, 110, 4, pp. 180-192.

Scheaffer, R. L., Gnanadesikan, M., Watkins, A. E., and Witmer, J.A. (1996), *Activity-Based Statistics : Instructor Resources*, New York : Springer.

Schoenfeld, A.H. (2007), "Method." in *Second Handbook of Research on Mathematics Teaching and Learning*, ed. F.K. Lester, Jr., Charlotte, NC: NCTM and Information Age Publishing, pp. 69-107.

Shulman, L.S. (1987), "Knowledge and Teaching: Foundations of the New Reform," *Harvard Educational Review*, 57, 1, pp. 1-23.

Stigler, J.W., and Hiebert, J. (1999), *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom*, New York: Free Press.

Tomlinson, C.A. (2005), *The Differentiated Classroom: Responding to the Needs of all Learners*, Upper Saddle River, NJ: Pearson.

Wassong, T., and Biehler, R. (2010, July), "A Model for Teacher Knowledge as a Basis for Online Courses for Professional Development of Statistics Teachers," in *Data and Context in Statistics Education: Towards an Evidence-Based Society*, Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8), ed. C. Reading, Ljubljana, Slovenia.

Randall E. Groth
Salisbury University
Department of Education Specialties
1101 Camden Ave.
Salisbury, MD 21801
regroth@salisbury.edu