

DEVELOPING STATISTICAL KNOWLEDGE FOR TEACHING DURING DESIGN-BASED RESEARCH

RANDALL E. GROTH
Salisbury University
regroth@salisbury.edu

ABSTRACT

Statistical knowledge for teaching is not precisely equivalent to statistics subject matter knowledge. Teachers must know how to make statistics understandable to others as well as understand the subject matter themselves. This dual demand on teachers calls for the development of viable teacher education models. This paper offers one such model, which relies upon engaging teachers in design-based research. Teachers collaborate with a researcher to design, implement, and analyze instruction to pursue desired statistical learning outcomes for students. The researcher allows teachers enough autonomy to make and learn from mistakes during the process. Unpacking and addressing the mistakes has value as a means of teacher learning. The model and a specific instance of its implementation are described along with reflections on how productive mistakes during design-based research provide opportunities for fostering the development of statistical knowledge for teaching.

Keywords: *Statistics education research; Teacher education models; NAEP tasks; Productive mistakes*

1. STATISTICAL KNOWLEDGE FOR TEACHING

A great deal of statistics education research is about understanding how students learn statistics (Garfield, 1995; Garfield & Ben-Zvi, 2007; Shaughnessy, 2007). A growing amount of statistics education research explores how teachers learn statistics (e.g., Makar & Confrey, 2004; Reading & Canada, 2011). Although research on how teachers learn statistics is important, an additional core issue needs attention: Along with learning statistics for themselves, teachers must learn how to help others understand it. In order to do so, teachers need to develop professional knowledge beyond statistical subject matter knowledge. For example, they must learn to assess and understand students' reasoning (Pfannkuch & Ben-Zvi, 2011), and select tasks suitable for advancing students' learning. Learning the subject matter of statistics is a necessary, but not sufficient, condition for teaching it.

Shulman (1987) coined the phrase *pedagogical content knowledge* to describe the knowledge teachers need in order to make subject matter comprehensible to students. He called it a "special amalgam of content and pedagogy that is uniquely the province of teachers" (p. 8). Researchers in mathematics and statistics education have appropriated, refined, and extended Shulman's notion of pedagogical content knowledge to describe the professional knowledge teachers need. One such notable effort was that of the Learning Mathematics for Teaching (LMT) project (Ball, Thames, & Phelps, 2008). The LMT model conceptualized mathematical knowledge for teaching as consisting of both subject matter

knowledge and pedagogical content knowledge, and made hypotheses about the nature of each of these knowledge domains.

In the LMT model, subject matter knowledge and pedagogical content knowledge are multi-faceted. Subject matter knowledge consists of common knowledge, specialized knowledge, and horizon knowledge. Common knowledge is that which is required across a variety of occupations, and is not unique to teaching (e.g., calculating statistics, understanding their meanings, etc.). Specialized knowledge is unique to the task of teaching. It allows teachers to appraise students' non-conventional representations and approaches to problems. Horizon knowledge entails knowing subject matter beyond the school curriculum and allows teachers to steer students' learning appropriately as opportunities arise (Ball et al., 2008).

The LMT model also identified three facets of pedagogical content knowledge: knowledge of content and students, knowledge of content and teaching, and curriculum knowledge. Knowledge of content and students provides teachers with insight on how students think about the discipline at hand, which allows teachers to anticipate students' thinking. Knowledge of content and teaching provides instructional strategies for making specific concepts understandable. Knowledge of curriculum enables teachers to sequence lessons and tasks in ways that support students' learning (Groth & Meletiou-Mavrotheris, in press).

Statistics education researchers have used and adapted the LMT model to describe and explore the knowledge teachers need for teaching statistics. A key consideration in adapting the LMT model is that statistics and mathematics are distinct disciplines. So, primarily non-mathematical elements of statistics, such as constructing survey questions and designing studies (Groth, 2007) are also important to statistics teacher knowledge development and should not be overlooked. Nonetheless, the LMT categories and adaptations of them have been used profitably to investigate and describe statistical knowledge for teaching (SKT) in regard to statistical investigations (Burgess, 2011), informal inference (Leavy, 2010), and sampling (Noll, 2011), among other topics.

2. PURPOSE

Although the LMT model and its adaptations are useful for the purpose of describing desirable learning goals for teachers, they are largely silent about how to help teachers attain the goals. Therefore, in order for the field of statistics teacher education to continue to mature, it is important to formulate, test, and refine models of how teacher educators might foster the development of SKT. The purpose of this paper is to offer one such model. The proposed model relies upon involving statistics teachers as partners in design-based research.

I begin by describing potential connections between design-based research and teacher education. I offer a model to facilitate teacher learning during design-based research, and provide an empirical example to illustrate the application of the model to a specific context. Instances of teacher mistakes during the process are used to exemplify potential opportunities for the development of SKT. Finally, implications for teacher educators and researchers are considered.

3. INVOLVING STATISTICS TEACHERS IN DESIGN-BASED RESEARCH

Design-based research is akin to engineering research. It aims to produce empirically tested theory about student learning in tandem with means for supporting student learning (Bakker & van Eerde, 2015). Researchers engineer these two products simultaneously. At

the outset of the process, researchers make conjectures about how student learning may occur and design instruction accordingly. They then empirically test the conjectures and the accompanying instructional approach with students, refine the instructional approach and conjectures in light of empirical data, and subject the refined approach and conjectures to further empirical testing with the students. These activities occur in continuous, repetitive cycles (The Design-Based Research Collective, 2003). A depiction of one such cycle appears in Figure 1.

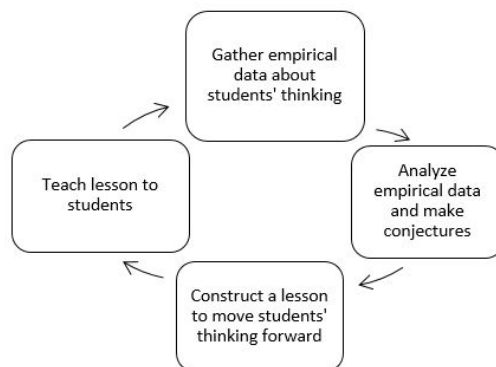


Figure 1. A research and development cycle to support design-based research

At the core of design-based research is Simon's (1995) notion of *hypothetical learning trajectory* (HLT), which consists of a goal for student learning, tasks used to promote learning, and hypotheses about how students' learning might develop. Researchers constructing HLTs make initial hypotheses about how students' thinking might develop and then progressively refine them as students' responses to tasks are observed. Researchers design and select subsequent tasks as necessary to help students attain intended learning goals (Steffe & Thompson, 2000). Teaching experiments undergirded by HLTs have contributed to the literature on students' understanding of statistical distribution (Cobb, 1999), data distributions as aggregates (Ben-Zvi & Arcavi, 2001), and sampling variability (Bakker, 2004).

Along with providing a means for designing theory and instructional materials to support students' learning in a given domain, design-based research offers a potential opportunity for collaboration between practitioners and researchers (e.g., McClain & Cobb, 2001; Smit & van Eerde, 2011). In such collaborative studies, researchers work closely with teachers to construct and re-construct HLTs. The construction and re-construction process involves optimizing instruction by collaboratively designing lessons, analyzing classroom data, making conjectures on how to build students' emergent understanding, and testing the conjectures by introducing instructional materials that embody the conjectures (Cobb, 2000). Teachers may also be involved in retrospective analysis that occurs after empirical testing has been completed (Bakker & van Eerde, 2015).

Given the demands of design-based research, it is optimal for teachers who collaborate with researchers for the purpose of publishing formal reports to have well-developed SKT. Such knowledge naturally lends itself to the tasks of co-constructing HLTs, delivering instruction during research and development cycles, and carrying out retrospective analysis. However, consider the case in which teachers' SKT development is the foremost goal rather than formal publication of research results. Engaging in design-based research has the potential to build teachers' SKT because of the careful attention to classroom data it entails. For example, analyzing students' thinking about a statistical concept provides a

means for developing knowledge of content and students. Developing teaching strategies responsive to students' needs and testing them empirically can lead to enhanced knowledge of content and teaching. Discussing the main statistical concepts at stake during a study with a research team can build teachers' subject matter knowledge.

Making teachers' SKT development the foremost priority in design-based research puts an extra burden on the teacher educator. The teacher educator must simultaneously participate in the construction and re-construction of two HLTs: one for students' learning and another for teacher learning. At times, these HLTs may overlap, as teachers have shown some of the same difficulties as students in regard to ideas such as average (Leavy & O'Loughlin, 2006), variability (Makar & Confrey, 2005), and distribution (Mooney, Duni, VanMeenen, & Langrall, 2014). At other times, it is reasonable to expect the HLTs to diverge, especially if teachers have some prior statistics coursework. Divergence unavoidably occurs in the goals driving the two HLTs. Whereas HLT goals for students' development consist only of subject matter knowledge, HLT goals for teacher development must include pedagogical content knowledge as well. In order for design-based research to reach its full potential as a means for teacher education, models and accounts of how teacher educators might simultaneously attend to the development of students' statistical knowledge and teachers' SKT are needed. Next, one particular model and an accompanying account of its implementation are offered.

4. MODEL AND EMPIRICAL SETTING

The empirical setting for the development of this theoretical model was an undergraduate research project focused on mathematics teacher education (Groth, Bergner, Burgess, Austin, & Holdai, 2016). As part of the project, I, a teacher educator, conducted collaborative design-based research with two undergraduates preparing to be secondary mathematics teachers, Rachel and Shantel (pseudonyms). The research involved developing the statistical thinking of four children who were entering Grade 6: Jonah, Rhonda, Mary, and Tyrone (pseudonyms). The specific learning goals we sought to help the children attain were:

- Display numerical data in plots on a number line, including dot plots, histograms, and line plots.
- Summarize numerical data sets in relation to their context, such as by:
 - a. Reporting the number of observations.
 - b. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.
 - c. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
 - d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010, p. 45).

The statistical content of these standards can be difficult for teachers and students alike (Groth & Bergner, 2006; Jacobbe, 2012; Leavy & O'Loughlin, 2006).

Because Rachel and Shantel were prospective teachers who each had taken introductory statistics but had not yet taught the subject, building their SKT was the primary goal of the design-based research experience. The model used to build their SKT (Figure 2) is now introduced and described in terms of its three main components: entry

point, core repetitive research and development cycle, and exit point. Although I illustrate the model in the context of work with prospective teachers, its components are relevant to the task of fostering teacher learning in general (hence the use of the word “teacher” rather than just “prospective teacher” in Figure 2).

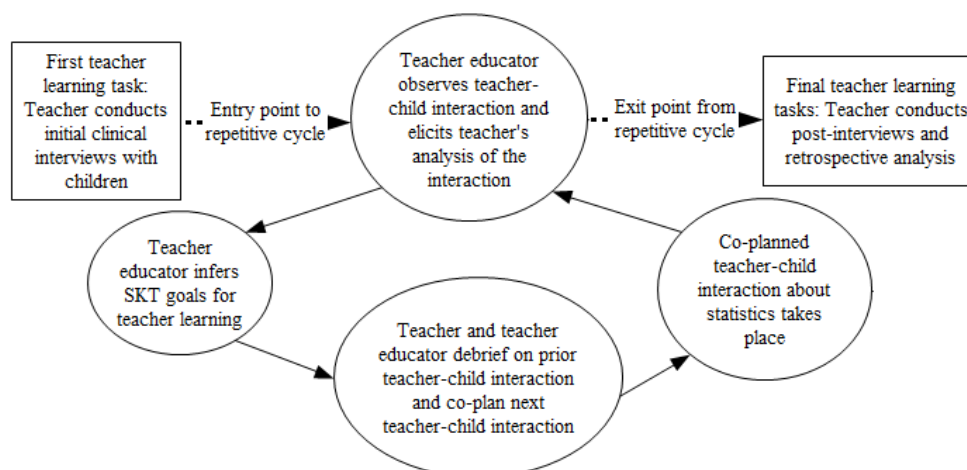


Figure 2. A model for fostering teachers' SKT development during design-based research

4.1. MODEL ENTRY POINT: CONDUCTING INITIAL CLINICAL INTERVIEWS

The first interaction Rachel and Shantel had with the children occurred as they conducted individual problem-solving clinical interviews with them. I hypothesized that conducting these interviews would help build their knowledge of content and of students. Knowledge of content and students provides a foundation for developing other aspects of pedagogical content knowledge, including knowledge of content and teaching and knowledge of curriculum (Groth, 2013). Research has consistently demonstrated that teachers with knowledge of students' cognition are better positioned to select teaching strategies and curricula that foster student learning (Sowder, 2007). Additionally, teachers tend to value opportunities to construct knowledge of how students think about content through firsthand interactions with students rather than solely through academic articles and books (McDonough, Clarke, & Clarke, 2002).

Before interacting with the children, Rachel and Shantel built background knowledge about conducting clinical interviews. I provided an article about logistics involved in conducting video-recorded clinical interviews (Ellemor-Collins & Wright, 2008) to help them anticipate key parts of their roles as interviewers. They also viewed video examples of clinical interviews being conducted (Learn NC & Wheatley, 2001). These examples allowed Rachel and Shantel to see interview techniques enacted by an experienced researcher. Additionally, I selected tasks for the initial interview script to help reveal the children's conceptual understanding of statistical ideas relevant to our design-based research. We discussed the tasks, their connections to the targeted curriculum standards, and possible student difficulties before interviews took place. I hypothesized that having these background experiences would help the two prospective teachers optimize their opportunities to develop knowledge of content and students during the initial interviews.

4.2. THE MODEL'S CORE REPETITIVE CYCLE

After conducting the initial clinical interviews, Rachel and Shantel entered the core repetitive cycle of the model (Figure 2). They produced transcripts of the interviews and analyzed them to identify the children's statistical strengths and needs. They used the insights they gained about the children's thinking to construct and carry out a one-hour lesson the following week. They video recorded the lesson, transcribed it, and retained the children's written work. In reviewing the classroom data, Rachel and Shantel looked for and summarized evidence of the children's strengths and weaknesses in regard to learning targeted statistical content. They then used these analyses to inform the construction of a subsequent lesson likely to help move the children's thinking forward. They implemented the lesson the following week, and again gathered and analyzed the video and written data from it. They continued to design, implement, test, and analyze lessons in this manner through the eighth week of the project.

I met with Rachel and Shantel after each lesson to hear their interpretations of video and written data from the interviews and lessons, share my own interpretations, and reconcile the interpretations as necessary. I directed the pair toward professional articles and teaching materials likely to help move the children's thinking forward and also to develop Rachel and Shantel's conceptual understanding of statistics content. I provided advice as they constructed lessons each week, but gave them responsibility for writing the lessons. After the first draft of each lesson had been written, I had them teach it to an audience of their peers and university faculty before implementing it with the children. These trial runs helped support thought experiments (Freudenthal, 1991) about the children's responses to our planned tasks, because audience members were to react to each lesson as they would expect the children to react. Although Rachel and Shantel were the lead teachers for the lessons they designed, at times I intervened as lead teacher when they struggled to guide the children toward intended learning goals.

4.3. MODEL EXIT POINT: CONDUCTING FINAL CLINICAL INTERVIEWS AND RETROSPECTIVE ANALYSIS

During the ninth week, Rachel and Shantel conducted post-assessment interviews with the children, once again using the script from the initial interviews. Comparing the final interview responses to the initial ones provided an assessment of children's overall progress and a starting point for retrospective analysis. During the tenth week, we completed collaborative retrospective analysis by identifying landmarks in children's learning that had occurred during the project. The poster and accompanying abstract constructed in the process became products which the pair presented at undergraduate research conferences.

I did an additional layer of retrospective analysis at the conclusion of the project to capture landmarks in SKT development that occurred as the collaborative design-based research took place. I reviewed weekly written feedback notes I prepared for Rachel and Shantel after each session, lesson and interview transcripts and video, and weekly written reports the pair had produced about the children's perceived statistical strengths and weaknesses. From these, I selected representative instances of opportunities for SKT development that arose during the design-based research to build a narrative describing an implementation of the model depicted in Figure 2. The implementation narrative appears next.

5. MODEL IMPLEMENTATION

5.1. WEEKS 1-3: DEALING WITH AGGREGATE DATA DISPLAYS AND EXAMINING DISTRIBUTIONS

One item administered during pre-assessment interviews (Figure 3) came from the National Assessment of Educational Progress (NAEP). The item required the children to choose measures of center to describe distributions. Students taking the NAEP generally struggled with the item. They often selected the mean as the best measure of center without examining the data distributions, believing that the mean is always better than the median for describing typical values (Zawojewski & Shaughnessy, 2000).

This question requires you to show your work and explain your reasoning. You may use drawings, words, and numbers in your explanation. Your answer should be clear enough so that another person could read it and understand your thinking. It is important that you show all of your work.

The table below shows the daily attendance at two movie theaters for 5 days and the mean (average) and the median attendance.

	<u>Theater A</u>	<u>Theater B</u>
Day 1	100	72
Day 2	87	97
Day 3	90	70
Day 4	10	71
Day 5	91	100
Mean (average)	75.6	82
Median	90	72

(a) Which statistic, the mean or the median, would you use to describe the typical daily attendance for the 5 days at Theater A? Justify your answer.

(b) Which statistic, the mean or the median, would you use to describe the typical daily attendance for the 5 days at Theater B? Justify your answer.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1996 Mathematics Assessment.

Figure 3. NAEP typical value item from interview script (U.S. Department of Education, Institute of Education Sciences, & National Center for Education Statistics, 2014)

I selected the NAEP task because of its potential to help Rachel and Shantel develop multiple aspects of SKT. As the item focuses on the uses of mean and median rather than their computation, I hypothesized that it would help them develop subject matter knowledge beyond a strictly procedural characterization of the two measures. This type of subject matter knowledge would be needed to help the children work toward the goal of “relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered” (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010, p. 45). I also hypothesized that allowing Rachel and Shantel to see how NAEP test-takers

often neglected examining distributions before choosing measures of center would help them develop knowledge of content and students and better anticipate similar thinking patterns among the participating children when planning lessons. Given that students may not naturally examine distributions before choosing measures of center, I also aimed to develop Rachel and Shantel's knowledge of curriculum by emphasizing the importance of drawing students' attention toward examining distributions if they do not do so on their own. A gradual progression from informal analysis of distributions to use of formal measures to describe center had also been emphasized in a learning progressions document I asked them to consider as they designed lessons (Common Core State Standards Writing Team, 2011). Finally, I hoped that extended attention to this task would help Rachel and Shantel develop knowledge of content and teaching in regard to posing statistical tasks. We discussed the positive aspect of the task in that it prompts students to move beyond strictly procedural knowledge of mean and median, but also considered how the manner in which the task associates "mean" with "average" might unduly influence the students' selection of an appropriate measure.

When Rachel and Shantel administered the task shown in Figure 3 to the children during pre-assessment clinical interviews, they struggled at times to formulate spontaneous probing questions to sustain conversations. As Rachel interviewed Tyrone, the following exchange occurred:

- Rachel: So, do you think the mean is a better representation of the numbers or the median is a better representation of the numbers?
 Tyrone: Median
 Rachel: Median, why do you say median?
 Tyrone: Because you did put them least to greatest; least to greater.
 Rachel: OK, so what about, same question, but with theatre B?

In this interview excerpt, Rachel did ask Tyrone to justify his initial response, but then did not probe further to ask how arranging things from least to greatest would help determine the typical value of the data set. When interviewing Mary, Rachel did go beyond the initial "why" question with a spontaneous probing question:

- Rachel: Which statistic, the mean or the median, would you use to describe the typical attendance for the five days at theatre A?
 Mary: The median.
 Rachel: OK, so why did you say the median?
 Mary: Because it's more than the mean.
 Rachel: OK, so how would that—being more than the mean—be a better description for the daily attendance?
 Mary: A lot more showed up that day.
 Rachel: OK, so, same question, but with theatre B...

Although Rachel spontaneously formulated one probing question in this instance, there was still a great deal more to learn about Mary's thinking. These missed opportunities provided chances for us to discuss how the children's thinking might be more effectively elicited, and how the elicitation of the children's thinking would be a crucial component of the inquiry-oriented lessons they would design and teach.

In another clinical interview question, Rachel and Shantel asked the children to produce graphs to display a data set showing the birth weights for 25 puppies born in a kennel (*Illustrative Mathematics*, n.d.). The children produced the graphs shown in Figure 4. Rachel and Shantel characterized Figures 4a and 4b as "bar graphs," placing them on equal footing. I challenged them to expand their interpretation of the two graphs by looking

beyond the cosmetic bar feature. The graph shown in Figure 4a aggregates the data, showing how many puppies had each birth weight (the birth weights are shown on the vertical axis). The graph shown in Figure 4b gives a bar to each individual value. So, for example, it shows that the first data value was 13, the second was 14, and so on (only a portion of the graph is shown). Although Figures 4a and 4b both make use of bars, Figure 4c is more similar to Figure 4a in the view it provides of the aggregate (graphs like Figure 4b are often called “case-value bars” (Bakker & Gravemeijer, 2004)). With the distinction between aggregate and non-aggregate displays more clearly drawn after our discussion of their differences, Rachel and Shantel decided to ask the children who naturally used aggregate displays to share their thinking with those who did not as they taught the first lesson.

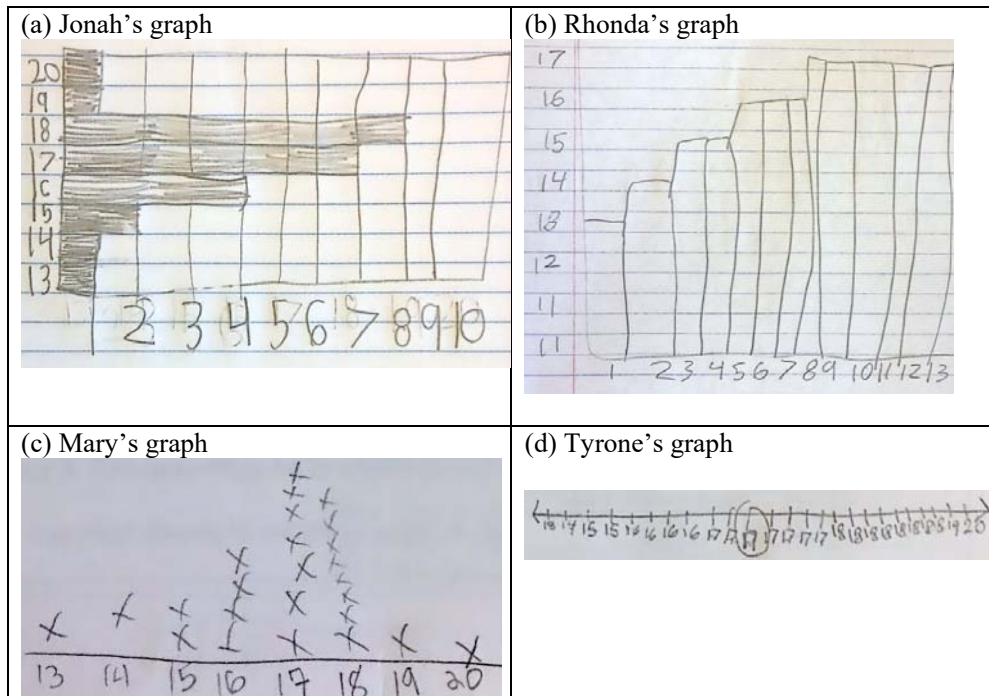


Figure 4. Children's use of graphical representations to organize a univariate data set

At the outset of the first lesson, Rachel and Shantel asked each of the children to construct a graph of their choosing for a set of scores obtained from playing a game that involved rolling two dice. Jonah used a graph similar to that shown in Figure 4a. Tyrone produced a graph like the one shown in Figure 4b, using a bar for each individual data value. When discussing the two graphs in class, Rachel and Shantel did not prompt the children to go beyond calling each representation a bar graph, indicating that they were still struggling to look beyond the cosmetic features of the two types of graphs.

When Mary constructed a dot plot (similar to Figure 4c), Rachel and Shantel asked her to show the graph to the rest of the group. They then asked the rest of the children to replicate Mary's strategy by working together at the board to re-construct the display using sticky notes to represent the dots in the dot plot. In subsequent tasks during the first two lessons, they asked the children to continue to use dot plots to represent univariate quantitative data.

One limitation of the first lesson was that constructing dot plots had become an end in itself rather than a means for examining distributions. As Rachel and Shantel planned their second lesson, I asked them to have the children begin to use aggregate data displays for the purpose of analyzing and describing characteristics of distributions. Rachel and Shantel then selected a word bank activity (Brodesky, Doherty, & Stoddard, 2008) to use with a *TinkerPlots* (Konold & Miller, 2011) file showing data from a group of cats. During the lesson, Rachel and Shantel asked the children to place words such as “cluster,” “skew,” and “hill” on univariate displays produced from quantitative data from a group of cats. They also began to ask the children to identify typical values in the cat data. Unfortunately, the activity of using the word “cluster” to describe portions of a distribution was largely separated from the activity of determining typical values. As a result, the children consistently identified the mode (the tallest stack) as the typical value in each distribution they were asked to examine rather than trying to identify central clusters. Rachel and Shantel also exhibited confusion about the statistical point of the lesson, telling the children at the end of the lesson that they would not have to construct dot plots anymore and would move on to something more enjoyable the next week. This was contrary to the progressions document they had studied earlier (Common Core State Standards Writing Team, 2011), which positioned dot plots as tools for analyzing distribution characteristics and comparing distributions. This led me to hypothesize that the next goal for Rachel and Shantel’s development of SKT would be to see the connections among the tasks of constructing dot plots, identifying clusters, and describing typical values.

5.2. WEEKS 4-5: DESCRIBING TYPICAL VALUES IN DISTRIBUTIONS

As Rachel and Shantel planned their third lesson, I suggested that they ask the children to look for central clusters in data and highlight them in some manner. I also suggested using a context in which the mode would not be a good indicator of typical value, in order to help prompt the children to look beyond just the tallest stack of dots in a dot plot. Rachel and Shantel decided to have the children gather data from scores obtained rolling dice. To make this a situation in which the mode would not be a good indicator of typical value, I suggested giving one of the children two loaded dice so that the sum of 12 would be obtained repeatedly. Rachel and Shantel took this suggestion and used it as a context for the children to gather data at the outset of the third lesson. The data the children gathered and the display they examined during class discussion are shown in Figure 5.

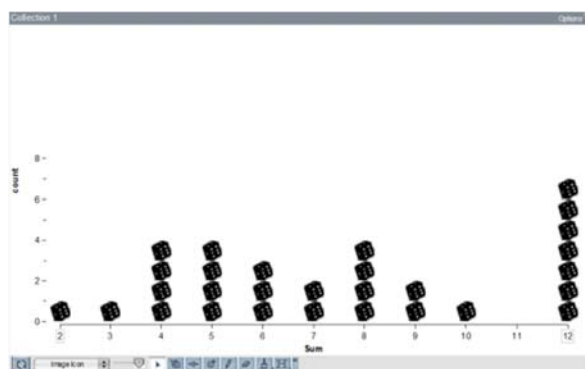


Figure 5. Scores obtained from rolling pairs of fair and loaded dice

As Rachel and Shantel questioned the children about the graph shown in Figure 5, they asked them to think about the word “cluster” from the word bank activity from the previous week and define what it would be. In response, Tyrone described a cluster as “when two things are close together...mushed.” Rhonda volunteered, “Like if you’re in an elevator and there’s like twenty people surrounding you and you can’t move cause you’re all close together.” After hearing these explanations, Rachel asked where they saw clusters in the graph. They gave a variety of answers, some as low as 2 and others as high as 12, since some were examining how close together the dots were within each vertical stack. Trying to draw their attention toward typical values, Rachel asked them to “circle the center.” This prompted Rhonda to circle the stack of dots above 6, simply because it was midway along the horizontal axis. Seeing that Rachel and Shantel were struggling to have the children identify a typical value for the data, I intervened in order to draw the children’s attention back to the context in which the data were gathered, asking, “What kinds of sums do people usually get when they play this game? When you look at the graph, what kinds of sums came up a lot?” They responded that 5 and 6 came up often, and that 12 did as well. When asked to think about why 12 came up so often, they rolled the dice again and noticed that one pair was producing almost all 12s. When they removed the loaded dice from the data and were asked again about scores for the game, they identified 6 and 7 as typical scores.

Given the events of the third lesson, I hypothesized that Rachel and Shantel next needed to develop a specific aspect of knowledge of content and teaching: leveraging the context of the data to help the children reason about typical values. Accordingly, I gave them the task of designing a lesson that engaged the children in thinking more deeply about context before and during data analysis. Rachel and Shantel designed two scenarios for their fourth lesson in response to this request. In the first, they told the children they had taken a poll of the allowances received by their classmates and then showed them the symmetric distribution it produced. The children were then to use the data to argue for a fair amount of allowance money. In the second scenario, the children were shown the heights of professional basketball players as well as the heights of a class of middle school students. They were then to compare the two data sets.

During the fourth lesson, Rachel and Shantel began by asking the children whether they received an allowance. Rachel then put a distribution of allowances she had invented for the task on the whiteboard and asked them to choose words from the word bank used in previous activities to describe parts of the distribution. They selected the informal words “hill,” “gap,” and “hole,” but did not select “cluster” or “center.” Rachel used the context of the problem to prompt the children to focus on the central cluster by asking them to look at the data and decide what a reasonable amount of allowance to request from their parents would be. At this point, some opted for purely contextual considerations, such as the difficulty of the work they were to do. Jonah, however, looked for the part of the graph with the most data values. Rachel then drew the children’s attention back to “center” in the word bank and asked where the center of the data would be. This prompted Tyrone to give an informal estimate of the center, and Mary suggested finding the median. Because the data were arranged in a dot plot, Rachel had Jonah and Mary work together to cross off dots on either side of the distribution until they met one another in the middle. This portion of the discussion portrayed the center of the data, the median, and informal estimates as concepts related to determining a typical allowance in the data.

Shantel presented the height data sets (professional basketball players and middle school students) during the second half of the fourth class session. During class discussion, the children gave several informal estimates of the center of each distribution. In order to build on the associations the children were beginning to make among the ideas of center, median, and visual estimates, I intervened to question the children on their thinking about

these connections. Specifically, I asked them to circle the informal estimates they had made during class discussion. I then asked them whether the median of each data set was inside or outside each circle. I also asked whether the tallest stack of dots could ever lie outside the central cluster. To conclude the lesson, children compared the central clusters of the two data sets and noticed that the center for the basketball team data was higher than that of the middle school student data.

5.3. WEEKS 6-8: CHOOSING AMONG MEASURES OF CENTER

For the final three lessons Rachel and Shantel taught, I altered their main task each week from designing an original lesson plan to implementing an existing set of lessons from a curriculum. The three lessons, from *Connected Mathematics* (Lappan, Fey, Fitzgerald, Friel, & Philips, 2004), introduced the mean as a data analysis tool and later asked children to decide on the best measure of center (mean, median, or mode) to use in different situations. I hypothesized that the task of implementing these lessons would foster their curriculum knowledge while simultaneously building the children’s abilities to choose from among different measures of center.

The first *Connected Mathematics* lesson introduced the mean as the fair share value for a set of data. Snap cube manipulatives were used to represent data values, and then cubes were to be re-distributed as necessary until each stack was of equal height. The height of each stack was the fair share value (i.e., the mean). So, during the first lesson in the sequence, Rachel and Shantel decided to give each child six stacks of cubes (Figure 6). Each stack of cubes represented a different family. The number of cubes in each stack represented the number of members for the given family.



Figure 6. A family size data set represented with snap cubes

When Rachel and Shantel did the trial run of the first Connected Mathematics lesson with their peers and other faculty members, several in the audience noted that the snap cube representation (Figure 6) looked very much like a dot plot. Since the children had already used dot plots extensively as tools to analyze distributions, there was strong potential for confusion during the lesson. The snap cube representation was more akin to the case-value bars some children had produced during initial clinical interviews and during the first lesson (e.g., Figure 4b), as each bar length represented one individual data value. Hence, it would be important to help children think about the distinction between dot plots and case value bars early in the lesson.

When teaching the lesson, Rachel did try to distinguish between dot plots and case value bars by telling the children, “We are going to be doing something a little bit different than our usual dot plots.” However, she did not provide or elicit further detail on how the snap cube representation differed from a dot plot. As a result, when she asked the children to find the median of the data set, they immediately began treating the snap cube representation as a dot plot. Rhonda and Tyrone, for example, suggested arranging the stacks from shortest to tallest and then snapping two blocks off at a time – one from the

stack furthest left and another from the stack furthest right – until arriving at a middle block. This mirrored the strategy for determining the median they used during the previous lesson with allowance data.

Rachel noticed the children's confusion of the two representations and attempted to intervene. However, her attempt reflected her own fragile, developing knowledge of the snap cube representation. She began by projecting the snap cube stacks (Figure 6) on a whiteboard. The stacks were arranged from shortest to tallest. Rachel drew a horizontal line under the stacks and then put the number 1 under the first stack, the number 2 under the second, etc., up to the sixth stack. She then took the stacks away, leaving the horizontal line and the integers 1-6 on the board. She asked children to look at the integers 1-6 and determine the median. Coincidentally, the median of 3.5 that was determined in this manner was the same as the actual median for the data set, potentially adding an extra layer of confusion.

The troubled implementation of the first Connected Mathematics lesson provided a number of teaching opportunities during mentoring sessions. Although Rachel and Shantel had begun to develop the knowledge of content and students necessary to anticipate that the children might interpret the snap cube representation as a dot plot, they did not do enough to address the potential confusion at the outset of the lesson. Knowledge of content and teaching, specifically pertaining to techniques for launching tasks (Jackson, Shahan, Gibbons, & Cobb, 2012) by eliciting and addressing children's thinking at the outset, was in need of further development. Specialized knowledge also needed to be addressed, as evidenced by the incorrect use of the snap cube representation on the whiteboard. (This is specialized knowledge in the sense that knowledge of the snap cube representation, unlike more conventional representations of data, is of more interest to teachers than other professionals who use statistics). Rachel and Shantel recognized the mistake with the snap cubes as we discussed it. The newness of the representation seemingly made it difficult for them to correct themselves in-the-moment while teaching.

One productive outcome of the first Connected Mathematics lesson was that the children were introduced to the arithmetic mean. The snap cube representation provided a tool for making sense of the add-and-divide algorithm, because stacks of cubes could all be snapped together and then separated back into the original number of smaller stacks in order to determine the fair share. Rachel and Shantel took advantage of the children's emerging knowledge of the mean during the next lesson. At the outset, they showed the children 24 candies. They then distributed them so that Rachel had 19 candies and everyone else had just one. Rachel and Shantel used this context to ask how many candies the typical person in the room had. They re-distributed the candies and asked children to re-evaluate their answers. The process of distributing and re-distributing was successful in prompting children to suggest strategies that aligned with the mean, median, mode, as well as more informal means for determining typical value. I intervened in the lesson only to start to push the children to begin making judgments about which types of strategies best described the typical value in each case.

For the final lesson, Rachel and Shantel selected a task from the Connected Mathematics sequence that required a judgment about the best measure of center to describe the typical income for individuals in a small town. The incomes were: \$0, \$0, \$0, \$0, \$0, \$0, \$0, \$0, \$200, \$200, \$200, \$200, \$200, \$200, \$200, \$30,600 (Lappan et al., 2004). The mayor of the town claimed that the typical income was \$2000. Two candidates running against him in an election thought that \$100 or \$0 would be better descriptions of typical income. The children were asked to decide which one-number summary was best. This prompted discussion of the mean, median, and mode, and of how outlying values can influence the mean. Rachel and Shantel asked the children to add and remove values from

the data set at several points during the lesson and observe how the mean changed each time. They also encouraged the children to debate the best choice of measure with one another as they modified the data set.

5.4. WEEKS 9-10: FINAL PROJECT SUMMARY

To close the project, Rachel and Shantel were given the tasks of re-administering the initial interview script to the children and then constructing a poster and giving an oral presentation to summarize their findings. As they gave their oral presentation to an audience of peers and university professors, aspects of SKT they had strengthened as well as some in need of further development were apparent.

When presenting graphs the children produced during pre-assessment interviews (Figure 4), Shantel still referred to Figures 4a and 4b as “bar graphs” as a way to distinguish them from the dot plot produced by one of the children (Figure 4c). She said, “Most of the students would start out with bar graphs, but we had one student who would use her procedural fluency and make a dot plot.” As happened earlier in the project, the cosmetic features (i.e., dots vs. bars) took precedence in Shantel’s characterization of the graphs rather than whether or not they aggregated the data. However, later in the presentation, she did show evidence of beginning to understand the aggregate vs. non-aggregate distinction when discussing the differences between the graphs one child produced on the pre- and post-assessment interviews. While showing Figure 4b to the audience, Shantel said

Here in the initial assessment we gave them a problem and it asked them to make an appropriate graph for the puppy birth weights. Here, one student actually made 17 case-value bars for each birth weight because she didn’t know how else to represent the graph.

Next, Shantel showed a dot plot the child had produced for the same set of data during the post-assessment interview and said, “They actually did gain procedural fluency and strategic competence. In the post-assessment, every student did make a dot plot for the data, using x s for each of the birth weights.” Although the crucial distinction for Shantel still appeared to be “dot plot vs. other type of graph” rather than “aggregate display vs. non-aggregate display,” she nonetheless attended to how the bars in Figure 4a showed just one data value each.

During her portion of the final presentation, Rachel explained how the research used data displays to support the children’s data analysis. While displaying Figure 5 for the audience, she explained that the children first organized the data into a dot plot. She then added

We had them actually come up and manipulate the middle clump, which is that grey area, so they would pick and choose where they thought that middle clump, that middle cluster, was. Then we built from that, how to find a middle data value from that clump.

Rachel’s remarks about the graph indicate growth from her view earlier in the project that dot plots were produced for their own sake rather than as tools for analyzing distributions. Rachel was also able to successfully explain how the snap cube model shown in Figure 6 could be used to determine the mean. Working with the model during a lesson with the children earlier in the project had proved to be problematic for her. Although she was successful explaining it during the presentation, her knowledge of the model did still appear to be in development. When an audience member asked her a follow-up question about the model, Rachel identified the number of stacks of cubes as representing the mean rather than the number of cubes in each stack.

In summary, the task of reflecting on the entire project orally helped reveal aspects of Rachel and Shantel’s SKT that had developed but still had room for improvement. Shantel

had improved her ability to interpret students' graphs, but perhaps had not yet fully recognized the importance of the aggregate vs. non-aggregate distinction. Rachel appeared to have better understanding of the snap cube model for the mean, but her knowledge of the model had not yet solidified to the point that she could consistently explain it correctly. Taken together with the events that occurred during design-based research, it appeared that both had made a degree of progress toward remedying and learning from mistakes made along the way.

6. PRODUCTIVE MISTAKES DURING DESIGN-BASED RESEARCH

In design-based research, as in all engineering research, mistakes are inevitable. The number of mistakes is likely to be greater when teachers who collaborate on statistics education research have relatively limited SKT. Rachel and Shantel made several such mistakes. I wish to unpack some of their mistakes to conclude the paper. Focusing on mistakes might, at first glance, seem counter-productive. Certainly, research on teacher learning that uses deficit models to describe teachers' knowledge is of limited value. However, my focus here is not simply on describing mistakes, per se, but on how conducting design-based research provided a space for making *productive mistakes*. It can be argued that productive mistakes are one of the chief means through which Rachel and Shantel had opportunities to develop SKT during the project.

Productive mistakes occur frequently in engineering research. Some famous mistakes led to the development of products such as sticky notes, penicillin, and rubber tires (Gojak, 2013). Making a mistake often leads one down a reasoning path that otherwise might remain unexplored (Petroski, 1985). Furthermore, making mistakes actually stimulates brain growth, and that growth is greatest when accompanied by a mindset that one can improve with effort (Boaler, 2016). Hence, it is important for teachers to help students see mistakes they make in doing statistics and mathematics as learning opportunities rather than avoiding mistakes at all costs (Hiebert et al., 1997). Similarly, teacher educators can help teachers use their mistakes during design-based research as opportunities for SKT development by drawing teachers' attention to the mistakes and using them as prompts for discussing how to improve future instruction.

6.1. PRODUCTIVE MISTAKE 1: LACK OF PROBING DURING INITIAL INTERVIEWS

One mistake Rachel and Shantel made while conducting initial clinical interviews was failing to probe the children's thinking in some cases when it was unclear. Noticing this, I asked them to explain why the children chose the mean in some cases and the median in others to describe the data shown in Figure 3. They realized they did not have enough information in the interview transcripts to know why some children chose as they did. This underscored the importance of formulating questions that carefully probed specific aspects of the children's responses. Fortunately, this mistake occurred near the outset of the project, providing time for them to re-think their questioning strategies before carrying out instruction. Although missed opportunities to probe the children's thinking still occurred at times during the lessons Rachel and Shantel taught, making the mistake of minimal probing during the initial clinical interviews and discussing it collaboratively helped place the elicitation and study of the children's statistical thinking at the center of the design-based research they conducted. Eliciting and studying the children's thinking subsequently provided multiple opportunities for Rachel and Shantel to enhance their knowledge of content and students.

6.2. PRODUCTIVE MISTAKE 2: LACK OF ATTENTION TO IMPORTANT DETAILS IN GRAPHS

Rachel and Shantel's lack of attention to details in some of the children's graphs suggested their specialized content knowledge needed development. The first such instance occurred when they categorized the children's graphs (Figure 4) according to whether or not they had bars rather than by whether or not they aggregated the data. The second instance was when Rachel portrayed case value bars represented with snap cubes (Figure 6) as if they formed a dot plot. As I drew their attention to these mistakes, they began to remedy them. By the end of their design-based research, they had not consistently avoided these mistakes across all contexts, but they did acknowledge the aggregate vs. non-aggregate and case value bar vs. dot plot distinctions as important and began to take them into account in planning lessons. Rachel's ability to explain the differences between case value bars and dot plots improved even though it was still in development.

6.3. PRODUCTIVE MISTAKE 3: FRAMING GRAPH PRODUCTION AS AN END IN ITSELF

Sensing that the children were tired of producing dot plot representations during the initial lessons they taught, Rachel and Shantel told the children they would not have to work with dot plots again in future lessons. At that point in the lesson sequence, the children had produced dot plots but had not begun to use them extensively to analyze and compare distributions. Discarding the dot plot representation at such a point in the lesson sequence would have resulted in the children having the ability to produce dot plots but not to use them for their intended purposes.

During our debriefing of the lesson in question, I had the opportunity to show Rachel and Shantel how dot plots ultimately could be used to help the children identify central clusters in distributions. We temporarily looked beyond the specific lesson they taught and focused instead on how it might fit within a series of lessons capable of helping the children reach our specified statistical learning goals. This broader perspective was introduced to help develop Rachel and Shantel's curriculum knowledge, and we discussed a collection of individual lessons that could progressively lead to a desired outcome. Subsequently, Rachel and Shantel began to design and execute their own lessons that portrayed producing dot plots as part of a data analysis strategy rather than an end in itself. Some examples of this were their lessons on scores from rolling dice, typical allowance, and comparing basketball teams' heights.

6.4. PRODUCTIVE MISTAKE 4: DIVORCING DATA ANALYSIS FROM CONTEXT

As Rachel and Shantel began to have the children use dot plots and other displays to analyze distributions, they made the mistake of divorcing the children's analyses from the context of the data. In the distribution of scores for the dice game (Figure 5), for example, they asked the children to identify the "center" without providing a context-based reason for doing so. Their struggles facilitating the children's analysis of the distribution provided an opportunity for me to intervene in their lesson and model how to ask questions that leveraged the context of the data to help the children reason about center. As we debriefed on the lesson, we discussed how the questions I asked differed from those they asked earlier in the lesson. Going forward, Rachel and Shantel began to frame their questions in terms

of the context of the data. For example, they took the lead on designing the lesson that involved distributing candies unequally to the children to use the idea of fairness to motivate the identification of typical values. The lesson then became an opportunity to use statistical tools to describe a situation at hand rather than an exercise in executing procedures or algorithms loosely tied to a context. Their ability to execute such lessons suggested improved knowledge of content and teaching.

7. CONCLUSION

Design-based research has strong potential to foster teacher learning because of the careful attention to classroom data it entails. As teachers construct and re-construct HLTs in collaboration with researchers, multiple opportunities to develop SKT can emerge. Researchers and teacher educators should bear in mind that many of the opportunities for teacher learning will come in the form of productive mistakes teachers make during the research process. Hence, when using design-based research primarily as a teacher education tool, eliminating all opportunities for mistakes should not be a goal. Rather, teacher educators and researchers can transform mistakes into opportunities for SKT development by unpacking the mistakes during lesson debriefing sessions and exploring how to avoid them in the future. Reflecting on teaching practices for the purpose of improving them provides robust opportunities for teacher learning (Ricks, 2011; Roback, Chance, Legler, & Moore, 2006).

Using mistakes as learning opportunities does, however, come with challenges. Not all mistakes are productive. Some can be harmful. The harm is potentially large in the teaching profession, because teachers' actions influence student learning. Therefore, it is not ethically tenable to set teachers up to make specific mistakes in the classroom in order to gain teacher learning opportunities. Instead, teacher educators must take advantage of productive mistakes as they naturally occur. Design-based research provides a controlled setting in which this can be accomplished. Carrying out the research with a small number of students and engaging in constant reflective dialogue about how to improve instruction can minimize the harm done by mistakes in the classroom. In the process, teachers become conscious of avoiding the identified mistakes in their own classrooms. Mistakes made in a controlled setting with a small number of students thus might not be repeated in classrooms where the teacher functions independently with a larger number of students. Research in natural classroom settings suggests that potentially harmful mistakes teaching statistics are prevalent (Groth & Meletiou-Mavrotheris, in press), so design-based research during pre-service teacher education may provide a means of heading off such mistakes before they occur.

Although a great deal can be learned from productive mistakes, teacher learning during design-based research may occur through other means as well. Design-based researchers function as tinkerers, or *bricoleurs* (Bakker & van Eerde, 2015), who piece together articles, instructional materials, and experiences with students to build theory and learning sequences. Engaging with such resources leads to a more comprehensive view of the discipline of statistics education. Teachers also hone lesson design skills as they attend closely to student thinking and participate in HLT construction. Hence, the model has value even if a relatively small number of productive mistakes occur as design-based research takes place.

Despite its potential, the model discussed in this paper presents a number of challenges. Teacher educators must grapple with when to intervene and when to allow enough room for teachers to struggle during the process. Overcoming the stigma associated with making mistakes is a constant challenge (Koestler, Felton, Bieda, & Otten, 2013), and teacher

educators must learn to skillfully address it. The types of teacher learning opportunities that will emerge during design-based research are somewhat unpredictable because many of them are contingent upon teachers' actions. This unpredictable nature can make it difficult to specify teacher learning outcomes in advance. The model is also very resource-intensive. In particular, teacher educators must have the time necessary for deep engagement with teachers during the experience. Each of these challenges to implementing the model merit further attention and thought. Nonetheless, as teacher educators and researchers attend to these challenges, we can begin to replace the problematic traditional dichotomy between teaching and research (Kennedy, 1997; Malara & Zan, 2002) with a vision of how conducting statistics education research can become an essential teacher learning experience.

ACKNOWLEDGEMENTS

The work described in this paper was supported by the National Science Foundation (NSF) under award DRL-1356001. The views expressed herein are those of the authors and do not necessarily reflect those of the NSF.

REFERENCES

- Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3(2), 64–83.
[Online: [http://www.stat.auckland.ac.nz/~iase/serj/SERJ3\(2\)_Bakker.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ3(2)_Bakker.pdf)]
- Bakker, A., & Gravemeijer, K. P. E. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147–168). Dordrecht, The Netherlands: Kluwer.
- Bakker, A., & van Eerde, H. A. A. (2015). An introduction to design-based research with an example from statistics education. In A. Bikner-Ahsbals, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education: Examples of methodology and methods* (pp. 429–466). New York: Springer.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389–407.
doi: 10.1177/0022487108324554
- Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, 45(1), 35–65.
doi: 10.1023/A:1013809201228
- Boaler, J. (2016). *Mathematical mindsets*. San Francisco, CA: Jossey-Bass.
- Brodesky, A., Doherty, A., & Stoddard, J. (2008). *Digging into data with TinkerPlots*. Emeryville, CA: Key Curriculum Press.
- Burgess, T. A. (2011). Teacher knowledge of and for statistical investigations. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics – Challenges for teaching and teacher education* (pp. 259–270). Dordrecht, The Netherlands: Springer.
- Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning*, 1(1), 5–43.
doi: 10.1207/s15327833mtl0101_1
- Cobb, P. (2000). Conducting teaching experiments in collaboration with teachers. In A.E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 307–333). Mahwah, NJ: Lawrence Erlbaum Associates.

- Common Core Standards Writing Team. (2011). *Progression for the common core state standards for mathematics (draft), K–5, measurement and data*.
[Online: http://commoncoretools.files.wordpress.com/2011/06/ccss_progression_md_k5_2011_06_20.pdf]
- Ellemor-Collins, D. L., & Wright, R. J. (2008). Student thinking about arithmetic: Videotaped interviews. *Teaching Children Mathematics*, 15(2), 106–111.
- Freudenthal, H. (1991). *Revisiting mathematics education: China lectures*. Dordrecht, The Netherlands: Kluwer.
- Garfield, J. (1995). How students learn statistics. *International Statistical Review*, 63(1), 25–34.
- Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372–396. doi: 10.1111/j.1751-5823.2007.00029.x
- Gojak, L. M. (2013). *The power of a good mistake*. NCTM President’s Messages Web post. [Online: http://www.nctm.org/News-and-Calendar/Messages-from-the-President/Archive/Linda-M_-Gojak/The-Power-of-a-Good-Mistake/]
- Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education*, 38(5), 427–437.
- 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), 121–145. doi: 10.1080/10986065.2013.770718
- Groth, R. E., & Bergner, J. A. (2006). Preservice elementary teachers’ conceptual and procedural knowledge of mean, median, and mode. *Mathematical Thinking and Learning*, 8(1), 37–63. doi: 10.1207/s15327833mtl0801_3
- Groth, R. E., Bergner, J. A., Burgess, C. R., Austin, J. W., & Holdai, V. (2016). Re-imagining education of mathematics teachers through undergraduate research. *Council on Undergraduate Research (CUR) Quarterly*, 36(3), 41–46.
- Groth, R. E., & Meletiou-Mavrotheris, M. (in press). Research on statistics teachers’ cognitive and affective characteristics. *First International Handbook of Research in Statistics Education*. Springer.
- Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K. C., Wearne, D., Murray, H., Olivier, A., & Human, P. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.
- Illustrative Mathematics (n.d.). *6.SP Puppy weights*.
[Online: <https://www.illustrativemathematics.org/content-standards/tasks/1026>]
- Jackson, K. J., Shahan, E. C., Gibbons, L. K., & Cobb, P. (2012). Launching complex tasks. *Mathematics Teaching in the Middle School*, 18(1), 24–29.
- Jacobbe, T. (2012). Elementary school teachers’ understanding of mean and median. *International Journal of Science and Mathematics Education*, 10(5), 1143–1161. doi: 10.1007/s10763-011-9321-0
- Kennedy, M. M. (1997). The connection between research and practice. *Educational Researcher*, 26(7), 4–12. doi: 10.3102/0013189X026007004
- Koestler, C., Felton, M. D., Bieda, K. N., & Otten, S. (2013). *Connecting the NCTM Process Standards and the CCSSM Practices*. Reston, VA: National Council of Teachers of Mathematics.
- Konold, C., & Miller, C. (2011). *TinkerPlots* (Version v2.0) [Computer software]. Emeryville, CA: Key Curriculum Press.
- Lappan, G., Fey, J. T., Fitzgerald, W. M., Friel, S. N., & Philips, E. D. (2004). *Data about us*. New York: Pearson.
- Learn NC & Wheatley, G. (2001). *Problem-centered math*.

- [Online: <http://www.learnnc.org/lp/editions/pcmath/cover>]
- Leavy, A. M. (2010). The challenge of preparing preservice teachers to teach informal inferential reasoning. *Statistics Education Research Journal*, 9(1), 46–67.
[Online: [http://iase-web.org/documents/SERJ/SERJ9\(1\)_Leavy.pdf](http://iase-web.org/documents/SERJ/SERJ9(1)_Leavy.pdf)]
- 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. doi: 10.1007/s10857-006-9003-y
- Makar, K., & Confrey, J. (2004). Secondary teachers’ statistical reasoning in comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 353–373). Dordrecht, The Netherlands: Springer.
- Makar, K. & Confrey, J. (2005). “Variation-talk”: Articulating meaning in statistics. *Statistics Education Research Journal*, 4(1), 27–54.
[Online: [http://iase-web.org/documents/SERJ/SERJ4\(1\)_Makar_Confrey.pdf](http://iase-web.org/documents/SERJ/SERJ4(1)_Makar_Confrey.pdf)]
- Malara, N. A., & Zan, R. (2002). The problematic relationship between theory and practice. In L.D. English (Ed.), *Handbook of international research in mathematics education* (pp. 553-580). Mahwah, NJ: Lawrence Erlbaum Associates.
- McClain, K. & Cobb, P. (2001). Supporting students’ ability to reason about data. *Educational Studies in Mathematics*, 45(1), 103–129. doi: 10.1023/A:1013874514650
- McDonough, A., Clarke, B., & Clarke, D. M. (2002). Understanding, assessing and developing children’s mathematical thinking: The power of a one-to-one interview for preservice teachers in providing insights into appropriate pedagogical practices. *International Journal of Educational Research*, 37(2), 211–226.
doi: 10.1016/S0883-0355(02)00061-7
- Mooney, E., Duni, D., VanMeenen, E., & Langrall, C. (2014). Preservice teachers’ awareness of variability. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*, Flagstaff, Arizona, USA. Voorburg, The Netherlands: International Statistical Institute.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*.
[Online: <http://www.corestandards.org/>]
- Noll, J. A. (2011). Graduate teaching assistants’ statistical content knowledge of sampling. *Statistics Education Research Journal*, 10(2), 48–74.
[Online: [https://www.stat.auckland.ac.nz/~iase/serj/SERJ10\(2\)_Noll.pdf](https://www.stat.auckland.ac.nz/~iase/serj/SERJ10(2)_Noll.pdf)]
- Petroski, H. (1985). *To engineer is human: The role of failure in successful design*. New York: St. Martin’s Press.
- Pfannkuch, M., & Ben-Zvi, D. (2011). Developing teachers’ statistical thinking. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics – Challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 323–333). Dordrecht, The Netherlands: Springer.
- Reading, C., & Canada, D. (2011). Teachers’ knowledge of distribution. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics – Challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 223–234). Dordrecht, The Netherlands: Springer.
- Ricks, T. E. (2011). Process reflection during Japanese Lesson Study experiences by prospective secondary mathematics teachers. *Journal of Mathematics Teacher Education*, 14(4), 251–267. doi: 10.1007/s10857-010-9155-7
- Roback, P., Chance, B., Legler, J., & Moore, T. (2006). Applying Japanese Lesson Study principles to an upper-level undergraduate statistics course. *Journal of Statistics Education*, 14(2).

- [Online: <http://www.amstat.org/publications/jse/v14n2/roback.html>]
- Shaughnessy, J. M. (2007). Research on statistics learning. In F.K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957-1009). Charlotte, NC: Information Age Press and National Council of Teachers of Mathematics.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23.
- Simon, M. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, 26(2), 114–145.
- Smit, J. & van Eerde, H. A. A. (2011). A teacher's learning process in dual design research: Learning to scaffold language in a multilingual mathematics classroom. *ZDM The International Journal on Mathematics Education*, 43(6-7), 889–900.
doi: 10.1007/s11858-011-0350-5
- Sowder, J. T. (2007). The mathematical education and development of teachers. In F.K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 157-223). Charlotte, NC: National Council of Teachers of Mathematics and Information Age Publishing.
- Steffe, L., & Thompson, P. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A. K. R. Lesh (Ed.), *Handbook of research design in mathematics and science education* (pp. 267–306). Hillsdale, NJ: Lawrence Erlbaum Associates.
- The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5–8.
doi: 10.3102/0013189X032001005
- U.S. Department of Education, Institute of Education Sciences, and National Center for Education Statistics (2014). *NAEP Questions Tool v4.0*.
[Online: <http://nces.ed.gov/nationsreportcard/itmrlsx/>]
- Zawojewski, J. S., & Shaughnessy, J. M. (2000). Mean and median: Are they really so easy? *Mathematics Teaching in the Middle School*, 5(7), 436–440.

RANDALL E. GROTH
Salisbury University
Department of Education Specialties
1101 Camden Ave.
Salisbury, MD 21801
USA