

## COLLABORATIVE DESIGN WORK OF TEACHER EDUCATORS: A CASE FROM STATISTICS

Hollylynne S. Lee  
North Carolina State  
University  
Hollylynne@ncsu.edu

Helen Doerr  
Syracuse University  
hmdoerr@syr.edu

Jonas Årlebäck  
Linköping University  
jonas.bergman.arleback@liu.se

Tyler Pulis  
North Carolina State  
University  
hpulis@ncsu.edu

*Our study, situated within graduate-level courses for teachers focused on statistics, illustrates work across institutions for making sense of teachers' statistical thinking. Using a common assessment instrument, we identify and discuss four items that indicate strong statistical thinking and two that highlight concepts with which teachers struggle. We discuss potential course elements that may be contributing to areas of success. Implications for collaborative course (re)design and shared assessment items are discussed.*

Keywords: Data Analysis and Statistics, Teacher Education-Inservice/Professional Development

### Introduction

Over the past 30 years, a considerable body of research has addressed students' statistical thinking from primary through tertiary schooling (cf., Shaughnessy, 2007). Over the same time frame, the role of statistics and probability has received attention in the secondary curriculum in the United States (National Council for Teachers of Mathematics, 2000; Common Core State Standards Initiative, 2010). However, as Shaughnessy and others have noted, any effort to improve the teaching and learning of statistics in secondary schools depends on the statistical knowledge of secondary teachers. To this end, many graduate programs in mathematics education and professional development efforts have begun to include more opportunities for teachers to develop their own statistical reasoning abilities, and to learn about pedagogical issues in teaching statistics. But the research base on secondary teachers' statistical reasoning is sparse (c.f., Batanero, Burrill, & Reading, 2011), and thus limits the capacity for teacher educators to design effective courses and other professional development experiences for teachers. In this paper, we report on our emerging work on collaborative design of a course for secondary teachers that focuses on teaching and learning statistical reasoning.

Working at two different institutions, the authors identified several important commonalities in the statistics education courses that we had developed independently. These commonalities included shared foci on key elements of statistical reasoning, such as the central role of variation, distribution, and the sampling distribution in developing students' inferential reasoning. Our common goals for our courses also included a focus on the central role that statistical software designed for the learning of statistics (such as *Fathom* (Finzer, 2001) and *TinkerPlots* (Konold, 2005) should play in supporting teachers in their own learning of statistics and in developing their approaches to teaching statistics. We shared a common commitment to taking the time needed to develop teachers' understandings and, hence, in both courses "coverage" of topics was on occasion sacrificed for depth of understanding. Finally, as teacher educators working in statistics, we had a common goal of gathering and analyzing data that would provide evidence about teachers' statistical reasoning and provide an empirical basis for a collaborative course design on the teaching and learning of statistics. Thus, we are not conducting a comparative course study, but rather we seek to investigate how evidence on teachers' statistical reasoning from both courses can be the basis for an on-going collaborative design of an effective statistics course for secondary teachers. In this paper, we focus on the following question: *How can assessment items on statistical reasoning provide a basis for the (re)design of learning experiences for teaching statistics at the secondary level?*

## Theoretical Grounding

Recently, researchers have discussed and investigated teachers' statistical knowledge for teaching, using various frameworks and perspectives in their work (e.g., Groth, 2007, Burgess, 2011; Lee & Hollebrands 2011; Noll, 2011). Each of the emerging frameworks has identified teachers' own statistical literacy and thinking as a foundational or cornerstone aspect of their ability to teach statistics. Not many would argue this point. Though our courses aim to more fully develop secondary teachers' understanding for teaching statistics, the focus in this paper is on several key concepts in statistics and teachers' reasoning abilities with these concepts, without regard to their understanding of how to teach these concepts. If statistical literacy and thinking are essential for teachers, then teachers should have a strong foundation that is above that of students who have taken a collegiate level introductory statistics course, most of which are at the level of an AP statistics course taught in high schools. Thus, we turned to literature on assessing statistical literacy, thinking and reasoning (e.g., Garfield & Chance, 2000) for assistance in how to best assess the literacy and thinking of our teachers. Several researchers have engaged in developing, validating, and administering test items that aim to assess conceptual understanding and statistical literacy and thinking, rather than skills. One coordinated effort (delMas, Garfield, Ooms, & Chance, 2007) resulted in the Comprehensive Assessment of Outcomes in a First Statistics course (CAOS, <https://apps3.cehd.umn.edu/artist/caos.html>) and the collection of items at the ARTIST website (Assessment Resource Tools for Improving Statistical Thinking).

## Methods

### Course Contexts and Participants

**Course1.** Authors 2 and 3 designed and taught a one-semester graduate-level course in mathematics education to engage teachers with a range of tasks involving the investigation and exploration of statistical concepts using the software package *Fathom* (Finzer, 2001). The statistical content of the course consisted of investigations into variation and distribution, sampling distributions, confidence intervals, and inferential statistics. In addition, the course included various readings and discussions about (a) the nature of statistical reasoning and how it compares to other forms of mathematical reasoning and about (b) secondary students' learning and statistical reasoning. *Fathom* was used to support teachers' learning by providing an interface that would allow them to flexibly explore multiple graphical representations (e.g. shifting between box plots, dot plots and histograms) while being able to easily compare data sets, and to make changes to the data so as to explore conjectures. *Fathom* also provided the simulation tools necessary to create sampling distributions and representations of a population, a sample, and the sampling distribution. We saw this as critical to developing the teachers' knowledge of sampling, in order to build an understanding of formal inference.

There were 13 teachers who participated from Course1. Four participants were pre-service teachers (in a graduate licensure program), five were in-service teachers, two were in a masters program full-time, and two were doctoral students in mathematics education. Eight of the participants were female and five were male. All participants had completed the equivalent of an undergraduate major in mathematics, with all but one having had at least one course in statistics.

**Course2.** At a different university, another graduate-level, semester-long course in mathematics education was designed and taught by Authors 1 and 4. The course was intended for secondary and tertiary teachers of introductory statistics courses. The course had similar content focus as Course1, except that the instructors deemed it was necessary to go deeper in several explorations with data early on, and thus confidence intervals and formal inference were not discussed. Teachers in Course2 also engaged in similar readings and discussions as those in Course1. *TinkerPlots* was the main tool used in the course, though teachers also had some experiences with *Fathom*. The choice of

*TinkerPlots* was purposeful to engage teachers in more exploratory data analysis and to emphasize reasoning about distributions and statistical measures in multiple visual ways.

The 16 teachers in Course2 consisted of two preservice teachers enrolled in an M.A.T. program; eight teachers in a masters program with six currently teaching (3 of whom taught AP Statistics) and two in graduate school full-time; six PhD students in Mathematics Education, all of whom had recent secondary teaching experience; and one Statistics PhD student currently teaching a college-level introductory statistics course. Twelve teachers were female and four were male, with three international students within their first year of graduate work in the US. All teachers had completed an equivalent of an undergraduate degree in mathematics or statistics, and all had taken at least a first level graduate course in statistics for social science majors (content about equivalent to Advanced Placement Statistics).

**Instrument**

In the final week of both courses, all participants completed a 20 item multiple choice test with items in six categories: graphical representations, sampling variation, inference, data collection and design, bivariate data and probability. This selection of categories represents key concepts in introductory statistical literacy and thinking, and potentially provides us with a range of topics for which we can use as a basis for the continuing (re)design of our courses. Most items were drawn from validated instruments such as the CAOS test, with a few selected from the larger ARTIST database (<https://apps3.cehd.umn.edu/artist>), and one item drawn from the work of Zieffler, et al. (2007) because of its focus on informal inference (emphasized in both courses). All items have been used previously with college students, and we have known results for 19 of the 20 items that would allow for a comparison to the results from our teachers.

**Analysis**

We were interested in identifying the common successes and struggles with statistical reasoning displayed by our teachers, across both courses and institutions, as evidenced in their responses to this 20-item assessment. The common successes and struggles would provide us with some evidence about the areas where teachers had strong understandings of the measured concept and areas where their struggles (known from the research on college students) persisted through to the end of both courses. This evidence in turn becomes the basis for the (re)design of the courses in ways that will build on the strengths of teachers’ statistical reasoning and will address areas of persistent difficulty. Thus, we first focused our attention on the items on which at least 75% of the teachers chose the correct response in both classes. We then attended to items that both sets of instructors felt assessed concepts they had spent considerable time developing in their courses, yet the response rates did not show evidence of strong statistical reasoning for either group of teachers. Due to space limitations, we discuss four questions from the strong statistical thinking category, and two items from the struggling statistical thinking category.

**Results**

The overall performance of our teachers on six selected items from the post-test is shown in Table 1. For comparison purposes, the last column indicates performance results of students from introductory statistics courses from the work of others. In all six items, our teachers demonstrated much higher abilities in statistical thinking.

**Table 1. Comparative Results for Percent Correct on Selected Items on Post-test.**

Item	Measured Learning Outcome	Course1 N=13	Course2 N=16	Comparison to Others’ Results (N varies)
------	---------------------------	-----------------	-----------------	--

Martinez, M. & Castro Superfine, A (Eds.). (2013). *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Chicago, IL: University of Illinois at Chicago.

Category: Strong Statistical Thinking				
4	Ability to correctly estimate standard deviations for different histograms. Understands highest standard deviation would be for graph with the most spread (typically) away from the center.	92%	75%	46.9% <sup>a</sup>
7	Ability to use a frequentist approach to estimate probability of events for tossing an irregular shaped object, rather than apply an equiprobable approach	100%	75%	59% <sup>b</sup>
14I	Understanding of the law of large numbers for a large sample by selecting an appropriate sample from a population given the sample size.	100%	100%	65.2% <sup>a</sup>
14II	Ability to select an appropriate sampling distribution for a population and sample size.	77%	75%	44% <sup>a</sup>
Category: Struggling Statistical Thinking				
3	Understanding that a distribution with a median larger than the mean is most likely skewed left.	46%	63%	39.7% <sup>a</sup>
5	Understanding that statistics from small samples vary more than statistics from large samples.	54%	63%	31.9% <sup>a</sup>
<sup>a</sup> Results from CAOS administration reported by delMas et al. (2007) with participants (N) varying per item from 724-749.				
<sup>b</sup> Results reported from ARTIST database (delMas, personal communication, January 31, 2013).				

**Evidence of Strong Statistical Thinking Across Groups**

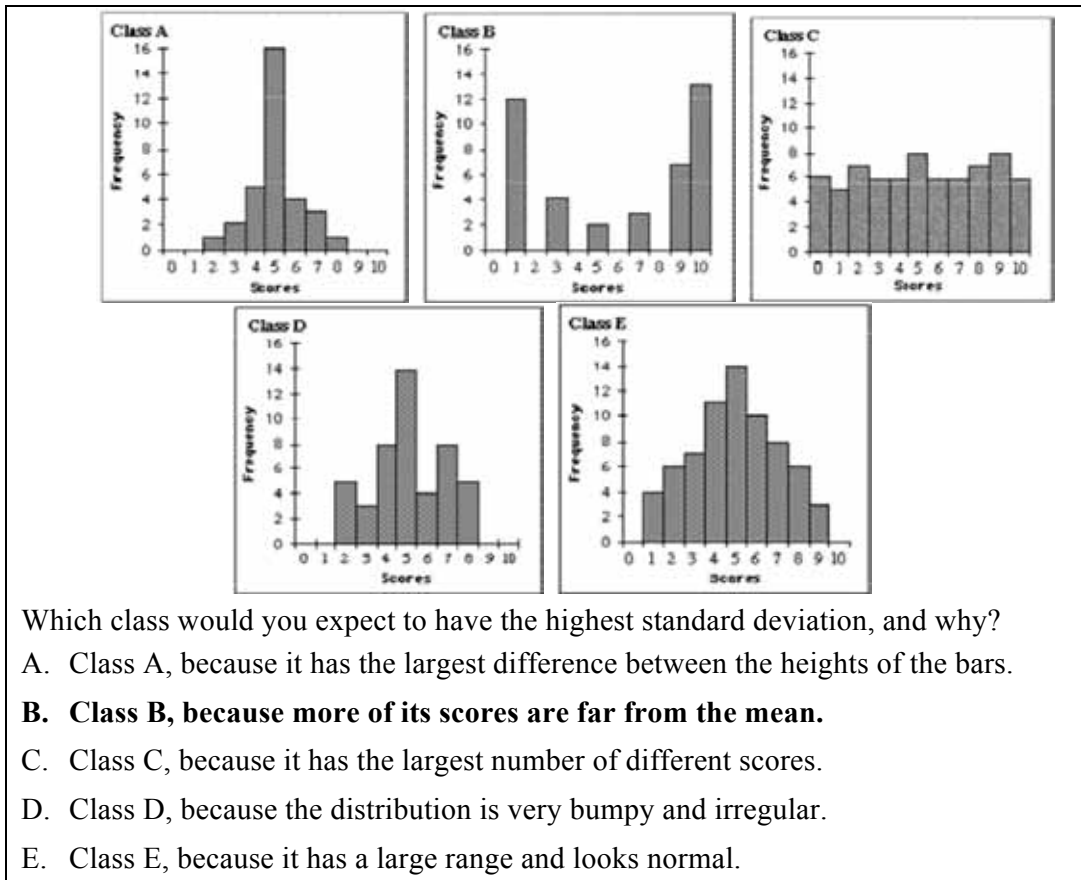
Most teachers in both groups (92% and 75%) were able to identify Class B (Figure 1) as the distribution that would likely have the largest standard deviation, because more of its scores are far from the mean. The concept of standard deviation was discussed explicitly in each course and teachers had an opportunity to examine research items similar to this in their reading and discussion of the work of delMas and Liu (2005). It is promising that these teachers tend to reason much better than students who have completed an introductory statistics class, of which less than 50% correctly responded to this item. Thus, explicit attention to reasoning about standard deviation of a distribution and reading about students’ reasoning with this concept may help support teachers’ reasoning.

Our teachers have no problem identifying a distribution for a single large sample chosen from a given population (100% correct in both courses), but did slightly less well on their ability to choose a correct graph representing a distribution of 500 sample means (Figure 2). On both parts of the item, our teachers did much better than the typical student as reported from delMas et al. (2007). This is especially encouraging since teaching the difference between a sample and a distribution of sample statistics is a particularly important, and hard to teach, concept in introductory level statistics at the collegiate and high school level (Saldanha & Thompson, 2002). Both courses spent considerable time engaging with these concepts using technological tools and reading literature about typical students’ difficulties with distinguishing between a sample and a distribution of sample statistics.

Five histograms are presented below. Each histogram displays test scores on a scale of 0 to 10 for one of five different statistics classes.

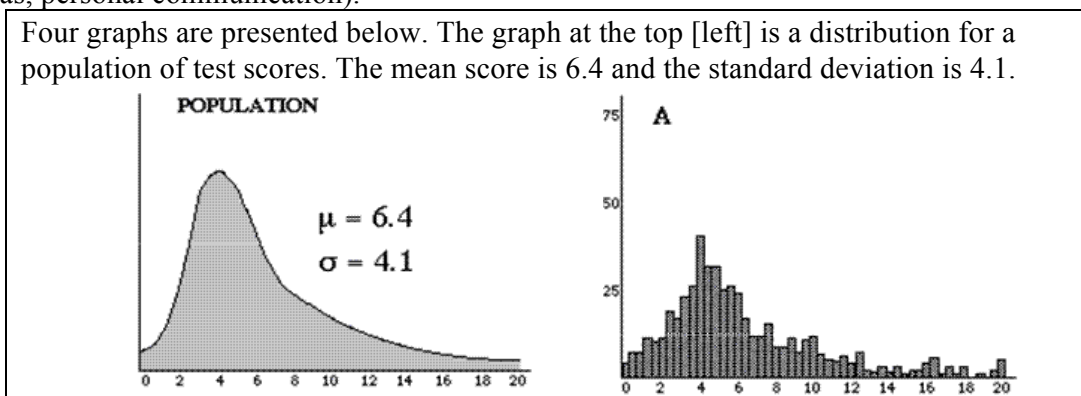
---

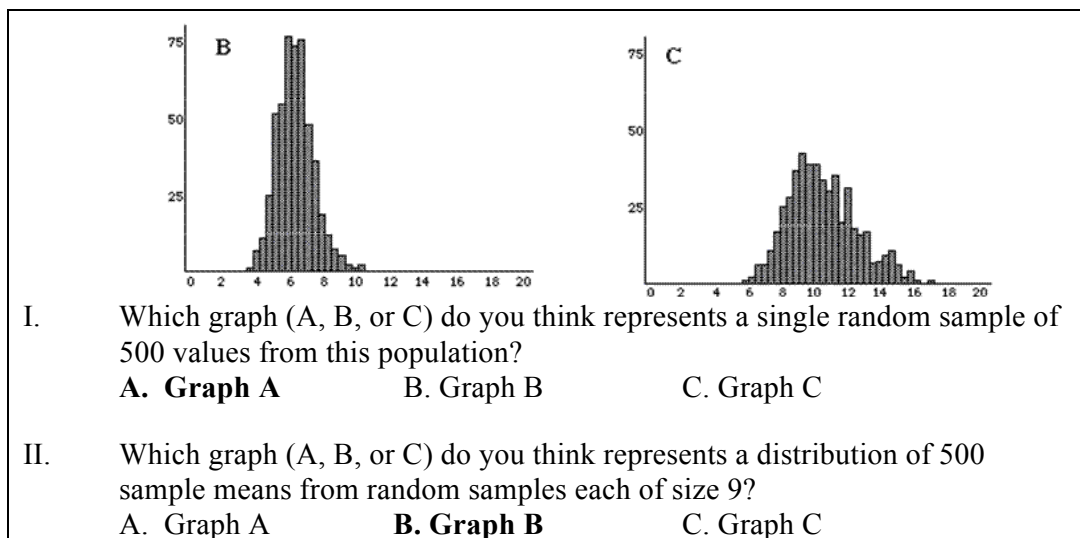
Martinez, M. & Castro Superfine, A (Eds.). (2013). *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Chicago, IL: University of Illinois at Chicago.



**Figure 1. Item 4 – Choosing Distribution with Highest Standard Deviation**

Teachers in both groups also demonstrated a strong understanding of how to estimate probability of events for tossing an irregular shaped object using a frequentist approach (Figure 3). None of the teachers chose the assignment of probabilities based on a classical equiprobable approach (choice A). Three students from Course2 chose none of the above for their response; thus their reasoning for how to approach this task is unclear. If we want teachers to teach their students that an equiprobable distribution is not always the best estimate for the probability of events, then the high success rate on this item is particularly promising. Only 59% of students completing the same item in the ARTIST database chose the correct response, and 22% thought a classical approach (choice A) was the best (delMas, personal communication).





**Figure 2. Item 14 -- Understanding Sample Distribution and Distribution of Sample Means**

A game company created a little plastic dog that can be tossed in the air. It can land either with all four feet on the ground, lying on its back, lying on its right side, or lying on its left side. However, the company does not know the probability of each of these outcomes. They want to estimate the probabilities. Which of the following methods is most appropriate?

A. Since there are four possible outcomes, assign a probability of 1/4 to each outcome.

**B. Toss the plastic dog many times and see what percent of the time each outcome occurs.**

C. Simulate the data using a model that has four equally likely outcomes.

D. None of the above.

**Figure 3. Item 7 – Probability Measurement from a Frequentist Approach.**

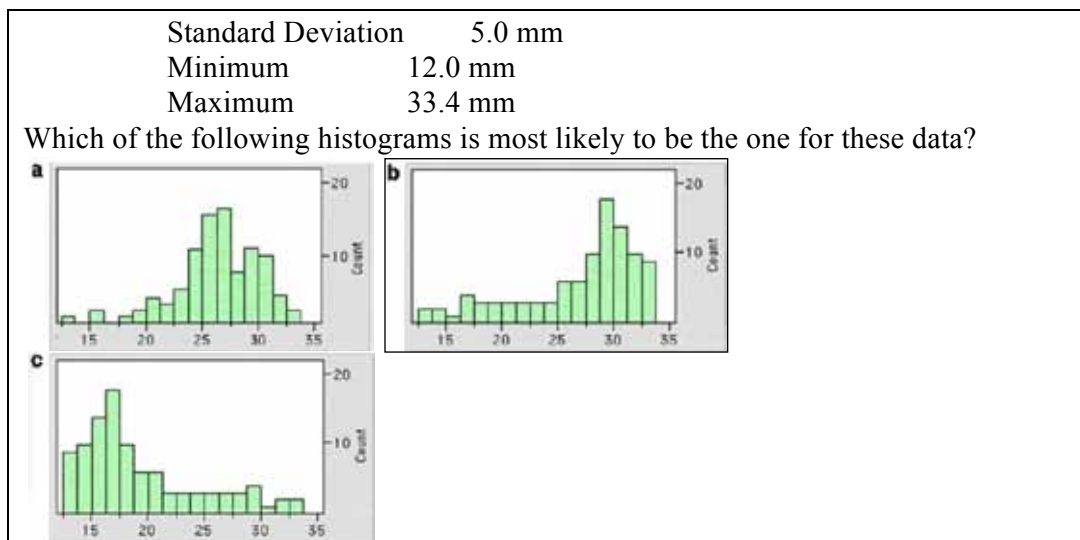
**Teachers’ Consistent Struggles Across Both Groups**

In both courses, significant attention was given to understanding the relationship between a distribution and its measures of center. The teachers in both courses engaged in exploratory data analysis with *Fathom* (Course1) or *TinkerPlots* (Course2) and experienced how dynamically moving a data value in a graph impacted the mean and deviations from a mean. In addition to the focus on standard deviation discussed earlier, both courses also included readings and discussions in which common conceptions of the mean were discussed (e.g., Shaughnessy, 2006; Zawojewski & Shaughnessy, 2000). Despite this focus, many teachers did not correctly choose the appropriate distribution (Histogram b) for the given statistical measures (see Figure 4).

A study examined the length of a certain species of fish from one lake. The plan was to take a random sample of 100 fish and examine the results. Numerical summaries on lengths of the fish measured in this study are given.

Mean	26.8 mm
Median	29.4 mm

Martinez, M. & Castro Superfine, A (Eds.). (2013). *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Chicago, IL: University of Illinois at Chicago.



**Figure 4. Item 3 – Choosing Distribution Given Statistical Measures**

Given our attention to the role of sample size in variation from expected proportions, we were surprised that many of our teachers did not respond correctly to item 5 (Figure 5). While our teachers did better as a group than introductory statistics students, the most common incorrect answer (E) chosen by introductory statistics students, was also the incorrect response chosen most often by our teachers. This choice demonstrates applying equiprobable reasoning to the situation rather than considering effect of sample size.

A certain manufacturer claims that they produce 50% brown candies. Sam plans to buy a large family size bag of these candies and Kerry plans to buy a small fun size bag. Which bag is more likely to have more than 70% brown candies?

A. Sam, because there are more candies, so his bag can have more brown candies.

B. Sam, because there is more variability in the proportion of browns among large samples.

**C. Kerry, because there is more variability in the proportion of browns among smaller samples.**

D. Kerry, because most small bags will have more than 50% brown candies.

E. Both have the same chance because they are both random samples.

**Figure 5. Item 5 – Understanding Role of Sample Size in Variability from Expected.**

### Discussion and Conclusions

Looking across the strengths and struggles of our teachers, two particular trends surfaced. First, teachers seemed to exhibit strong distributional reasoning when reasoning *from* graphs (items 4 and 14), but had difficulty predicting a graph of a distribution when given only statistical measures (item 3). Similarly, teachers were able to apply non-equiprobable approaches to a probability estimation task (item 7), but were prone to resort to applying equiprobable reasoning in a context about comparing likelihood of results from different sample sizes (item 5).

These results point to implications for us as course designers and for item and task development to measure our teachers’ reasoning more systematically. First, we want to be sure to include a stronger focus in our courses for teachers to reason from graphical representations of distributions, as

---

Martinez, M. & Castro Superfine, A (Eds.). (2013). *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Chicago, IL: University of Illinois at Chicago.

well as towards predicting graphs of data with certain statistical characteristics. We also need to draw teachers' attention to students' use of equiprobable reasoning and how it can interfere with probability judgments. Such course improvements can include more targeted readings and discussions, as well as purposeful task design, particularly with technology tools. Our results also point to a need for multiple choice items and open response tasks that may better assess teachers' (a) reasoning from and to graphs of data, and (b) ability to apply equiprobable reasoning appropriately. With our focus on using dynamic statistics software, we aim to develop new assessment tasks and items that take advantage of these tools.

### References

- Batanero, G. Burrill, & C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study* (pp. 259-270). New York: Springer.
- 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: A Joint ICMI/IASE Study* (pp. 259-270). New York: Springer.
- Common Core State Standards Initiative (CCSSI). (2010). *Common core state standards for mathematics*. Retrieved from <http://www.corestandards.org/the-standards/mathematics>
- delMas, R., Garfield, J. Ooms, A., Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28-58, <http://www.stat.auckland.ac.nz/serj>.
- delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, 4(1), 55-82. <http://www.stat.auckland.ac.nz/serj>.
- Finzer, W. (2001). *Fathom!* [Computer Software]. Emeryville, CA: KCP Technologies.
- Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education*, 38, 427-437.
- Konold (2005) *TinkerPlots*. [Computer Software]. Emeryville, CA: KCP Technologies.
- Lee, H. S., & Hollebrands, K. F. (2011). Characterizing and developing teachers' knowledge for teaching statistics. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Teaching Statistics in School Mathematics - Challenges for Teaching and Teacher Education: A joint study of the International Commission of Mathematics Instruction and International Association of Statistical Education* (pp. 359-369), Springer.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- Noll, J. A. (2011). Graduate teaching assistants' statistical content knowledge of sampling. *Statistics Education Research Journal*, 10(2), 48-74.
- Saldanha, L. A. & Thompson, P. W. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51, 257-270.
- Shaughnessy, J. M. (2006). Research on students' understanding of some big concepts in statistics. In G. Burrill (Ed.), *NCTM 2006 Yearbook: Thinking and reasoning with data and chance* (pp. 77-95). Reston, VA: NCTM
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *The Second Handbook of Research on Mathematics* (pp. 957-1010). Reston, VA: National Council of Teachers of Mathematics (NCTM).
- 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.
- Zieffler, A., delMas, R., Garfield, J., & Gould, R. (2007, August). *Studying the development of college students' reasoning about statistical inference*. Paper presented at the Fifth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-5), University of Warwick, UK.