INVESTIGATING SECONDARY TEACHERS' STATISTICAL UNDERSTANDINGS

Helen M. Doerr & Bridgette Jacob

Syracuse University

Most secondary teachers are familiar with the procedures of basic descriptive statistics, but they have not necessarily been prepared to interpret graphical representations of data or to reason about sampling distributions. In this exploratory research study, we investigate the understandings of eleven teachers who participated in a semester-long course, using Fathom, to develop their understanding of these concepts. We present the analysis of a pre- and post-test of content knowledge, the teachers' performance on two tasks, and their use of Fathom as a tool to simulate and represent sampling distributions.

INTRODUCTION

Over the last twenty years, researchers have made substantial progress in understanding students' conceptions of probability and statistics. Over this same time frame, there have been calls for a greater emphasis in schools on the inclusion of statistics topics throughout the curriculum (NCTM, 2000; Pfannkuch & Begg, 2004). However, a vision of improved teaching and learning of school statistics relies heavily on the knowledge and skills of teachers to enact instruction that engages students in developing statistical reasoning. Unfortunately, considerably less progress has been made in understanding teachers' statistical reasoning. As Shaughnessy pointed out in his recent review of the research on statistics learning and reasoning, "More research is needed on teachers' conceptions of statistics. Teachers have the same difficulties with statistical concepts as the students they teach." (2007, p. 1000). The broad goal of the research reported in this paper is to contribute to the research base on teachers' understandings of statistics, with an eye towards characterizing the knowledge that teachers bring to tasks involving graphical representations of data distributions, including sampling distributions. These understandings are key to developing a robust understanding of statistical inference, a topic that is taught at the upper secondary level in the United States. In particular, we are interested in two questions: (1) how do secondary teachers interpret graphical representations of data and (2) how do secondary teachers reason about sampling distributions?

BACKGROUND

The statistics background for most secondary mathematics teachers in the United States is very limited (Shaughnessy, 2007). Some pre-service teachers will have had limited formal coursework in statistics, and often their experience of this coursework is somewhat removed from the statistical content that they will need to teach in the secondary school. Such formal coursework in statistics would rarely address the specialized kinds of statistical knowledge that is needed for teaching that is different from just more statistical content, such as the potential misunderstandings that arise from students in the classroom (c.f., Shaughnessy & Chance, 2005). While many secondary teachers will be fluent with the procedures of descriptive statistics such as those in the study by Makar and Confrey (2004), teachers are likely to struggle with using graphical interpretations to data distributions. These researchers found that teachers were able to develop a robust understanding of distribution when working with data in a meaningful context, namely, interpreting test results for students, but the teachers encountered difficulty in distinguishing between the variability in a data set from that of a related sampling distribution. In their study with secondary teachers using *Tinkerplots*, Rubin and colleagues (Rubin et al., 2005) found that the shape of the distribution influenced the strategies that teachers used when comparing two distributions and that the teachers were more confident in their conclusions about symmetric distributions than about skewed distributions. As Pfannkuch (2006) points out, reasoning about data sets from their graphical representations is a complex task, requiring one "to attend to a multiplicity of elements within and between box plots, and to make judgements." (p. 29). For the teacher, the complexity of this task is further layered with the necessity of generating the kinds of talk that will communicate the concepts represented in box plots in ways that will build towards informal inference.

While the most widely available technological tool for data analysis in the secondary classroom in the United States is the graphing calculator, this tool is more limited in its capabilities for learning data analysis than currently available software. Software designed for the learning of statistics (such as *Tinkerplots* or *Fathom*) provides the learner with opportunities to flexibly explore the data. The "landscape-type" design (Bakker, 2002) of *Fathom*, in contrast to "route-type" software tools, does not assume a particular learning trajectory for the teachers, but provides many routes for exploration. In her study of secondary teachers' comparing distributions, Madden (2008) argues that route-type tools can scaffold teachers' learning of both statistical content and new technology environments by moving from physical experiment to route-type tools to landscape-type tools. In addition, land-scape tools have the potential to support an "expressive" approach to data modeling (Doerr & Pratt, 2008) that would allow teachers to create meaningful representations and interpretations of data and sampling distributions. In this study, we assume the reciprocal relationship between representations and models described by Rubin et al. (2005): "Not only does the model of data a person currently holds influence the representation she chooses to use, but the representation in turn influences the model of data she is developing." In this sense, the simulation and representational capabilities of *Fathom* can reveal the person's current way of thinking and support the development of that thinking.

DESIGN AND METHODOLOGY

This exploratory study was designed to gain insight into secondary teachers' knowledge about the graphical representation of data and sampling distributions. To this end, authors designed and taught a one-semester course that would 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. However, this study focuses only on the teachers' own statistical knowledge, not on their knowledge of students' learning.

The choice of *Fathom* was intended to support the 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 the population, the sample, and the sampling distribution. We saw this as critical to developing the teachers' knowledge of sampling.

There were 11 subjects who participated in this study. Eight of the participants were pre-service teachers, two were in-service teachers, and one was a doctoral student in mathematics education. Eight of the participants were female and three 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. All participants completed a 20 item pre- and post-test of their statistical knowledge in six categories: graphical representations, sampling variation, inference, data collection and design, bivariate data and probability. These items were drawn from the Comprehensive Assessment of Outcomes in a First Statistics course (CAOS, https://app.gen.umn.edu/artist/caos.html). These items have been used with college students, and this study extends those results to this group of teachers.

All participants completed two tasks prior to specific instruction that (a) required them to compare the standard deviation of two distributions, based on their graphs (delMas & Liu, 2005), and (b) to analyze the relationship between a population and a sampling distribution (Chance, delMas, & Garfield, 2004). We anticipated, as Shaughnessy (2007) alluded to, that the teachers would have many of the same difficulties as college students did with the CAOS items and with the tasks. Finally, the participants engaged with a *Fathom*-based activity investigating confidence intervals and how the number of samples and the size of samples affect the sampling distribution.

RESULTS

In this section, we first report on the results of the pre- and post-test. We then present the findings on two tasks on graphical interpretations. This is followed by a brief analysis of one participant's use of *Fathom* to represent her understanding of sampling distributions.

Post-Test Results

The post-test results suggest that there was an overall improvement in the teachers' understanding of the statistical concepts, as measured by the 20 items on the test, and this gain was significant ($p<0.05$), as shown in Table 1, $n = 11$. The only sub-area of the test where there was a significant difference from the pre-test to the post-test was in the area of graphical representations with six items. This result likely reflects the emphasis given to graphical representations and the extensive use of *Fathom* within the course.

	pre-test mean	post-test mean	difference	p-value
overall	11.09	12.54	1.45	0.027
graphical representation	4.09	5.00	0.91	0.033

Table 1. Pre-post test results for overall concepts and graphical representations.

The first item among the six graphical representation items addressed the ability to describe and interpret a distribution displayed in a histogram. There was no change in this item from the pre- to the post-test, with 82% correct. The persistent error was misinterpreting the magnitude of the standard deviation of a near normal distribution as too small. Performance on the second item increased from 64% correct to 73% correct; the most common error was tending to select a normal distribution that did not make sense in the context of the problem. Items three and four addressed the ability to interpret the median and the quartiles in a box plot; these two items went from 73% to 91% correct and 55% to 82% correct. The error made on item four was incorrectly reasoning that the boxplot with a longer upper whisker would have a higher percentage of data above the median. On item five, which tested the understanding that a distribution with the median larger than the mean is likely skewed to the left, the correct response rate went from 64% to 82%. All participant errors on item five (both pre- and post-test) were incorrectly selecting a somewhat symmetric, mound-shaped graph. On item six, which addressed the ability to estimate standard deviations for different histograms, there was a change from 73% to 91% correct. Taken together, these results seem to suggest a tendency for some teachers to have more difficulty interpreting skewed distributions than symmetric distributions and a tendency to incorrectly choose a normal distribution.

Graphical Interpretations of Standard Deviation

Early in the course, we gave the participants the task from delMas and Liu (2005) on interpreting standard deviations graphically to the participants. Three of these items (#4, 8, and 10) are shown in Figure 1. The mean and the standard deviation are displayed for the graph on the left, but only the mean is displayed for the graph on the right. The participants were asked to determine whether the standard deviation for the graph on the right was greater than, less than, or equal to the standard deviation for the graph on the left. As delMas and Liu point out, test item 4 "was specifically designed to see if students understood that given the same frequencies and range, a distribution with a stronger skew tended to have a larger standard deviation" (2005, p. 63). According to delMas and Liu, test items 8 and 10 were "designed to challenge the belief that a perfectly symmetric and bell-shaped distribution will always have a smaller standard deviation. Students were expected to find these items more difficult than the others." (2005, p.63). Unlike the students in delMas and Liu's study, the participants in our study had more difficulty with item 10 than item 8.

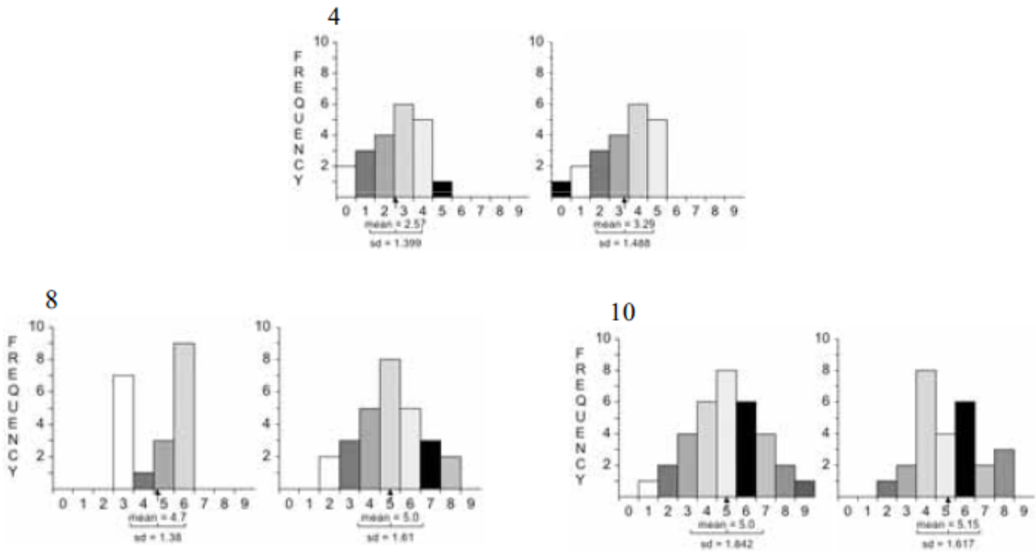

Figure 1: Comparing standard deviations graphically.

Of the ten participants who completed this task, four incorrectly answered that the standard deviations of the two distributions shown in item #4 were equal. This suggests that these participants are not attending to how skew affects the standard deviation. For item 8, four of our participants incorrectly concluded that the graph on the right had a smaller standard deviation, despite the fact that the graph on the left has a smaller range and represents a smaller number of values. This suggests that these participants might have reasoned that the symmetry of the bell-shaped curve with a large portion of the density centered about the mean would result in a smaller standard deviation.

For item 10, the graph on the right appears to have less density around the mean, but at the same time, it has a smaller range and represents a smaller number of values. As delMas and Liu point out, a reasonable response to both items 8 and 10 could be that the standard deviations needed to be calculated in order to determine how they differ. In their study, many of the students used calculations to come to a correct answer. Those who did not calculate came to the same incorrect conclusion as nearly all of the participants in this study. Only one of the ten participants answered this item correctly. Seven concluded that the distribution on the right had the greater standard deviation. One of the remaining two students concluded that the standard deviations were equal and the other answered with a question mark. These results suggest that these participants might be assuming that a symmetric normal distribution minimizes the standard deviation. The results also point to the difficulties in determining the standard deviation from a graph when having to interpret combined effects of density about the mean, range, and frequency.

Graphical Interpretations of Population Distributions and Samples

We asked our participants to compare the shape and the variability of a sampling distribution to a population, based on a task described in Chance, delMas and Garfield (2004). Prior to specific instruction, we asked our participants which of the graphs in A through E represented a distribution of sample means for 500 samples of size 4 and of size 16, based on the population distribution shown in the upper left in Figure 2. We asked them to state whether these sampling distributions would have less, more or the same variability as the population and as each other. The results of this task are shown in Table 2 and Table 3.

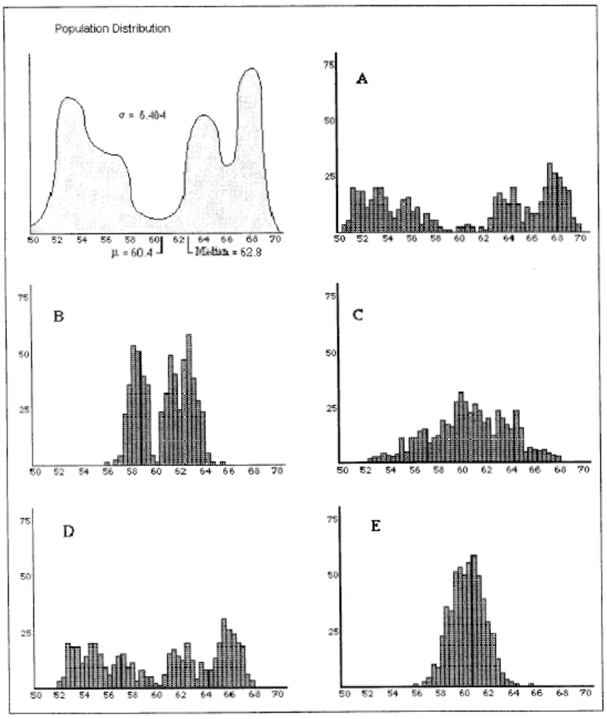

Figure 2: Population distribution from Chance, delMas, & Garfield (2004), p. 321

Only 4 of the 11 participants chose the correct response (*C*) when asked for the distribution of sample means for 500 samples of size 4. When asked for the distribution of sample means for 500 samples of size 16, only 5 of the 11 participants chose the correct response (*E)*. For both items, the majority of participants chose distributions that indicate a belief that the sampling distribution should look like the population. This is a common misconception held by students (Chance, delMas & Garfield, 2004). The choice of response *B* for the second item would indicate that some participants believed that the sampling distribution continued to look like the population as the sample size increased, but with reduced the variability.

responses	*A*	*B*	*C*	*D*	*E*
sample means of size 4	2	0	4	5	0
sample means of size 16	2	4	0	0	5

Table 2: Possible distributions of sample means drawn from a population

As shown in Table 3, most participants (7 out of 11 and 10 out of 11) correctly compared the variability of the samples of size 16 to both the population variability and the variability of the samples of size 4. However, when comparing the variability of samples of size 4 to the variability of the population, 5 of the participants incorrectly expected the samples of size 4 to have *more* variability. This misconception is particularly interesting given that nearly all of the participants correctly compared the variability of a sample of size 4 to a sample of size 16. We speculate that these participants might be confusing the variability of a single sample with the variability of the sampling distributions. This needs further investigation.

responses	*less*	*same*	*more*
samples of size 4 compared to population	6	0	5
samples of size 16 compared to population	7	3	1
samples of size 16 compared to samples of size 4	10	0	1

Table 3: Comparing the variability of the sample distributions

Representing the sampling distribution using *Fathom*

Given the difficulties that a number of participants had with the previous task, we designed a task using *Fathom* to help participants better understand the three levels of abstraction that are present in a sampling simulation: (1) the population; (2) a particular sample from the population; and (3) the collection of measures that result from repeated sampling. This task consisted of two parts: the first part of the task asked the participants to set up a simulation in *Fathom* and investigate how the sampling distribution of a summary statistic compared to the population distribution and how changing the number of samples and the sample size affects the sampling distribution. The participants successfully completed this first part of the task using a collection of 100 random rectangles (Key Curriculum Press, 2002, p. 127).

The second part of the task asked the participants to create a display that could be used to help students understand the difference between the population distribution, the sample, and the sampling distribution. This creation of a display that could be understandable to a learner was a substantial shift in the nature of the task that the participants were engaged in. Rather than convincing themselves about the relationships involved, the participants now shifted their attention to create a representation that could be used to illuminate the relationships among the population, a sample and the sampling distribution of a particular statistic. The work of one pre-service teacher is shown in Figures 4 and 5 and is representative of the kind of displays most participants created.

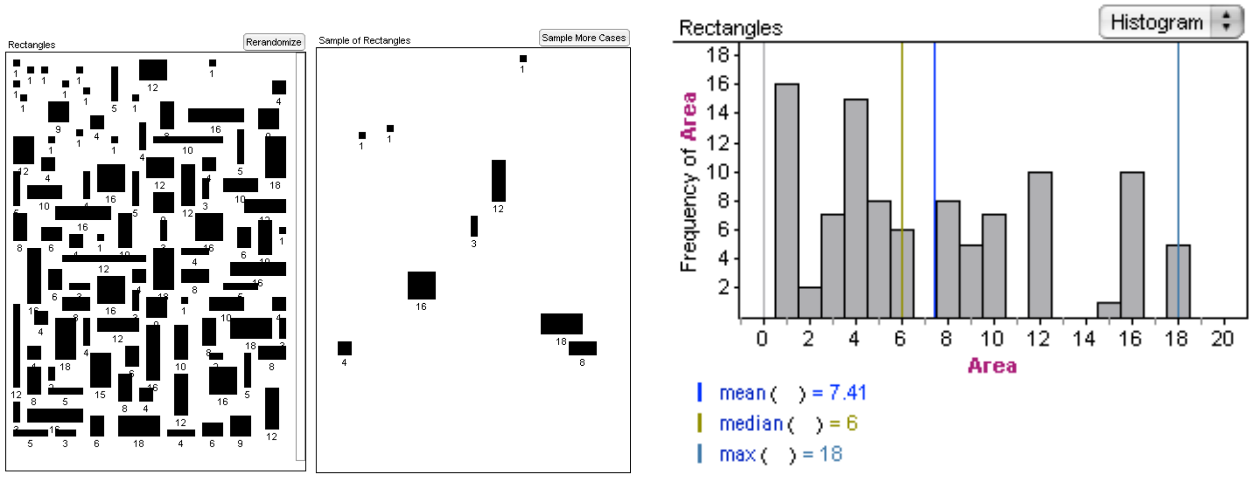

Figure 4: Juxtaposing the population, the sample and the population distribution

This pre-service teacher displayed the entire population of 100 rectangles, next to a randomly chosen sample of 10 rectangles along with a histogram showing the distribution of the areas along with the summary statistics for the entire population (as shown in Figure 4). The next section of her display (shown in Figure 5) juxtaposed a table that summarized the dimensions of each of the ten rectangles in the randomly drawn sample and the distribution of the areas for that sample. Just below this, she positioned a table (lower left in Figure 5) that showed the mean, median and maximum area for each sample of 10 rectangles up to 101 such samples. Next to the table is the graph of the distribution of the sample means. The pre-service teacher created this display with the animation feature of *Fathom* turned on so that she could see the sample distribution changing with each sample of 10 rectangles, while the sampling distribution was slowly being built and taking on the shape of a normal distribution. All participants introduced dynamic elements of animation into their displays, which are not fully captured by static snapshots.

This display suggests that this pre-service teacher was separating the three levels (or tiers as Madden (2008) refers to them) of abstraction that one needs to understand in order to grasp the concept of the sampling distribution. By using the simulation capabilities of *Fathom*, along with its flexibility in selecting representations, this pre-service teacher has clarified the distinction between the population, the sample, and the collection of sample means. The animation feature seemed to make visible how the sampling distribution is built over time as samples are collected. This particular display has the potential to help the pre-service teachers avoid confusing the population distribution with the sampling distribution as seen in the second task reported above. On a related item on the pre- and post-test, that asked participants to select an appropriate sampling distribution for a particular population and sample size, we found that the response rate went from 45% to 64% correct.

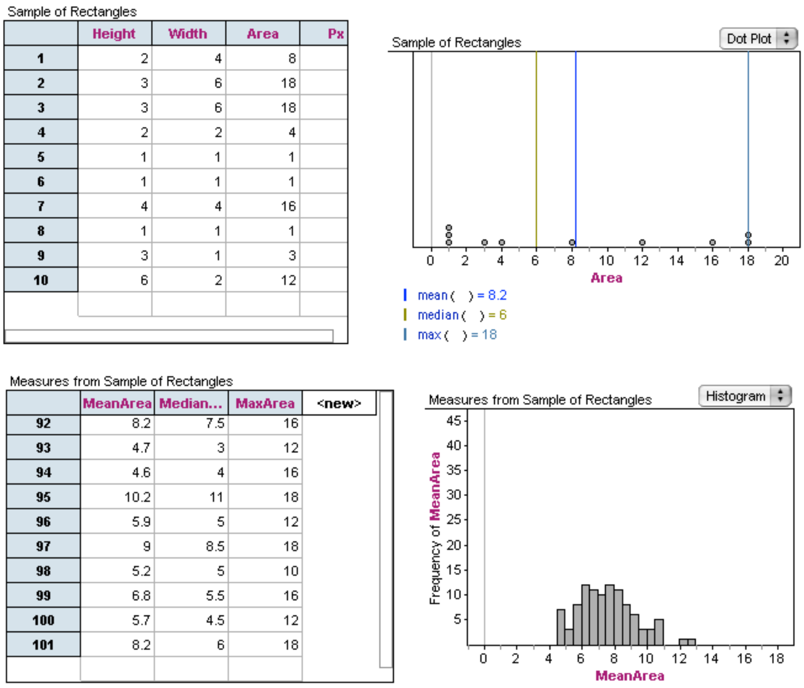

Figure 5: A snapshot of a dynamic display building the sampling distribution

DISCUSSION AND CONCLUSIONS

The results of this study provide some evidence that, as Shaughnessy (2007) argued, teachers have some of the same difficulties with statistics as do students. We note that a limitation of this study is the small number of participants and that while there was a statistically significant gain on the post-test, this gain was small and largely in the area of graphical representations. Consistent with the findings of Makar and Confrey (2004) and Rubin et al. (2005), we found that some teachers had difficulty interpreting skewed distributions and tended to inappropriately choose symmetric normal distributions. This was evidenced when having to interpret how the combined changes in density about the mean, range and frequency affected the standard deviation. It is likely that these secondary teachers' prior learning of statistics focused largely on the computational algorithm for standard deviation. Extending the results of Chance et al (2004), we found some teachers, like students, had a tendency to see the sampling distribution as having the same shape as the population distribution. Most teachers correctly reasoned about the variability of the distribution of larger samples, but incorrectly expected greater variability in the sampling distribution of smaller samples when compared to the distribution of the population. We found nearly all participants were able to create animated displays with *Fathom* that brought clarity to the three levels of abstraction that are present in any sampling distribution. This suggests that a land-scape tool such as *Fathom* has the potential to make visible teachers' models of statistical concepts.

REFERENCES

Bakker, A. (2002). Route-type and landscape-type software for learning statistical data analysis. In B. Phillips (Ed.), *Developing a statistically literate society: Proceedings of the Sixth International Conference on Teaching Statistics.* Voorburg, The Netherlands: International Statistical Institute.

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking*, (pp. 295-323). Dordrecht, Netherlands: Kluwer Academic Publishers.

delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal, 4*(1), 55-82.

Doerr, H. M., & Pratt, D. (2008). The learning of mathematics and mathematical modeling. In M. K. Heid & G. W. Blume (Eds.), *Research on technology in the teaching and learning of mathematics: Syntheses and perspectives: Vol. 1. Mathematics Learning, Teaching and Policy*, pp. 259-285. Charlotte, NC: Information Age.

Finzer, W. (2001). *Fathom Dynamic Data Software.* (Version 2.1) [Computer Software]. Emeryville, CA: Key Curriculum Press.

Key Curriculum Press. (2002). *Exploring statistics with Fathom*. Emeryville, CA: Author.

Madden, S. (2008). Dynamic technology scaffolding: A design principle with potential to support statistical conceptual understanding. Paper presented at the *11th International Congress on Mathematical Education*, 6-13 July. Monterrey, Mexico. Retrieved from http://tsg.icme11.org/tsg/show/15

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, Netherlands: Kluwer Academic Publishers.

Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal, 5*(2), 27-45.

Rubin, A., Hammerman, J. K., Campbell, C., & Puttick, G. (2005). The effect of distributional shape on group comparison strategies. In K. Makar (Ed.), *Reasoning about distribution: A collection of current research studies* (SRTL-4), University of Auckland, NewZealand, 2-7 July. Brisbane: University of Queensland.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 2, pp. 957-1009). Charlotte, NC: Information Age Publishing

Shaughnessy, J. M. & Chance, B. (2005). *Statistical questions from the classroom.* Reston, VA: NCTM.