Identifying Statistical Concepts Associated with High and Low Levels of Self-Efficacy to Teach Statistics in Middle Grades

Leigh M. Harrell-Williams
University of Memphis

M. Alejandra Sorto
Texas State University

Rebecca L. Pierce
Ball State University

Lawrence M. Lesser
The University of Texas at El Paso

Teri J. Murphy
Northern Kentucky University

Journal of Statistics Education Volume 23, Number 1 (2015),
www.amstat.org/publications/jse/v23n1/harrell-williams.pdf

Key Words: SETS instrument; GAISE; Common Core State Standards for Mathematics (CCSSM); Pre-service teachers; Middle grades.

Abstract

Previous mathematics and science education research indicates that knowledge and beliefs, including teaching efficacy, affect teachers' actions and effectiveness in a classroom. Our middle grades and high school Self-Efficacy to Teach Statistics (SETS) instruments, aligned with the statistical concepts in national and state guidelines such as the *GAISE Pre-K-12 Report* and the *Common Core State Standards for Mathematics* (CCSSM), were developed for use in teacher education research. This study focuses on the middle grades SETS instrument, which measures pre-service teachers' self-efficacy to teach topics at GAISE levels A and B as well as K–8 CCSSM statistics topics. The items ask teachers to rate their self-efficacy to teach a particular

concept on a Likert scale from 1 ("not confident at all") to 6 ("completely confident"). Data were collected at four public institutions of higher education in the United States. Rasch modeling was used to order the items by difficulty of endorsement to gain knowledge regarding pre-service teacher perceptions of difficulty, with the goal of identifying priorities for increasing pre-service teachers' self-efficacy with statistical topics.

1. Introduction

The *Mathematical Education of Teachers II report* (Conference Board of the Mathematical Sciences 2010) recommends that the statistical preparation of future Pre-K-12 mathematics teachers "focus on the data collection, analysis, and interpretation needed to teach the statistics outlined in the [Common Core State Standards] CCSS" (p. 18). This recommendation is supported by the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Pre-K-12 Report* (Franklin et al. 2007) in the sense that it provides a developmental path of these essential statistical ideas and by the CCSSM document itself, which provides the expectations in terms of content and mathematical practices.

Even though teacher preparation programs may design and implement courses aligned with these recommendations, pre-service teachers do not always leave teacher preparation programs with the desired knowledge and beliefs (Pfannkuch and Ben-Zvi 2011; Pierce and Chick 2011). An example of this is illustrated by the hypothetical vignette (Lesser 2013) in which a teacher has difficulty discussing nuances of the measures of central tendency in response to students' reasonable questions, even though he has presented the students with textbook definitions for each measure. For this reason, many mathematics and statistics educators are interested in measuring knowledge and beliefs as well as the extent to which they influence teaching practices and student achievement. More specifically, research about teacher efficacy in mathematics and science education indicates that this construct is related to teachers' content knowledge, pedagogical content knowledge, and beliefs and attitudes regarding content (Çakiroğlu 2000; Gresham 2008; Huinker and Madison 1997; Swars 2005; Wenta 2000). In statistics education, there are no studies that give evidence of these relationships, but the work has started with the development of the Self-Efficacy to Teach Statistics (SETS) instruments by the authors (Harrell et al. 2009; Harrell-Williams, Sorto, Pierce, Lesser, and Murphy 2014).

As a first step to establishing relationships between teachers' statistical knowledge and beliefs and the measurement of their impact on teaching practices, we examine the extent to which SETS might be utilized in determining statistical topics that future teachers feel more efficacious or less efficacious to teach.

2. Studies of Teachers' Self-efficacy to Teach Statistics

Studies related to teachers' self-efficacy to teach statistics are very limited. Estrada, Batanero, and Lancaster (2011) described two studies related to teacher attitudes towards their ability to teach statistics. The first study was conducted with 34 Australian in-service and pre-service elementary teachers (Begg and Edwards 1999). Their framework included teachers' attitudes and beliefs about statistics, teachers' statistical knowledge, and teachers' attitude and beliefs about teaching statistics. As part of this last construct, Begg and Edwards investigated teachers' confidence by asking how they felt about teaching statistics. They found that teachers' lack of confidence to teach certain concepts was related to their lack of familiarity with the newer ideas in their statistics curriculum, such as stem and leaf graphs and probability. At the other end, teacher confidence was highest on the more familiar topics such as data collection, bar graphs, and pictographs. The study also investigated, qualitatively, the relationship between teacher confidence level and their perception of the essential knowledge for teaching. They found that "teachers themselves do not necessarily see their lack of statistical education as a problem for their teaching." (Begg and Edwards 1999, p. 9). Further, the teachers appear more concerned about the pedagogy than gaining more statistical professional development.

The second study also involved Australian teachers. Watson (2001) surveyed 43 elementary and secondary teachers with the purpose of profiling teachers' competence and confidence to teach statistics and probability. Watson found that teacher confidence was highest for "graphical representations" and lowest for "odds." We note that both of the aforementioned studies used similar methodology to investigate the level of teachers' confidence, including interviews and surveys. Given the overall goal of both studies, their surveys were not designed to provide a single measure of teachers' confidence but instead a holistic assessment of teachers' characteristics.

Our study adds to the body of existing research on teacher efficacy to teach statistics in two distinct ways. The previous studies were conducted prior to the development of the GAISE and CCSSM and were based on statistical concepts in the Australian primary school curriculum in the mid-1990s. Our study measures efficacy to teach statistical concepts that may not have been captured in the previous studies, but that are currently relevant to recent changes in the K-12 curriculum in the United States. Additionally, the type of teachers targeted in the studies differs. The two previously mentioned studies focused on smaller samples of mostly Australian in-service teachers; only 12 pre-service teachers were included in the first study. Some of these did not have formal statistics training. Our study focuses on pre-service teachers in the United States, who have completed some formal training in statistics as part of their teacher preparation program.

3. Methods

3.1 Research Questions

Of the statistical topics on the middle grades Self-Efficacy to Teach Statistics (SETS) instrument, which are pre-service teachers most likely to indicate that they have a higher level of confidence to teach? Correspondingly, which topics are they most likely to indicate that they have a lower level of confidence to teach?

3.2 SETS Instrument for Middle Grades

The middle grades Self-Efficacy to Teach Statistics (SETS) instrument asks pre-service teachers to rate their "confidence in teaching middle grades students the skills necessary to complete" the 26 tasks using a six-point Likert scale, where 1 indicates "not at all confident" and 6 indicates "completely confident." The items are based on the GAISE Pre-K-12 Level A and B items and align with the CCSSM for grades 5 to 8. An informal author-conducted review of state standards for student and teacher knowledge for middle school mathematics that were available online in 2007, as well as the GAISE Pre-K-12 document, led to the selection of representative behaviors to be measured by the middle grades SETS instrument. While the items were written prior to the initial release of the CCSSM, the middle grades SETS items align with the CCSSM, partially due to the influence of the GAISE authors during the development of the CCSSM. The terminology used in the items was evaluated by 12 in-service elementary and middle school teachers who were enrolled in a Mathematics Specialist program, with their suggestions leading to some changes in item wording (Harrell et al. 2009; Harrell-Williams et al. 2014).

A sample of items from the SETS instrument is presented in the results section and the entire middle grades SETS instrument is presented in the Appendix. Evidence supports the existence of two subscales: "Reading the Data" and "Reading Between the Data." Reliability of separation estimates for the subscales scores and the composite efficacy score are .87, .91 and .94, respectively. The correlation between the subscales is estimated to be .85. Harrell et al. (2009) and Harrell-Williams et al. (2014) describe the development and validation of the scores from this instrument, using a sample of pre-service teachers working towards certification that included middle grades mathematics.

3.3 Sample

A sample of n = 309 pre-service teachers was obtained from four large-sized public institutions of higher education in the United States with teacher education programs. The universities were chosen to represent a variety of the types of certification programs for middle school mathematics (see Table 1). The participants were enrolled either in an introductory statistics course or in a designated mathematics education course, and were recruited because they were pursuing a teacher education degree. The pre-service teachers were approached in their classes at the end of the semester to complete the SETS instrument on a voluntary basis, making this a convenience sample. Approximately 78% of the sample was female. Participants reported the following ethnicities: White (88.3%), Hispanic (5.9%), Black (1.3%), Asian/Pacific Islander (0.6%), American Indian/Alaskan Native (0.3%), Biracial (1%), with a 2.6% non-response rate. The pre-service teachers in the sample were at various stages in their teacher education program; the majority of those who responded indicated that they had the preliminary type of certification status. Approximately 84% of pre-service teachers reported an intended certification that covered one or more of the middle grades years, with an 11.3% non-response rate. Of the 88.7% who responded to this item, 94.2% reported an interval including one or more of the middle grades years. Although about 2/3 of the teachers had taken one college-level course in statistics, none had taken a course in statistics pedagogy.

Table 1. Type of Teacher Certification Program by Institution

Institution	Major	Department offering courses	Teaching Certification (Grades)
#1	Middle School Mathematics Specialist	Mathematics	4 – 8
#2	Elementary Education	Mathematics and Statistics	P – 5
	Middle Grades Education	Mathematics and Statistics	5 – 9
	Mathematics	Teacher Education	8 – 12
#3	Middle School Mathematics Teaching	Mathematical Sciences	K – 8
	High School Mathematics Teaching	Mathematical Sciences	6 – 12
#4	Middle School Mathematics Specialist	Mathematics and Statistics	4 – 8

3.4 Rasch Measurement Framework

The SETS instrument was developed using a Rasch framework (Wolfe and Smith 2007a, 2007b). The Rasch framework for measurement is based on the idea of developing instruments that generate data that conform to a specific measurement model (Wilson 2003). Unlike a traditional statistical approach or the item response theory (IRT) framework to modeling, the emphasis is not on finding the best model that fits or explains that data. Within a Rasch framework, the focus of the analysis is on the identification of items or persons who are considered anomalies (i.e., a person with higher ability incorrectly answering an easy question). This type of analysis facilitates changes in the instrument to improve the measurement of the intended construct, such as removing or rewording an item. The Rasch family of models includes models for unidimensional latent traits (Andrich 1978; Masters 1982; Rasch 1960) and newer models that can explicitly model multidimensionality, such as the Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM) (Adams, Wilson, and Wang 1997; Briggs and Wilson 2003).

As the items on the SETS employ an ordinal response scale and because there is evidence of multidimensionality (Harrell-Williams et al. 2014), the data from this study were scaled to the Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM) (Adams, Wilson, and Wang 1997; Briggs and Wilson 2003). Similar to other Rasch and item response theory models, this model simultaneously estimates person and item parameters on the same logit scale. Equation 1 displays the MRCMLM, which models a person's ability (or other latent trait) on each of the *D* distinct traits (or dimensions) based on their response pattern to the items on the instrument. There is a unique person ability estimate for each of the D dimensions, denoted by $\boldsymbol{\theta} = (\theta_1\ \theta_2, ..., \theta_D)$. The instrument consists of N items (i=1,…, N), where the ith item has k+1 (k =

0, 1, 2, …, k) response levels (categories). The random variable X_{ik} takes on the value 1 if the response to item i equals category k, which occurs with probability π_{Xik}. The column vectors $\mathbf{b}_{ik}$ map the person's response on each item, and are collected into a matrix **B,** which represents the person's scoring matrix for the entire instrument. The design vectors $\mathbf{a}_{ik}$ indicate which of the D dimensions item i is theorized to load on and are combined into the design matrix **A**. The vector ξ contains the item parameters, including the thresholds which indicate the location on the logit scale where the current response category (k) is more likely than the previous adjacent category (k-1).

Equation 1. The Multidimensional Random Coefficients Multinomial Logit Model

$$\pi_{X_{ik}=1;\mathbf{A},\mathbf{B},\xi|\boldsymbol{\theta}} = \frac{\exp\left(\mathbf{b}_{ik}\boldsymbol{\theta} + \mathbf{a}'_{ik}\xi\right)}{\sum_{k=1}^{K_i}\exp\left(\mathbf{b}_{ik}\boldsymbol{\theta} + \mathbf{a}'_{ik}\xi\right)}$$

The resulting item difficulty parameters indicate the difficulty of responding to a particular item with higher-scored categories. Unlike a classical test theory difficulty statistic, which ranges from zero to one, IRT and Rasch difficulty estimates are centered at zero and can be positive or negative in value. Like a z-score, the difficulty estimates can take on any real number but values beyond 3 or –3 are unusual because those items have limited practical value for measurement purposes. Item difficulty parameter estimates that are the largest (positive and larger in absolute value) indicate items that are hardest for the pre-service teachers to respond to with higher response categories, such as "completely confident." In other words, pre-service teachers are less likely to say they are "completely confident" to these items. Item parameter estimates that are lowest (negative and larger in absolute value) indicate items that are easiest to respond to using the higher response categories, meaning that most pre-service teachers are likely to respond with "completely confident." The person parameters rank orders the participants based on their item response patterns. Lower person parameter estimates indicate that a participant displays less of the measured trait, while higher estimates indicate that the participant has more of the measured trait.

Using the rating scale version of the MRCMLM entails two main assumptions. The local independence of items is assumed, meaning that a response on one item should not influence the response on another item (after conditioning on person ability). For instance, local independence is generally violated by having multiple reading items that share a common passage. There is no formal test for this assumption as the choice of the Rasch family model for analysis depends on the design of the items in the instrument. However, all of the SETS items were written as stand-alone items. Secondly, the model assumes that the data were generated to fit the model, which is evaluated using residual-based item fit statistics. The results of the analysis of residuals from the scaling of the data used in this study to the MRCMLM can be found in Harrell-Williams et al. (2014). In summary, none of the item fit statistics indicated problems with the SETS items.

3.5 Analysis

The data from this study were scaled to Multidimensional Random Coefficients Multinomial Logit Model using Conquest (Wu, Adams, Wilson, and Heldane 2007) to obtain person and item parameters. The analysis in this study is exploratory and descriptive in nature as the purpose is to identify which statistical concepts are “easier” and “more difficult” for pre-service teachers to state they are “completely confident” to teach using the item difficulty estimates.

4. Results

4.1 Person Parameter Estimates on the Wright Map

Figure 1 presents the person-item map or Wright map (Wilson 2005) from our data. A Wright map is a commonly used Rasch figure for simultaneously plotting both the item and person estimates on the logit scale. The plot is divided down the center by a dashed line, with the left side displaying a vertical histogram of the person estimates and the right side displaying a vertical histogram of item difficulty. The “#” symbols on the left side of the plot form the distribution of the person estimates of overall self-efficacy to teach statistics, with each symbol representing two or more people. The top of the person distribution is identified with the label “more” and the bottom of the distribution is labeled “less.” This indicates that pre-service teachers with the most self-efficacy to teach statistics are closer to the top of the figure, while those with the least self-efficacy to teach statistics are closer to the bottom of the figure. The “M” marker along the left side of the dashed line indicates the mean of the self-efficacy trait levels. The “S” markers indicate the pre-service teachers within one sample standard deviation of the mean, while the “T” markers indicate those between one and two sample standard deviations from the mean. The shape of the distribution indicates variability among the pre-service teachers in terms of their level of self-efficacy to teach statistics. For the purpose of this analysis, the important result to note is that some amount of variability exists among the participants; for example, not all of the pre-service teachers are grouped at one end of the logit scale.

Figure 1. Wright map of person and item estimates on logit scale

```
                <more>|<rare>
 4             .##    +
                 .    |
                      |
                 .    |
                      |
                      |
                 #    |
 3               .    +
                      |
              #### T  |
                 #    |
               ###    |
                 #    |
              .###    |
 2             .##    +
               .##    |
          .####### S  |
            .#####    |
         .########    |
           .######    |
           #######    |
 1     .##########    +
            ######    |
         .######## M  |T I0023
     #############    |  I0012
           .######    |S I0018  I0020  I0025
        .#########    |  I0017  I0019  I0021  I0026
          ########    |  I0013  I0014
 0         .######    +M I0001  I0006  I0024
            ######    |  I0008  I0009  I0015  I0016
            .##### S  |  I0003  I0005  I0007  I0011
             #####    |S I0002  I0022
             .####    |  I0004
                .#    |T
                 .    |  I0010
-1               #    +
                 .    |
                 # T  |
                      |
                 #    |
                 .    |
                      |
-2                    +
                 #    |
                      |
                      |
                      |
                      |
                 .    |
-3                    +
                <less>|<frequ>
```

4.2 Item Parameter Estimates on the Wright Map

The item parameter locations are plotted on the vertical histogram on the right side of the graph in relationship to their location on the logit scale. The top of the item difficulty distribution is identified with the label “rare” and the bottom of the distribution is labeled “frequent.” This indicates that items closer to the top of the figure (i.e., the harder or more difficult items) are those in which the use of the highest response category “completely confident” is used less frequently, as opposed to those items at the bottom of the figure, where the use of the highest response category is used more frequently. Similar to the person estimates, the “M” marker along the right side of the dashed line indicates the mean of the item difficulties, and the “S” and “T” markers indicate one and two sample standard deviations, respectively, from the item mean. Approximately 2/3 of the item difficulties are within one standard deviation of the mean, as would be expected since the item difficulty estimates follow a bell-shaped distribution. The spacing of the items indicates that SETS instrument measures rather uniformly along the continuum of the intended self-efficacy to teach statistics trait.

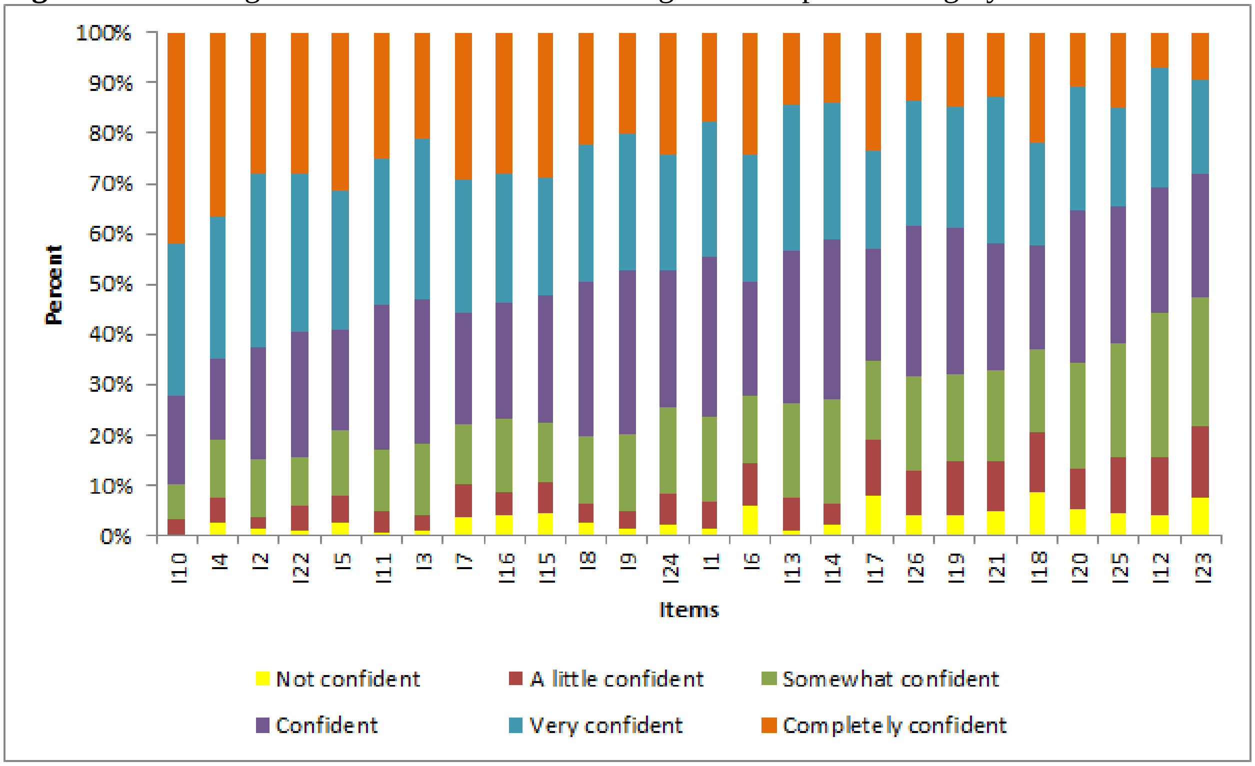
4.3 Results Regarding Items Associated with High and Low Levels of Teaching Efficacy

Looking at the top of the distribution of the vertical item histogram on the right side of the Wright map (Figure 1), the most difficult items for pre-service teachers to endorse as "completely confident" to teach are seen to be (in decreasing order of difficulty) items 23, 12, 25, 20, 18, and 21. The statistical concepts in these items, along with the difficulty estimates and their standard errors, are presented in Table 2. To illustrate this idea, Figure 2 presents the percentages of pre-service teacher responses for each of the six points of the rating scale for each item, sorted in ascending order of difficulty. Item 23 (indicated by the bar to the far right on the figure) has the highest item difficulty estimate of all 26 items on the SETS, and the "not confident" and "a little confident" categories were used with more frequency than on any other item, while the "very confident" and "completely confident" categories were used less often than in most of the other items. These six items identified as being associated with lower level of teaching efficacy are all associated with GAISE Level B, which requires a higher level of knowledge of the concept than items associated with Level A. These items concern association between two variables, the comparison of two groups, and the development of a question to test a hypothesis.

Table 2. Item Difficulty Estimates for the Items Associated with Low Teacher Efficacy

Item Number	Item Difficulty Estimate	S.E. of Estimate	Item Description
23	0.74	0.06	Interpret measures of association.
12	0.61	0.06	Distinguish between a question based on data that vary and a question based on a deterministic model.
25	0.43	0.06	Distinguish between "association" and "cause and effect."
20	0.41	0.06	Describe numerically the strength of association between two variables using linear models.
18	0.36	0.06	Use interquartile range, five-number summaries, and boxplots for comparing distributions.
21	0.31	0.06	Explain the differences between two or more groups with respect to center, variability, and shape.

Figure 2. Percentage of Pre-Service Teachers Using Each Response Category



The items that were most likely to be rated as "completely confident" by pre-service teachers are seen to be 10, 4, 2, 22, and 5. The responses to these items are shown on the left end of the graph, and show the "very confident" and "completely confident" categories were used more than 50% of the time. With the exception of item 22, these items correspond to GAISE Level A. These items concern the variability in results and their generalizability as well as using graphical measures and tables to summarize univariate data. The statistical concepts in these six items, along with the difficulty estimates and their standard errors, are presented in Table 3.

Table 3. Item Difficulty Estimates for the Items Associated with High Teacher Efficacy

Item Number	Difficulty Estimate	S.E. of Estimate	Item Description
5	-0.34	0.06	Use dotplot, stem and leaf plot, and tables (using counts) for *describing* distributions.
22	-0.41	0.06	Recognize that a sample may or may not be representative of a larger population.
2	-0.48	0.06	Recognize that there will be natural variability between observations for individuals.
4	-0.50	0.07	Create dotplot, stem and leaf plot, and tables (using counts) for *summarizing* distributions.
10	-0.90	0.07	Recognize that statistical results may be different in another class or group.

5. Discussion

5.1 Summary

The topics the pre-service middle grades teachers felt most confident about teaching were related mostly to GAISE Level A topics: creating and using tables and graphs to summarize and describe univariate data, and recognizing the existence of variability in the data collected among individuals in the same group as well as from different groups. The topics the pre-service middle grades teachers are least confident to teach include making distributional comparisons across groups, using and interpreting measures of association, and developing a research question, all level B topics. These results are not surprising since the Level B topics build on the skills learned in Level A topics, making them more complex in nature to learn and to teach.

These results are supported by recent research studies, which give plausible explanations for these findings. Most of the current generation of pre-service teachers are a product of a school curriculum that focused mainly on calculations, procedures, and creating graphs (Sorto 2006). There is some evidence that suggests that pre-service teachers, as students themselves, have higher competence about how to read graphs (Gonzáles, Espinel, and Ainley 2011) and compute measures (Jacobbe and Carvalho 2011; Sánchez, da Silva, and Cautnho 2011), than how to interpret or reason about these. Additionally, pre-service teachers have awareness that students experience difficulties when learning to interpret graphs and statistical results (Leavy, Hannigan, and Fitzmaurice 2013).

Research on secondary pre-service and in-service teachers' understanding of distribution, measures of association, and statistical investigations suggests that teachers have had fewer opportunities to learn about these topics during their preparation (Makar and Fielding-Wells 2011) and that even when teachers were exposed to the content and a variety of teaching environments, their understanding is not guaranteed (Makar and Confrey 2004). In the latter study, secondary teachers participated in a six-month professional development program analyzing their students' results on a state mandated high-stakes academic test and solving problems involving sampling distribution. When teachers were given a task of comparing two groups, they were able to describe the variation *within* each distribution and variation *between* groups by separately reporting similarity or differences in means and standard deviations but struggled to interpret the difference between these two types of variation. With respect to teachers' understanding of measures of association, Batanero, Estepa, and Godino (1997) concluded in their study with 19 pre-service teachers that in a computer-based teaching environment most students overcome a deterministic conception of association (association determined by a mathematical function disregarding the random variation). However, most pre-service teachers still believe that a strong association is enough to draw causal conclusions. Also utilizing technology support but from a modeling perspective, Engel, Sedlmeier, and Wörn (2008) exposed 78 pre-service teachers to explore real, bivariate numerical data. They found that the emphasis on the signal-to-noise approach helps students develop statistical thinking skills -- in particular, the handling of variation in scatterplot representations.

In addition to lack of sufficient exposure to the topics, other explanations may exist, such as teacher beliefs regarding the importance and usefulness of certain statistical topics. For instance, Australian secondary teachers indicated that it is important to teach graph interpretation and construction, central measures, variation of data, and probability (Watson 2001). Multiple studies (Estrada 2002; Sedlmeier and Wassner 2008) suggest that mathematics teachers, in particular those with less experience, prefer teaching strategies that relate statistics with daily issues and with other (non-mathematical) topics (Gonzáles et al. 2011).

Lastly, the pre-service teachers' statistical knowledge for teaching was not measured in this study, although it may contribute to their self-efficacy to teach statistics. At the time of data collection, no validated measure of pre-service teachers' statistical knowledge for teaching existed and we are not aware of one that is widely available for use at this time either. While the authors of the Mathematical Knowledge for Teaching measure (Hill, Schilling, and Ball 2004) developed a "Probability, Data and Statistics" module, the majority of the items focus on probability and the items on data and statistics have not been validated as a separate instrument for research or assessment purposes.

5.2 Limitations

This study was completed using data from a convenience sample, which may limit the generalizability of these results. While the predominance of Caucasian (88%) and female (78%) pre-service teachers makes the sample quite similar in these demographics to pre-service teachers in other studies on teacher efficacy (Duffin, French, and Patrick 2012; Fives and Buehl 2010; Knoblauch and Woolfolk Hoy 2008; Tschannen-Moran and Woolfolk Hoy 2001), this sample may not be representative of the entire population of pre-service middle grades mathematics teachers.

5.3 Implications

The identification of statistical topics associated with high and low levels of teachers' self-confidence have several implications for teacher preparation and development. Teacher preparation programs that are in the process of revising their curriculum to include the teaching and learning of statistics at the middle school level should be informed by the evidence of this and other studies that teachers need more opportunities to develop research questions, to use and interpret measures of association, and to compare groups using distributional concepts. In addition to opportunities to develop these topics, teacher educators and developers need to consider teaching and learning issues associated with them and the difference between teachers as learners and as students. For example, when teaching the concept of group comparisons, research results suggest that teachers struggled to interpret the difference between *within* and *between* variation and hence need to discuss frequently sources of variation in both data and measures to "engender a tolerance for variation both within and between distributions" (Makar and Confrey 2004, p. 371). When teaching measures of association, pre-service teachers tend to have a deterministic, unidirectional, and casual conception of association (Estepa and Batanero 1996). Although these results can be generalized to students in general, teachers see themselves as experts and may be less likely to be open to learning and discussing their reasoning. The administration of SETS as a diagnostic or evaluation tool is less intimidating than a content test, hence the instrument can serve as a planning tool to uncover self-perceived teachers' weaknesses and strengths while maintaining a collegial, professional atmosphere. However, research has shown that pre-service teacher beliefs and teaching efficacy are not always highly correlated with specialized content knowledge for teaching (Leader-Janssen and Rankin-Erickson 2013; Swars, Hart, Smith, Smith, and Tolar 2007). Although the SETS instruments have been used solely with pre-service teachers thus far, it is reasonable to imagine professional development designers and facilitators for in-service teachers using the SETS instrument to identify topics of focus during in-service training days in a school or district. If a professional development program is already designed, the SETS instrument might serve to evaluate the impact of the program on specific topics.

5.4 Future Work

Data collection for validating the high school SETS instrument is underway. As the 26 items from the middle grades SETS instrument are contained in the high school instrument, this additional data can be used to confirm the current findings as well as rank order the statistical concepts that relate to the high school statistics strands of the CCSSM. Additionally, open-ended questions are being added to the middle and high school versions of the SETS to collect data regarding why the pre-service teachers are rating specific topics with low or high efficacy. As content assessments that align with CCSSM become available, investigations regarding the relationship between content knowledge for teaching statistics in K-12 and self-efficacy to teach statistics will be conducted.

Appendix - All Items on the Middle Grades SETS Instrument

The instrument below uses the two-factor structure presented in Harrell-Williams et al. (2014). Applying the language of Friel, Curcio, and Bright (2001), the two-factor structure in the column headings represents levels A and B, respectively, of the preK-12 GAISE (Franklin et al. 2007). Readers interested in using the SETS instrument should contact the authors.

Factor 1 Items: "Reading the Data"	**Factor 2 Items: "Reading Between the Data"**
1. Collect data to answer a posed statistical question in contexts of interest to middle school students.	12. Distinguish between a question based on data that vary and a question based on a deterministic model (for example, specific values of rate and time determines a particular value for distance in the model $d = r \times t$).
2. Recognize that there will be natural variability between observations for individuals.	13. Identify what variables to measure and how to measure them in order to address the question posed.
3. Select appropriate graphical displays and numerical summaries to compare individuals to each other and an individual to a group.	14. Describe numerically the variability between individuals within the same group.
4. Create dotplot, stem and leaf plot, and tables (using counts) for *summarizing* distributions.	15. Create histograms for summarizing distributions.
5. Use dotplot, stem and leaf plot, and tables (using counts) for *describing* distributions.	16. Use histograms for comparing distributions.
6. Create boxplots for *summarizing* distributions.	17. Compute interquartile range and five-number summaries for summarizing distributions.
7. Use boxplots, median, and range for *describing* distributions.	18. Use interquartile range, five-number summaries, and boxplots for comparing distributions.
8. Identify the association between two variables from scatterplots.	19. Recognize the role of sampling error when making conclusions based on a random sample taken from a population.
9. Generalize a statistical result from a small group to a larger group such as the whole class.	20. Describe numerically the strength of association between two variables using linear models.
10. Recognize that statistical results may be different in another class or group.	21. Explain the differences between two or more groups with respect to center, spread (for example, variability), and shape.
11. Recognize the limitation of making inference (i.e. generalization) from a classroom dataset to any population beyond the classroom.	22. Recognize that a sample may or may not be representative of a larger population.
	23. Interpret measures of association.
	24. Distinguish between an observational study and a designed experiment.
	25. Distinguish between "association" and "cause and effect."
	26. Recognize sampling variability in summary statistics such as the sample mean and the sample proportion.

Acknowledgement

The researchers gratefully acknowledge support from the Consortium for the Advancement of Undergraduate Statistics Education (under NSF DUE award #0618790).

References

Adams, R. J., Wilson, M., and Wang, W. C. (1997), "The Multidimensional Random Coefficients Multinomial Logit Model," *Applied Psychological Measurement*, 21(1), 1-23.

Andrich, D. A. (1978), "A Rating Formulation for Ordered Response Categories," *Psychometrika*, 43(4), 561-573.

Batanero, C., Estepa, A., and Godino, J. D. (1997), "Evolution of Students' Understanding of Statistical Association in a Computer-Based Teaching Environment," in J.B. Garfield, and G. Burrill (Eds.) (1996), *Research on the Role of Technology in Teaching and Learning Statistics:* 1996 IASE Round Table Conference (pp. 183-198), University of Minnesota: IASE.

Begg, A. and Edwards, R. (1999), "Teachers' Ideas About Teaching Statistics," in *Proceedings of the 1999 Combined Conference of the Australian Association for Research in Education and the New Zealand Association for Research in Education,* Melbourne: AARE and NZARE. Available at http://publications.aare.edu.au/99pap/beg99082.htm

Briggs, D. C. and Wilson, M. (2003), "An Introduction to Multidimensional Measurement Using Rasch Models," *Journal of Applied Measurement*, 4(1), 87-100.

Çakiroğlu, E. (2000), *Preservice Elementary Teachers' Sense of Efficacy in Reform Oriented Mathematics* (Unpublished doctoral dissertation), Indiana University, Bloomington, IN.

Conference Board of the Mathematical Sciences (2010), *The Mathematical Education of Teachers II* (Issues in Mathematics Education, 17), Providence, RI: American Mathematical Society. Available at: http://www.cbmsweb.org/MET2/met2.pdf

Duffin, L. C., French, B. F. and Patrick, H. (2012), "The Teachers' Sense of Efficacy Scale: Confirming the Factor Structure with Beginning Pre-service Teachers," *Teaching and Teacher Education*¸ 28(6), 827-834.

Engel, J., Sedlmeier, P., and Woern, C. (2008), "Modeling Scatterplot Data and the Signal-Noise Metaphor," in C. Batanero, G. Burrill, C. Reading, and A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*, Monterrey, Mexico: International Commission on Mathematical Instruction and International Association for Statistical Education. Available at: http://iase-web.org/documents/papers/rt2008/T4P8_Engel.pdf.

Estepa, A. and Batanero, C. (1996), "Judgments of Correlation in Scatter Plots: Students' Intuitive Strategies and Preconceptions," *Hiroshima Journal of Mathematics Education*, 4, 21-41.

Estrada, A. (2002), *Análisis de las Actitudes y Conocimientos Estadísticos Elementales en la Formación del Profesorado (Analyzing Attitudes and Elementary Statistical Knowledge in the Training of Teachers)*, unpublished doctoral dissertation, Universidad Autónoma de Barcelona, Spain.

Estrada, A., Batanero, C. and Lancaster, S. (2011), "Teachers' Attitudes Toward Statistics," in C. Batanero, G. Burrill and C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study* (pp. 163-174), New York: Springer.

Fives, H. and Buehl, M. M. (2010), "Examining the Factor Structure of the Teachers' Sense of Efficacy Scale," *The Journal of Experimental Education*, 78(1), 118-134.

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

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

Gonzáles, M. T., Espinel, M. C. and Ainley, J. (2011), "Teachers' Graphical Competence," in C. Batanero, G. Burrill and C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study* (pp. 187-197), New York: Springer.

Gresham, G. (2008), "Mathematics Anxiety and Mathematics Teacher Efficacy in Elementary Pre-service Teachers," *Teaching Education*, 19(3), 171-184.

Harrell, L. M., Pierce, R. L., Sorto, M. A., Murphy, T. J., Lesser, L. M., and Enders, F. B. (2009), "On the Importance and Measurement of Pre-Service Teachers' Efficacy to Teach Statistics: Results and Lessons Learned from the Development and Testing of a GAISE-Based Instrument," in *Proceedings of the Joint Statistical Meetings, Section on Statistical Education*, (pp. 3396-3403), Alexandria, VA: American Statistical Association. Available at: http://www.statlit.org/pdf/2009HarrellEtAlASA.pdf

Harrell-Williams, L. M., Sorto, M. A., Pierce, R. L., Lesser, L. M., and Murphy, T. J. (2014), "Validation of Scores from a Measure of Teachers' Self-efficacy to Teach Middle Grades Statistics," *Journal of Psychoeducational Assessment*, 32(1), 40-50.

Hill, H. C, Schilling, S. G., and Ball, D. L. (2004), "Developing Measures of Teachers' Mathematics Knowledge for Teaching," *The Elementary School Journal*, 105(1), 11-30.

Huinker, D. and Madison, S. (1997), "Preparing Efficacious Elementary Teachers in Science and Mathematics: The Influence of Methods Courses," *Journal of Science Teacher Education*, 8(2), 107-126.

Jacobbe, T. and Carvalho, C. (2011), "Teachers' Understanding of Averages," in C. Batanero, G. Burrill and C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study* (pp. 199 – 209), New York: Springer.

Knoblauch, N. A. and Woolfolk Hoy, A. (2008), "'Maybe I Can Teach Those Kids.' The Influence of Contextual Factors on Student Teachers' Efficacy Beliefs," *Teaching and Teacher Education*, 24(1), 166-179.

Leader-Janssen, E. M. and Rankin-Erickson, J. L. (2013), "Preservice Teachers' Content Knowledge and Self-Efficacy for Teaching Reading", *Literacy Research and Instruction,* 52(3), 204-229.

Leavy, A. M., Hannigan, A. and Fitzmaurice, O. (2013), "If You're Doubting Yourself Then, What's the Fun in That? An Exploration of Why Prospective Secondary Mathematics Teachers Perceive Statistics as Difficult," *Journal of Statistics Education* [online], 21(3), 1-26. Available at http://www.amstat.org/publications/jse/v21n3/leavy.pdf

Lesser, L. M. (2013), "One 'Mean' Class," *Statistics Teacher Network* [online], 81, 20. Available at http://www.amstat.org/education/stn/pdfs/STN81.pdf.

Makar, K. and Confrey, J. (2004), Secondary Teachers' Reasoning about Comparing Two Groups," in D. Ben-Zvi and J. Garfield (Eds.), *"The Challenges of Developing Statistical Literacy, Reasoning, and Thinking"* (pp. 353-373), Dordrecht: Kluwer.

Makar, K. and Fielding-Wells, J. (2011), "Teaching Teachers to Teach Statistical Investigations," in C. Batanero, G. Burrill and C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study* (pp. 347-358), New York: Springer.

Masters, G. N. (1982), "A Rasch Model for Partial Credit Scoring," *Psychometrika*, 47(2), 149-174.

Pfannkuch, M. and Ben-Zvi, D. (2011), "Developing Teachers' Statistical Thinking," in C. Batanero, G. Burrill and C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study* (pp. 323-333), New York: Springer.

Pierce, R. and Chick, H. (2011). "Teachers' Beliefs about Statistics Education," in C. Batanero, G. Burrill and C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study* (pp. 151-162), New York: Springer.

Rasch, G. (1960), *Probabilistic Models for Some Intelligence and Attainment Tests*, Copenhagen: Danish Institute for Educational Research.

Sánchez, E., Borim da Silva, C., and Cautnho, C. (2011), "Teachers' Understanding of Variation," in C. Batanero, G. Burrill and C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study* (pp. 211-221), New York: Springer.

Sedlmeier, P. and Wassner, C. (2008), "German Mathematics Teachers' Views on Statistics Education," in C. Batanero, G. Burrill, C. Reading, and A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*, Monterrey, Mexico: International Commission on Mathematical Instruction and International Association for Statistical Education. Available at: http://iase-web.org/documents/papers/rt2008/T2P1_Sedlmeier.pdf

Sorto, M. A. (2006), "Identifying Content Knowledge for Teaching Statistics," in A. Rossman and B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil: International Statistical Institute and International Association for Statistical Education. Available at: http://iase-web.org/documents/papers/icots7/C130.pdf

Swars, S. L. (2005), "Examining Perceptions of Mathematics Teaching Effectiveness Among Elementary Preservice Teachers with Differing Levels of Mathematics Teacher Efficacy," *Journal of Instructional Psychology*, 32(2), 139-147.

Swars, S. L., Hart, L. C., Smith, S. Z., Smith, M. E. and Tolar, T. (2007), "A Longitudinal Study of Elementary Service Pre-Service Teachers' Mathematics Beliefs and Content Knowledge," *School Science and Mathematics*, 107(8), 325-335.

Tschannen-Moran, M. and Woolfolk Hoy, A. (2001), "Teacher Efficacy: Capturing an Elusive Construct," *Teaching and Teacher Education*, 17(7), 783-805.

Watson, J. M. (2001), "Profiling Teachers' Competence and Confidence to Teach Particular Mathematics Topics: The Case of Chance and Data," *Journal of Mathematics Teacher Education*, 4(4), 305-337.

Wenta, R. G. (2000), *Efficacy of Preservice Elementary Mathematics Teachers*, Unpublished doctoral dissertation, Indiana University.

Wilson, M. (2003), "On Choosing a Model for Measuring", *Methods of Psychological Research Online*, 8(3), 1-22.
Available at http://www.dgps.de/fachgruppen/methoden/mpr-online/issue21/mpr122_8.pdf

Wilson, M. (2005), *Constructing Measures: An Item Response Modeling Approach*, Mahwah, NJ: Erlbaum.

Wolfe, E. W. and Smith, E. V., Jr. (2007a), "Instrument Development Tools and Activities for Measure Validation Using Rasch Models: Part I—Instrument Development Tools," in E. V. Smith, Jr. and R. M. Smith (Eds.), *Rasch Measurement: Advanced and Specialized Application* (pp. 202-242), Maple Grove, MN: JAM Press.

Wolfe, E. W. and Smith, E. V., Jr. (2007b), "Instrument Development Tools and Activities for Measure Validation Using Rasch Models: Part II—Validation Activities," in E. V. Smith, Jr. and R. M. Smith (Eds.), *Rasch Measurement: Advanced and Specialized Application* (pp. 243-290), Maple Grove, MN: JAM Press.

Wright, B. D. and Masters, G. N. (1982), *Rating Scale Analysis*, Chicago: MESA Press.

Wu, M. L., Adams, R. J., Wilson, M. R., and Heldane, S. A. (2007), *ACER ConQuest: Generalized item response modeling software (Version 2.0)*, Melbourne: Australian Council for Educational Research.

Leigh M. Harrell-Williams
Department of Counseling, Educational Psychology & Research
University of Memphis
3720 Alumni Avenue
Memphis, TN 38152
Leigh.Williams@memphis.edu

M. Alejandra Sorto
Department of Mathematics
Texas State University
601 University Drive
San Marcos, TX 78666
sorto@txstate.edu

Rebecca L. Pierce
Department of Mathematical Sciences
Ball State University
2000 W University Avenue
Muncie, IN 47306
rpierce@bsu.edu

Lawrence M. Lesser
Department of Mathematical Sciences
The University of Texas at El Paso
500 W. University Avenue

El Paso, TX 79968
Lesser@utep.edu

Teri J. Murphy
Department of Mathematics & Statistics
Northern Kentucky University
Nunn Drive
Highland Heights, KY 41011
murphytj1@nku.edu

Volume 23 (2015) | Archive | Index | Data Archive | Resources | Editorial Board | Guidelines for Authors | Guidelines for Data Contributors | Guidelines for Readers/Data Users | Home Page | Contact JSE | ASA Publications