2016 NCTM RESEARCH CONFERENCE
April 11-13 • San Francisco

Paper Title: Preparedness of Preservice Secondary Mathematics Teachers' to Teach Statistics

Author(s): Jennifer N. Lovett

Session Title: Preparedness of Preservice Secondary Mathematics Teachers' to Teach Statistics

Session Type: Interactive Paper

Presentation Date: April 13, 2016

Presentation Location: San Francisco, California

Authors/presenters retain copyright of the full-text paper. Permission to use content from this must be sought from the copyright holder.

Preparedness of Preservice Secondary Mathematics Teachers' to Teach Statistics

Introduction

Statistics and data science are two of the fastest growing and most popular fields (Hardin et al., 2015). The pipeline to prepare the workforce for these disciplines begins in K-12, particularly in high school. Over the last decade there has been increased emphasis on statistics in standards documents meant to guide K-12 curriculum (e.g., National Governors Association Center for Best Practice & Council of Chief State School Officers, 2010). However, research has shown that inservice teachers are not prepared to teach statistics (e.g., Burrill & Biehler, 2011) and struggle understanding how the statistical content should progress across grade levels (Jones & Tarr, 2010). Teachers often teach statistics procedurally, focusing on computations of statistical measures (Makar & Confrey, 2004) and creating graphical representations (Sorto, 2006). Therefore, students who enroll in secondary mathematics teacher education programs in the U.S. have likely had minimal experience with statistics in their own K-12 education. They have not had many opportunities to develop a conceptual understanding of the topics they are now expected to teach compared to other areas of mathematics; thus, preservice teachers are likely no more prepared than inservice teachers to teach statistics (Franklin et al., 2007).

While examining knowledge needed to teach is important, researchers should also consider the non-cognitive aspects that teachers draw upon and how these are related to a teacher's preparedness to teach statistics (e.g., Ball, Thames, & Phelps, 2008). Teachers' affect plays a crucial role in the pedagogical approaches they use, the time spent on a subject, and thus can impact students' learning (e.g., Wilkins, 2008). Affect includes a teacher's beliefs, attitudes, and emotions towards statistics. However, there is a lack of research on secondary teachers' affect in regards to teaching statistics (Batanero, Burrill, & Reading, 2011). The limited research that has been conducted has been with elementary teachers. Researchers have found that a teacher's beliefs and attitudes towards statistics were related to their prior experiences with statistics, impact the choice of instructional tasks, and students' attitudes and beliefs towards statistics (e.g., Begg & Edwards, 1999). Since a teacher's beliefs and attitudes plays a large role, it is crucial when considering PSMTs' preparedness to teach statistics that PSMTs' affect as well as statistical knowledge is examined.

Background and Research Questions

Statistical Knowledge

In 2007, the GAISE report laid out the statistical concepts that students need to develop in K-12 schooling and thus setting a minimum for the statistical knowledge PSMTs need to know to teach statistics. The GAISE framework consists of three levels A, B, and C (Franklin et al., 2007). Although there are not explicit definitions given for each level in the GAISE framework, the levels increase in statistical sophistication and become more abstract. Each level is aligned to specific statistical content. The content in level A represents topics for early or novice learners of statistics (no matter what grade level, but often introduced in elementary and middle school), level B represents slightly more advanced statistical content (often taught in middle school or early high school), and level C represents even more advanced content (typical taught in high school or introductory college courses) (Franklin et al., 2007).

The GAISE report recommends that within each of the three levels, students should learn statistical topics through engaging with the statistical investigative cycle (Wild & Pfannkuch, 1999). Though the investigative cycle is described slightly differently in different countries, in the US, and in the GAISE framework, four components are emphasized: posing questions, collecting data, analyzing data, and interpreting results. To effectively prepare high school students to increase their sophistication across three levels with all phases of a statistical investigation, PSMTs also need deep statistical knowledge across all three levels and all phases.

PSMTs' statistical knowledge

Over the last 25 years, there has been little research about the statistical knowledge of PSMTs or the misconceptions they develop (Batanero et al., 2011), even recently with the increased emphasis on statistics in high school mathematics with the adoption of CCSSM. The majority of research on preservice teachers' statistical knowledge has focused on elementary teachers (e.g., Browning, Goss, & Smith, 2014; Hu, 2015). The limited research conducted on PSMTs' statistical knowledge has been small-scale studies, from a small number of institutions on specific statistical content (e.g., Doerr & Jacob, 2011; Makar & Confrey, 2005). For example, a recent study by Casey and Wasserman (2015) examined 11 preservice teachers' statistical knowledge of informal lines of best fit from three universities. From these studies, research has shown that preservice secondary teachers focus on procedures, computations, and algorithms, lack statistical reasoning skills, and have difficulty interpreting graphical representations. To date, only one large-scale study, conducted by Lee et al. (2014) examined how 204 preservice mathematics teachers from eight universities used dynamic statistical tools to conduct a statistical investigation. They found that preservice teachers who pose a broad statistical question engaged in more graphical augmentations (e.g., adding shaded regions, reference lines, or statistical measures) using dynamic statistical software. These graphical augments allowed preservice teachers to dive deeper into the data analysis and make connections to the context to support claims. For the field to truly understand PSMTs' statistical knowledge and pedagogical statistical knowledge, more small and large-scale studies are needed.

Statistics Teaching Efficacy

As mentioned before, one's preparedness to teach not only relies on cognitive aspects, but also affective constructs such as beliefs, attitudes and self-efficacy. In our work, we focus on preservice secondary mathematics teachers' self-efficacy to do the job they are preparing for, teaching statistics to high school students. Self-efficacy has grown from Bandura's (1977) social cognitive theory. An individual's self-efficacy originates from the construct of efficacy expectations. Bandura (1986) defines self-efficacy as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performance (p.391). Judgments of one's own self-efficacy are task-specific and change over time (Bandura, 1977). A teacher has two types of self-efficacy for each content area they teach: self-efficacy to know the content themselves and self-efficacy to teach the topic to students. In mathematics education, researchers have defined these two types of self-efficacy as: mathematics self-efficacy and mathematics teaching efficacy (Bates, Latham, & Kim, 2011). Mathematics self-efficacy can be defined as a teacher's belief in his/her ability to do mathematics (Hackett & Betz, 1989) and mathematics teaching efficacy can be defined as a teacher's belief in his/her ability to teach mathematics to bring about student learning (Ashton, 1985). Applying this to teaching statistics, teachers have two types of self-efficacy: statistics self-efficacy and statistics teaching efficacy. Statistics self-efficacy is a "teacher's belief in his/her ability to do statistics" and statistics teaching efficacy as "a teacher's belief in his/her ability to teach statistics to bring about student learning" (Lovett, Doerr, Thrasher, & Lee, under review, p. 4).

Only one instrument (known to the researchers) has been developed to measure statistics *teaching* efficacy: the Self-Efficacy for Teaching Statistics (SETS) instrument. This instrument has two versions: one to measure the statistics teaching efficacy to teach middle school students (Harrell-Williams, Sorto, Pierce, Lesser, & Murphy, 2013) and the other to measure the statistics teaching efficacy to teach high school students (Harrell-Williams, Sorto, Pierce, Lesser, & Murphy, 2014). Due to the recentness of this instrument, only a few studies have been conducted on statistics teaching efficacy on preservice teachers. One current study being conducted by the authors of SETS is to determine the relationship of preservice teachers' statistics teaching efficacy using the middle school SETS instrument and statistics self-efficacy using the CSSE instrument. Pearson correlation coefficient was computed to examine the relationship of the total SETS score and CSSE to score to be 0.819 (Harrell-Williams, personal communication). This provides evidence that an individual's self-efficacy to do statistics plays a crucial role in their statistics teaching efficacy.

Purpose of the Study

Given the context of statistics education in the US and increased demands on secondary teachers for teaching statistics, this study examines the preparedness of PSMTs to teach statistics as they enter student teaching. Therefore, the research questions that will be addressed are:

1. What is PSMTs' statistical knowledge of the high school content they are expected to teach using the phases of a statistical investigation?
2. What is PSMTs' statistics teaching efficacy for teaching high school statistics?

Methodology

This paper is part of a mixed methods study on preparedness of PSMTs to teach statistics Author (2016). The study utilizes an explanatory design, first quantitatively examining PSMTs' statistical knowledge and statistics teaching efficacy, and then qualitatively seeking factors and experiences that influence PSMTs' confidence through analysis of open-ended responses and interviews.

Participating Institutions

The population of interest for this study is PSMTs prepared through university-based teacher preparation programs in the United States. Since a list of all universities in the US that prepare PSMTs is not readily accessible, a purposeful sampling was used rather than using a random sample. This study focused on PSMTs who currently attend institutions where at least one faculty member participated in either a National Science Foundation (NSF)-funded or American Statistical Association (ASA)-funded program to increase the emphasis of statistics education of teachers at that institution between 2002-2014. This narrowed the possible sample to 57 institutions whose faculty had participated in the NSF-funded program, Preparing to Teach Mathematics with Technology (PTMT, ptmt.fi.ncsu.edu), and/or the ASA-funded Math/Stat Teacher Education: Assessment, Methods, and Strategies (TEAMS, http://www.amstat.org/sections/educ/newsletter/v9n1/TEAMS.html). All 57 institutions were contacted, and 18 institutions agreed to participate in this study during the 2014-2015 school year. The majority of institutions (61.1%) had an enrollment profile of high undergraduate and the majority of participants attended institutions with a basic classification of research universities (very high), research universities (high), or Master's college and university with a larger program.

Participants

Across 18 institutions, there were 235 PSMTs who participated in some aspect of the study. Participants who didn't complete all aspects of the study and those who took exceptionally less time to complete the content assessment than recommended by the authors of the assessment (less than ten minutes) were eliminated (Jacobbe, personal communication). This resulted in a sample size of 217 PSMTs consisting of undergraduate juniors and seniors, or graduate students earning initial licensure enrolled in their last mathematics education course prior to student teaching. The majority of PSMTs were female (71%) and 88% were Caucasian. The majority of PSMTs (59%) had taken one or two statistics courses at the time of the study.

Instruments

Two instruments were used in this study to measure PSMTs' statistics teaching efficacy and their content understanding of statistics. The instruments were administered online in the final few weeks of the participants' last mathematics methods course before student teaching. Participants took the statistics teaching efficacy instrument first, and then shortly afterwards took the content assessment. Details about each instrument follow below.

Statistics teaching efficacy. To examine PSMTs' statistics teaching efficacy, the high school version of the Self Efficacy to Teach Statistics (SETS; Harrell-Williams et al., 2014) instrument was administered. This instrument was chosen because it collects both qualitative and quantitative data about PSMTs' statistics teaching efficacy. Furthermore, the SETS instrument is aligned with the GAISE framework, which reflects the content that PSMTs are expected to teach to high school students. Hence, there is a close correspondence between the teaching efficacy instrument and the specific topics PSMTs need to know and to teach.

The instrument contains 44 six-point Likert scale items. An earlier version of this instrument with 26 items aligned with levels A and B of GAISE was validated for use in measuring changes in elementary and middle grades preservice teachers' self-efficacy as a result of interventions, such as a course (Harrell-Williams et al., 2013). The high school version contains the previous 26 items aligned to GAISE levels A and B and contains an additional 18 items validated and aligned to GAISE level C (Harrell-Williams, personal communication). In addition to an overall score, the instrument provides sub-scale scores that correspond to Levels

A, B and C in the GAISE framework. There are 11 Likert items for level A, 15 items for level B and 18 items for level C. For all Likert items, the stem of the question was

> "Rate your confidence in teaching high school students the skills necessary to complete successfully the task given by selecting your choice on the following scale: 1 = not at all confident, 2 = only a little confident, 3 = somewhat confident, 4 = confident, 5 = very confident, 6 = completely confident" (Harrell-Williams et al., 2014).

Therefore, for each participant, the SETS instrument produces a confidence score to teach high school students statistics for GAISE level A, GAISE level B, GAISE level C, and an overall score. There may be concern that PSMTs would have a tendency to present favorable images of their statistics teaching efficacy since they are self-reporting (Ross, 1989); however research has shown that there is little motivation to misreport since the confidentiality was preserved (Baldwin, 2000).

Statistical Knowledge. To examine PSMTs' statistical knowledge, the Levels of Conceptual Understanding of Statistics (LOCUS) assessment (Jacobbe, Case, Whitaker, & Foti, 2014) was administered online (locus.statisticseducation.org). The LOCUS assessment is aligned with the CCSSM and assesses understanding across the three levels of development in the GAISE framework (Franklin et al., 2007). A 23-item paper version of the LOCUS assessment has been validated as a measure to reliably assess current statistical understanding (Jacobbe, Case, et al., 2014). Participants in this study took the Intermediate/Advanced Statistical Literacy version of the assessment, which was designed for students in grades 10 – 12. This test was chosen because this assessment represents the content that PSMTs are expected to teach to their students in the near future when they begin teaching. This assessment consists of 30 questions: 20 questions that align with levels B and C of the GAISE framework and 10 equator questions aligning with all three levels. These 30 questions contain the previously validated 23 items and an additional 7 items.[1] These 30 questions are also aligned with the four phases of an investigative cycle: Forming Questions, Collecting Data, Analyzing Data, and Interpreting Results. For a description of each type of question and sample questions see Jacobbe, Foti, Case, and Whitaker (2014). Therefore, for each participant, the LOCUS assessment produces a score for GAISE level B, GAISE level C, Forming Statistical Questions, Collecting Data, Analyzing Data, Interpreting Results, and an overall score.

Analysis of Data

Analysis of the LOCUS scores began by generating descriptive statistics and distributions for PSMTs' LOCUS overall scores, Level B scores, Level C scores and scores for each portion of the statistical investigative cycle. Paired samples t-tests were used to test for differences between PSMTs' statistical knowledge in GAISE Levels B and C, and a repeated measures ANOVA used to test for significant differences in PSMTs' statistical knowledge between the four phases of a statistical investigation. Paired samples t-tests were appropriate since the samples are independent and identically distributed. A repeated measures ANOVA was used since the assumptions of normality and sphericity were not validated.

When analysis began of PSMTs' statistics teaching efficacy, it was necessary to account for missing data, since every PSMT did not complete every SETS item. Since data was primarily missing due to nonresponse of certain items this was addressed by using multiple imputation (Allison, 2002). Nine imputed data sets were created using SPSS, since approximately nine cases were missing (White, Royston, & Wood, 2011). Following imputation, an overall score was calculated for each PSMT by calculating the sum of his/her Likert scores for all of the items and

---

[1] This 30-item version of LOCUS has been validated, with manuscript currently under review (Jacobbe, personal communication).

then dividing by the total number of items. Sub-scale scores for each GAISE level were also calculated for each PSMT by calculating the sum of the Likert scores and dividing by the number of items in each level. This resulted in four scores for each PSMT that corresponded to the six-point Likert scale for each of the nine imputed data sets. Using Rubin's (1987) rules, analyses run on each imputed data set were pooled and were examined to compare pooled values to the original data. The results of the pooled data were similar to the original data thus imputed results will be presented. To test for significant differences in PSMTs' statistics teaching efficacy between the three GAISE levels, repeated measures ANOVA tests were used. A repeated measures ANOVA was used since the assumptions of normality and sphericity were not validated.

Results

Statistical Knowledge

These results describe what PSMTs from 18 universities currently understand about the statistics content they will soon be responsible for teaching. Table 1 reports summary statistics for PSMTs' scores on the LOCUS assessment overall, for GAISE Levels B and C. With an overall mean score of 69% and a standard deviation of 14.06, these results suggest that these PSMTs do not demonstrate a strong conceptual understanding of the statistics content they will soon teach high school students. Figure 1 shows the distribution of PSMTs' LOCUS scores. As seen in the boxplots for the overall scores, there are at least some PSMTs who scored in the 90-100%, indicating a strong statistical knowledge of topics they will soon be responsible to teach. However, only one fourth of PSMTs scored above 77%, and one fourth scored below 57% overall.

Table 1. *PSMTs' statistics content scores*

	Number of items	Mean	SD
Overall Score	30	68.61	14.06
GAISE Levels			
Level B Score	11	70.85	17.69
Level C Score	17	64.87	14.16
Phases of Statistical Investigative Cycle			
Formulating Questions	5	80.37	21.51
Collect Data	7	70.40	19.70
Analyze Data	7	63.34	22.22
Interpret Results	11	60.48	16.25

Note. Two items are classified as GAISE level A, but were not analyzed due to the small number.

In terms of the GAISE level subscores, PSMTs scored, on average, significantly higher on the Level B items than on Level C items ($t=5.772$, $p<0.001$), demonstrating that their statistical knowledge is weaker as items increase in sophistication. The boxplots show that for the subscores, there are at least some PSMTs who demonstrated a strong grasp of the statistical concepts they will soon be responsible to teach for both GAISE levels; however, there is a concern for the majority of PSMTs since more than 75 percent of them scored below 80% overall. The interquartile range for the overall score and Levels B and C are similar, though more outliers exist for Level B (the star in *Figure* refers to an extreme outlier, which is a value more than three times the interquartile range less than either Q1 or Q3.)

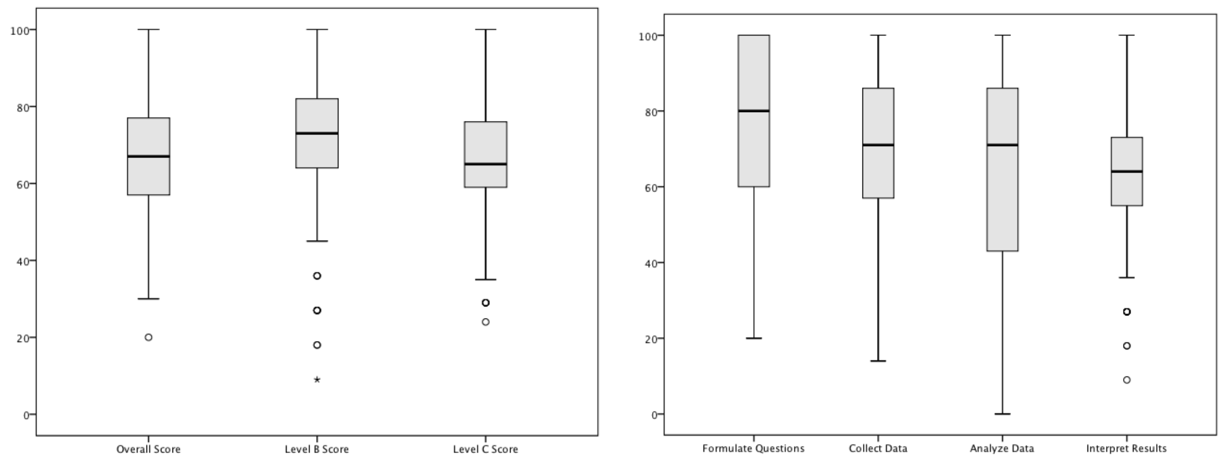

Figure 1. Distribution of PSMTs' LOCUS scores

Examining scores by the phases in the statistical investigative cycle, Table and **Error! eference source not found.** shows that PSMTs scored highest on Formulating Questions and the lowest on Interpreting Results items. A repeated measures ANOVA determined that mean scores differed significantly between the four phases of the cycle [F(3,648)=64.73, p<0.001]. Post hoc tests using a Bonferroni correction revealed that PSMTs scored significantly lower on questions as the cycle progressed (p<0.001). However, there was no real difference between mean scores for Analyze Data and Interpret Results (p=0.32). The boxplots in figure 1 show that for all four phases, there are some PSMTs who scored very well, indicating that those PSMTs likely have the content knowledge needed for teaching that phase of an investigative cycle. On Formulating Questions items, at least half of PSMTs scored 80% or higher, and a quarter of those scored 100%, indicating stronger understanding for these PSMTs about Formulating Questions. However, there are concerns regarding the other phases of the investigative cycle. Half of PSMTs scored below 71% on Collecting Data and Analyzing Data items, and half scored below 64% on Interpreting Results items. Even being conservative, this result is convincing that the majority of PSMTs do not have the statistical foundation needed to teach students key concepts related to Collecting Data, Analyzing Data, and Interpreting Results.

An examination of individual items identified several strengths and weakness of PSMTs' statistical knowledge. As previously mentioned, PSMTs scored the highest on average for Formulating Questions items, in these items PSMTs' demonstrated a strength in their ability to read a description of a study and measurements taken to identify the statistical question of interest. For Collecting Data items, one strength of PSMTs' was their ability to identify a data collection plan based on a study description. In terms of Analyzing Data and Interpreting Results, PSMTs are proficient at identifying an appropriate measure of center for a given context and comparing distributions in a context using the center and spread, topics that are heavily emphasized in school mathematics. This strength in understanding measures of center suggests that PSMTs' should be well equipped to assist their future students develop stronger conceptions of measures of center beyond the standard algorithms.

Weaknesses demonstrated by the largest number of PSMTs were seen in the Analyze Data and Interpret Results items. This is not surprising since PSMTs' scored on average the lowest in these two phases. These weaknesses involve issues in understanding variability, sampling distributions, *p*-values, and confidence intervals. Even though emphasis on these topics in high school has increased with the adoption of CCSSM, PSMTs' statistics and mathematics education courses have not developed a deep understanding of these topics.

Statistics Teaching Efficacy

PSMTs completed the SETS instrument by rating their confidence to teach statistics from 1 to 6 so that 1 = not at all confident, 2 = only a little confident, 3 = somewhat confident, 4 = confident, 5 = very confident, 6 = completely confident" (Harrell-Williams et al., 2014). Table 2 reports summary statistics for PSMTs' confidence scores on the SETS instrument overall and for GAISE Levels A, B and C. These results show that on average PSMTs are *confident* to teach high school students statistics and are most confident in GAISE Level A items. On average, teachers only felt between *somewhat confident* and *confident* in their ability to teach level C topics. PSMTs' confidence decreased as the statistical sophistication of the items increased. A repeated measures ANOVA determined that mean scores differed significantly between the three levels [$F(2,432)=60.04$, $p<0.001$]. Post hoc tests using a Bonferroni correction indicate that PSMTs scored significantly lower as the statistical sophistication of items increased ($p<0.001$).

Table 2. PSMTs' confidence scores for teaching statistics

	Number of items	Mean	SD
Overall Score	44	4.10	0.78
GAISE Levels			
Level A Score	11	4.54	0.79
Level B Score	15	4.12	0.82
Level C Score	18	3.80	0.89

Discussion and Conclusion

There are several findings of this study that are significant for statistics educators and mathematics teacher educators working with PSMTs. First, because of the increased emphasis of statistics in the U.S. high school curriculum, it is important that PSMTs are prepared to teach statistics. The results of this study provide insight on the current landscape of PSMTs' statistical knowledge and statistics teaching efficacy. PSMTs in this study were chosen from a purposeful sample of universities that had faculty members who have participated in programs to increase statistics education. Thus we expected that these might be receiving more emphasis on statistics in their mathematics teacher education programs than the average PSMT. However, these

PSMTs still generally do not have strong conceptual understandings of the statistics content they will be expected to teach and do not feel confident to teach statistics. Additionally, PSMTs' statistical understandings decrease as the investigative cycle progresses, that is, they are far less knowledgeable about analyzing data and interpreting results than they are about formulating questions and collecting data. Previous research has shown a similar trend with inservice teachers' and students' statistical knowledge measured by LOCUS (Jacobbe, 2015; Jacobbe, Foti, et al., 2014). Similar to statistical knowledge, PSMTs' statistics teaching efficacy decreased as topics get more sophisticated, being only *somewhat confident* to teach high school level statistics content. Thus, the experiences that PSMTs' are engaging in during the mathematics teacher education programs are not adequate in preparing PSMTs for all aspects of the statistical investigative cycle and especially with content at a GAISE Level C in sophistication.

This study had a number of limitations, which should be taken into consideration when interpreting the findings. First, this study was a purposeful sample on institutions across the U.S. and was not a random sample of all PSMTs. While most studies in teacher education that have explored PSMTs' content knowledge and/or teaching efficacy have been conducted at one institution or a small number of institutions, our study was conducted across 18 institutions varying in program size and location in the U.S., which may expand the potential usefulness of the results. However, of importance, is that our purposeful sample was from those institutions with at least one faculty member that had participated in a project aimed at increasing preparation of mathematics teachers to teach statistics. While we have no evidence of how the various teacher education programs actually attend to the preparation of secondary mathematics teachers to teach statistics, our results suggest that these efforts may not be having a strong impact on the current cohort of PSMTs represented in our study. This suggests that efforts are needed on a larger and more sustainable scale that can truly transform mathematics teacher education programs in ways that provide the needed preparation for the increased demands on high school teachers for teaching statistical content. The statistics education community needs to develop sustainable and large-scale models for infusing statistics and statistics teaching as a core component of all secondary mathematics teachers' preparation. There is a need to further examine *how* these specific mathematics teacher education programs are preparing their PSMTs to teach statistics. Such case studies could provide recommendations to make large-scale changes for mathematics teacher education programs across the country. Such large-scale changes that increase the emphasis on statistics in secondary mathematics teacher preparation are what is needed to stop the cycle of new mathematics teachers being unprepared to teach the statistical standards in today's curriculum.

References

Allison, P. D. (2002). *Missing data*. Thousand Oaks: SAGE.

Ashton, P. T. (1985). Motivation and the teacher's sense of efficacy. In C. Ames & R. Ames (Eds.), *Research on motivation in education* (Vol. 2, pp. 141-171). Orlando, FL: Academic Press.

Baldwin, W. (2000). Information no one else knows: The value of self-report. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 3-7). Mahwah, NJ: Erlbaum.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389-407.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Bandura, A. (1986). *Social foundations of thought and actions: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Batanero, C., Burrill, G., & Reading, C. (Eds.). (2011). *Teaching statistics in school mathematics-Challenges for teaching and teacher education: A joint ICMT/IASE study*. The Netherlands: Springer.

Bates, A. B., Latham, N., & Kim, J. A. (2011). Linking preservice teachers' mathematics self-efficacy and mathematics teaching efficacy to their mathematical performance. *School Science and Mathematics, 111*(7), 325-333.

Begg, A., & Edwards, R. (1999). *Teachers' ideas about teaching statistics*. Paper presented at the joint conference of the AARE & NZARE, Melbourne.

Browning, C., Goss, J., & Smith, D. (2014). Statistical knowledge for teaching: Elementary preservice teachers. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July 2014) Flagstaff, Arizona, USA*. Voorburg, The Netherlands: International Statistical Institute.

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-Challenges for teaching and teacher education* (pp. 57-69). The Netherlands: Springer.

Casey, S. A., & Wasserman, N. H. (2015). Teachers' knowledge about informal line of best fit. *Statistics Education Research Journal, 14*(1), 8-35.

Doerr, H., & Jacob, B. (2011). Investigating secondary teachers' statistical understandings. In M. Pytlak, T. Rowland, & E. Swoboda (Eds.), *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education* (pp. 776-786).

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Schaeffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) Report: A Pre-K-12 curriculum framework*. Alexandria, VA: American Statistical Association.

Hackett, G., & Betz, N. E. (1989). An exploration of the mathematics self-efficacy/mathematics performance correspondence. *Journal for Research in Mathematics Education, 20*(3), 261-273.

Hardin, J., Hoerl, R., Horton, N. J., Nolan, D., Baumer, B., Hall-Holt, O., . . . Ward, M. D. (2015). Data science in statistics: Preparing students to "think with data". *The American Statistician, 69*(4), 343-353.

Harrell-Williams, L., M., Sorto, M. A., Pierce, R., Lesser, L. M., & Murphy, T. J. (2014). Using the SETS instruments to investigate sources of variation in levels of pre-service teacher efficacy to teach statistics. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July 2014) Flagstaff, Arizona, USA*. Voorburg, The Netherlands: International Statistical Institute.

Harrell-Williams, L., M., Sorto, M. A., Pierce, R. L., Lesser, L. M., & Murphy, T. J. (2013). Validation of scores from a new measure of preservice teachers' self efficacy to teach statistics in the middle grades. *Journal of Psychoeducational Assessment, 20*(10), 1-11.

Hu, S. (2015). *Pre-service teachers' understanding of association and correlation in an inquiry-based introductory statistics course.* Paper presented at the Thirty-seventh annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, East Lansing, MI.

Jacobbe, T. (2015). Developing K-12 teachers' understanding of statistics. Retrieved from https://www.causeweb.org/webinar/teaching/2015-04/

Jacobbe, T., Case, C., Whitaker, D., & Foti, S. (2014). Establishing the validity of the LOCUS assessments through an evidenced-centered design approach. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July 2014) Flagstaff, Arizona, USA*. Voorburg, The Netherlands: International Statistical Institute.

Jacobbe, T., Foti, S., Case, C., & Whitaker, D. (2014). High school (ages 14-18) students' understanding of statistics. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July 2014) Flagstaff, Arizona, USA*. Voorburg, The Netherlands: International Statistical Institute.

Jones, D., & Tarr, J. E. (2010). Recommendations for statistics and probability in school mathematics over the past century. In B. J. Reys, R. E. Reys, & R. Reubenstein (Eds.), *Mathematics curriculum: Issues, trends, and future directions* (pp. 65-75). Reston, VA: National Council of Teachers of Mathematics.

Lee, H. S., Kersaint, G., Driskell, S., Harper, S. R., Jones, D., Leatham, K., . . . Kwaku, A.-G. (2014). Teachers' use of transnumeration in solving statistical tasks with dynamic statistical software. *Statistics Education Research Journal, 13*(1), 25-52.

Lovett, J. N., Doerr, H. M., Thrasher, E., & Lee, H. S. (under review). Developing teacher’s efficacy to teach statistics. *Submitted to Journal of Mathematics Teacher Education*.

Makar, K., & Confrey, J. (2004). Secondary teachers' reasoning about comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking*. Dordreht, The Netherlands: Kluwer Academic Publisher.

Makar, K., & Confrey, J. (2005). Variation talk: Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27-54.

National Governors Association Center for Best Practice & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington D.C.: Author.

Ross, S. H. (1989). Parts, wholes, and place value: A developmental view. *The Arithmetic Teacher, 36*(6), 47-51.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Sorto, M. A. (2006). Identifying content knowledge for teaching statistics *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil*. Voorburg, The Netherlands: International Statistics Institute.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple impuation using chained equations: Issues and guidance for practice. *Statistics in Medicine, 30*(4), 377-399.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-248.

Wilkins, J. L. (2008). The relationship among elementary teachers' content knowledge, attitudes, beliefs, and practices. *Journal of Mathematics Teacher Education, 11*(2), 139-164.