THE INFLUENCE OF A GRADUATE COURSE ON TEACHERS' SELF-EFFICACY TO TEACH STATISTICS

Emily Thrasher
North Carolina State University
epthrash@ncsu.edu

Tina Starling
North Carolina State University
ttstarli@ncsu.edu

Jennifer N. Lovett
North Carolina State University
jnickel@ncsu.edu

Helen M. Doerr
Syracuse University
hmdoerr@syr.edu

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

This paper explores the impact on teachers' self-efficacy to teach statistics from a graduate course aimed to develop teachers' knowledge of inferential statistics through engaging in data analysis using technology. This study uses qualitative and quantitative data from the Self-Efficacy to Teach Statistics Survey (Harrell-Williams et al., 2013) to provide data about teachers' confidence to teach statistical topics. The survey was given to 27 participants from two different institutions before and after the graduate course. We found that participants' self-efficacy to teach statistics increased after participation in the graduate course and references to specific course activities will be identified.

Keywords: Data Analysis and Statistics; Teacher Education-Inservice (Professional Development); Teacher Beliefs

Considerable research has addressed students' statistical thinking (Shaughnessy, 2007), and statistics continues to receive attention in the secondary US mathematics curriculum (NCTM, 2000; Common Core State Standards Initiative, 2010). However, there is a lack of research on secondary teachers' statistical reasoning and beliefs (Batanero, Burrill, & Reading, 2011). In fact, very little is known about teachers' self-efficacy to teach statistics (Harrell-Williams, Sorto, Pierce, Lesser, & Murphy, 2013). This study is situated in self-efficacy for teaching within a graduate course aimed at developing knowledge of the teaching and learning of statistics.

Researchers have investigated the statistical knowledge needed for teaching, using various frameworks (e.g., Groth, 2007). Each of these frameworks has identified teachers' own statistical reasoning as a foundational aspect of their ability to teach statistics. Thompson (1992) argues that researchers should not separate the study of teachers' beliefs from teachers' knowledge since they are intertwined. Thus this study aims to look at self-efficacy to teach as another component of teachers' readiness to teach statistical concepts to their students.

Background and Research Focus

Self-efficacy is often defined as "people's judgments of their capabilities to organize and execute courses of action required to designated types of performance" (Bandura, 1986, p.391). Self-efficacy to teach can be defined as a teacher's "belief to bring about student learning" (Ashton, 1985, p.142). Not only is self-efficacy to teach a central component of a teacher's beliefs (Greshman, 2008; Smith, 1996), it has been has been linked to positive influences on students' learning, the use of more innovative teaching strategies, and time spent teaching certain topics (e.g., Czerniak & Chiarelott, 1990). With these connections, it seems important to improve teachers' self-efficacy to teach. However, it has been suggested that it is hard to impact self-efficacy after teachers enter the classroom (e.g., Smith, 1996).

Bandura (1997) argued that there are four types of sources that may impact one's self-efficacy: mastery experiences, vicarious experiences, verbal persuasion, and physiological responses. For the purpose of this study, the focus is on how mastery experiences impact one's self-efficacy to teach. Mastery experiences are prior experiences in performing a task that are perceived to be a success

(Bandura, 1997). In terms of self-efficacy to teach there are two forms of mastery experiences: classroom teaching experiences and cognitive mastery (Palmer 2011). Arguably, classroom teaching experiences are the most crucial source of self-efficacy to teach because individuals can only assess their ability to teach by participating in the act of teaching (Tschannen-Moran, Hoy, & Hoy, 1998). Cognitive mastery refers to a teacher's perceived success in understanding the content and pedagogy to teach a specific topic (Palmer 2011). The cognitive mastery framework underpins our study to measure the development of self-efficacy to teach statistics from a graduate course aimed at developing subject matter knowledge and pedagogical content knowledge.

Our research is situated within the design and implementation of a graduate course that was largely influenced by Pfannkuch and Ben-Zvi's (2011) recommendations for designing experiences to develop teachers' statistical thinking, as well as the Guidelines for Assessment and Instruction in Statistics Education (GAISE) reports (Franklin et al., 2007; Garfield et al., 2007) and the Mathematical Education of Teachers II report (CBMS, 2012). Over two academic years, a team of four instructors from two institutions designed, and taught, a 15-week course which provided participants opportunities to develop a deeper understanding of a few statistical ideas. Two instructors taught at one university while the other two taught at the other university creating as similar of a course as possible at both institutions through continuous co-planning and reflection.

Throughout the semester-long course, participants implemented the cycle of statistical investigation (Friel, O'Connor, Mamer, 2006) as they engaged in with real data and tasks designed to develop their understandings of variation, distribution, samples and sampling distributions, and inferential statistics, especially randomization approaches using simulations. The course used dynamic software, *Fathom* (Finzer, 2005) and *TinkerPlots* (Konold & Miller, 2011), and online applets such as StatKey (lock5stat.com/statkey/). Assigned readings and discussions centered on (a) the nature of statistical reasoning and how it compares to mathematical reasoning, and (b) students' learning and reasoning related to the aforementioned topics. The software tools, new to most participants, were used to support their learning by allowing them to flexibly explore graphical representations, easily compute statistical measures, compare data sets, and make changes to data to explore conjectures. The software also provided simulation tools necessary to create representations of a population, a sample, and an empirical sampling distribution. This study addresses the following questions: 1) To what extent is teachers' self-efficacy to teach statistics changed from a graduate course focused on teaching and learning statistics? and 2) What learning experiences do teachers identify that influenced their self-efficacy to teach statistics?

Methodology

Participants

Participants came from all the teachers participating in either course across the two institutions. The course served a variety of graduate students (n=27). Participants consisted of one undergraduate pre-service teacher, six pre-service teachers in a masters program; 11 in-service teachers enrolled in a master's program; one full-time master's student in mathematics education; and eight doctoral students in mathematics or mathematics education. Twenty-one participants were female and six were male. Six participants indicated that English was their second language. Most participants had completed the equivalent of an undergraduate degree in mathematics, and all but two had at least one prior course in statistics. Hereafter we refer to course participants as teachers.

Data Collection and Analysis

To examine changes in teachers' self-efficacy to teaching statistics, the Self-Efficacy to Teach Statistics (SETS) survey was administered (Harrell-Williams, Sorto, Pierce, Lesser, & Murphy, 2014). This survey was chosen because it collects both qualitative and quantitative data about teachers' self-efficacy to teach statistics. Researchers argue both data sources are needed within the self-efficacy research (Wyatt, 2014). SETS was administered prior to the first day of class and during the last week of class. The SETS survey contains 44 six-point Likert scale items and six open response items that are aligned with the GAISE framework (Franklin et. al., 2007). An earlier version of this instrument 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). In addition to an overall score, the instrument provides sub-scale scores that correspond to Levels A-C in the GAISE framework. Although there are not explicit definitions given for each level in the GAISE report, each level is aligned to specific content. The content in level A is considered more concrete and level C is considered the most abstract. For example, in level A students are asked to compare groups without generalization while in level C students answer comparison questions and make generalizations (Franklin, 2007). There were 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). For the open-ended portion of SETS, in each GAISE level category, teachers were asked to identify an item which they felt least and most confident to teach to high school students and to explain their reasoning (total of six open-ended items).

The analysis of the SETS data was completed in two phases. The first phase focused on the responses to the 44 six-point Likert scale items. For both the pre- and post-survey, each teacher was given a total score calculated as the sum of his/her Likert scores. Sub-scale scores were also calculated for each teacher. The totals were divided by the number of items, which resulted in a final score that corresponded to the six-point Likert scale. Additionally, a gain score was calculated for each teacher as the difference of pre- and post- scores for each item. Means were computed for pre, post, and gain scores and a Wilcoxon Signed Rank Test for Matched Pairs was conducted to test for the significance from pre to post. Finally, the gain scores were averaged for each teacher and for each item. The item averages were examined for highest average gains in relationship to course content. Teachers with missing values within specific calculations were removed from the sample for that calculation.

The second phase of data analysis focused on analyzing the open-ended responses for ways in which the course influenced teachers' self efficacy to teach statistics. The code "course" identified responses that explicitly referred to specific course activities.

Results

First, we report the extent of the change in self-efficacy to teach statistics through the quantitative data from the pre and post SETS survey. Second, we report our findings from the qualitative responses to identify the course experiences that the teachers identified as influencing their self-efficacy to teach statistics.

Influence of Professional Development on Self-efficacy

We investigated the general influence of the course on self-efficacy to teach statistics by examining the mean scores for the pre and post survey and the mean gains by teacher and by item. Teachers began the course between somewhat confident (score of 3) and confident (score of 4) for each item; however, the teachers finished the course describing their self-efficacy to teach statistics as between confident and very confident (see Table 1). With the exception of one teacher, all teachers showed a positive average item gain in self-efficacy to teach statistics. The highest average item gain was 1.68 Likert points, which was recorded by two teachers. Figure 1 shows the distribution of average item gains by teacher. On average, teachers' self-efficacy for to teach statistics for each item increased one Likert point (0.95 increase). After accounting for missing values and using a Wilcoxon Signed Rank Test for Matched Pairs, the increase in self-efficacy between the pre and post surveys is considered statistically significant (see Table 2).

Table 1: Descriptive statistics of Likert scale items

		Pre	Post	Gain
Overall	N	26	24	23
	Mean	3.62	4.65	0.95
	Standard deviation	0.82	0.64	0.49
Level A topics	N	26	27	26
	Mean	3.95	5.10	1.14
	Standard deviation	0.80	0.51	0.55
Level B topics	N	27	26	26
	Mean	3.75	4.70	0.94
	Standard deviation	0.81	0.71	0.55
Level C topics	N	27	25	25
	Mean	3.39	4.35	0.97
	Standard deviation	0.96	0.74	0.63

Similar results were found for all three GAISE sub-scores. Level A had the highest pre score average of 3.95. This suggests that teachers started out confident in their ability to teach topics within that level. Interestingly, these topics are also the areas were teachers' confidence grew the most with a statistically significant (Wilcoxon signed rank test, p>0.001) average gain of 1.14. Level B had a pre survey mean score of 3.75 and a post survey mean of 4.70. Accounting for missing values, the average gain for level B was 0.94, which was also statistically significant. Finally level C started at the lowest confidence at 3.39, implying that most teachers on average feel only somewhat confident in their ability to teach statistics. The post mean score was 4.35, which is a growth in confidence to teach the level C topics in statistics. The average growth for level C was 0.97 points. Similarly, according to a Wilcoxon signed rank test this growth was statistically significant (Table 2). In

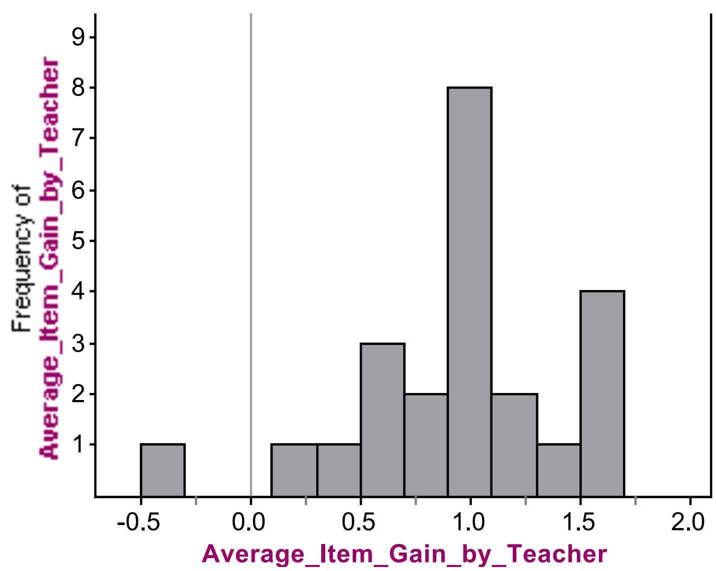

Figure 1: Distribution of Each Teachers' Average Likert Item Gain

Table 2: Wilcoxon Signed Rank Test for Matched Pairs

	Test Statistic	p-Value
Overall	4.14	0.000
Level A topics	4.44	0.000
Level B topics	4.46	0.000
Level C topics	4.07	0.000

addition to average gains at all levels, the standard deviation decrease at all levels indicating a decrease in variability in post confidence ratings.

Examining the average gain by item shed light on the specific content aligned to the teachers' growth in self-efficacy to teach statistics. The items that showed the greatest gain on average across teachers related directly back to the course goals: Item 44 (*Determine if the difference between two population means or proportions is statistically significant using simulations*) had an average gain of 1.85 and Items 9 (*Generalize a statistical result from a small group to a larger group*), 37 (*Evaluate whether a specified model is consistent with data generated from a simulation*), and 43 (*Compare two treatments from a randomized experiment by exploring numerical and graphical summaries of data*) all had an average gain in self-efficacy to teach of 1.42 points. All four of these items address the course foci of inferential statistics using randomization approaches, sampling distributions, and variation. The item with the lowest overall gain (0.48) was Item 33 (*Fit an appropriate model using technology for a scatterplot of two quantitative variables*), which was not a topic explicitly addressed during whole-class activities or discussions within the course.

Teachers' Reflection on Learning Experiences

In the open response items of the SETS instrument used at the end of the course, teachers identified course experiences when describing what, in each level, they felt most confident about. At both institutions, the course began with lengthy discussions on the cycle of statistical investigation (Friel et al., 2006). This cycle became a theme of the course as teachers gained repeated experience with posing statistical questions, collecting data, and data analysis and interpretation. Early in the course, teachers had opportunities to deepen their understanding of distribution through a series of tasks related to interpreting graphical representations. One such task asked teachers to match box plots to corresponding dotplots. This task revealed that a given boxplot could have underlying dotplot distributions that looked somewhat differently. About this activity, one teacher wrote

> "I feel most confident about working with box plots; the [activity] showed both the advantages and disadvantages of boxplots and [how] we can use them to describe data."

Based on research by delMas and Liu (2005), a second task teachers experienced was a game in *Fathom* designed to enhance teachers' conceptualization of the relationship between a distribution and its standard deviation. Teachers remembered this game at the end of the course. For example,

> "After doing the activity of "What Makes the Standard Deviation Larger or Smaller", I noticed a couple of patterns for justifying the characteristics of normal distributions with different centers, shapes, standard deviation, and so on."

As the course progressed, simulations became a means by which teachers developed understanding about variability and sampling distributions. The SETS item (44), which showed the greatest gain in self-efficacy focused on simulations. In the open response items, teachers remembered learning from the simulations with and without technology:

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

> "[Simulation] is something that we spent a lot of time on in the course. There are a lot of different ways to approach [it] with students such as hands-on simulations or technology simulations."

One hands-on simulation used physical devices, some of which did not have equiprobable outcomes. In the activity, each group had to describe a repeatable action that could produce an unpredictable result and the possible outcomes from this repeatable action. After the event(s) of interest was selected, for which results could be examined from the repeatable actions, each teacher in the group collected a sample of 10 trials. The activity continued with more samples being collected and a sampling distribution being created. While it is a familiar activity for statistics educators, it was not for the participating secondary teachers. One teacher wrote,

> "I liked the activity we did in class of having each person collect data for a sample of 10…I think I have a good conceptual understanding of the relationship between samples, distributions of samples, and populations."

The simulation focus continued as the course ended with randomization techniques. One teacher shared that she

> "already knew about randomization tests, but I feel more confident having multiple pieces of software that can perform the simulation for me. Before I was just using statcrunch and showing my students the output, but now I can actually have them do it!"

A specific course experience referenced in the SETS open responses was the Dolphin Therapy task (Rossman, 2008). This task required a re-randomization technique to test the difference of proportions. Teachers were given index cards to use in the design and simulation of the problem. Eventually, they used StatKey and TinkerPlots for a greater number of samples.

Another course experience that was highlighted by teachers in the open responses of the SETS survey was the course mid-term project. For the assignment, teachers self-assigned themselves to a working group. Each small group examined best practices for teaching learning a specific statistical topic. They applied research literature to create or adapt meaningful tasks and implemented one task with a group of students. Projects were shared through oral presentations and a course wiki. Topics for the midterm project included: Comparing Distributions, Sampling Techniques and Study Design, Sampling Distributions, Hypothesis Testing, Linear Regression/Covariation and Correlation, Using Probability to Make Decisions, Subjective Probability and Bayes Theorem.

The course experiences described above were ones specifically linked by teachers to content in which they felt most confident. In the survey, teachers were also asked to identify particular areas where they felt least confident. Mostly, teachers responded with comments such as *"these items were not specifically discussed in the course" or "I do not think I had a lot of practice … in the course."* Other times, however, teachers provided more insight into particular self-reported deficiencies (e.g. box plots, error, randomization, inference, sampling). Several teachers even suggested that they wanted more time with topics or would continue to refer to course materials to develop a deeper understanding. One teacher wrote,*"Sampling!!! I don't feel very confident teaching it yet. I began to develop a better internal understanding of it in class. I wish I could study it some more in a similar environment as was created in [my course]."* And, another teacher wrote, *"I am confident that randomization is highly important but I still second guess myself...Since I second guess myself, I am somewhat confident because I at least know that I have resources that I can reread."* Despite the fact that all teachers showed gains in self-efficacy overall, the open-ended responses provided details for instructors at each institution regarding potential pivotal experiences for teachers' own development of statistical understanding during the graduate course that seem to influence their statistics teaching efficacy.

Discussion and Conclusions

The results indicate that the course had a statistically significant positive impact on teachers' self-efficacy to teach statistics. These results were seen on the overall level and at all three GAISE levels. This suggests that a graduate course focused on the teaching and learning of statistics can impact a teachers' self-efficacy to teach statistics, and furthermore suggests that teachers can gain in self-efficacy to teach statistics from focusing on content knowledge and pedagogical content knowledge for teaching statistics. Additionally, our data also show that teachers have decreasing confidence to teach from level A to level C. This result holds for both the pre and post surveys and suggests that more abstract material corresponds to lower self-efficacy to teach. This result is similar to results found by Harrell-Williams et al., (2013) with pre-service teachers.

In addition to an overall gain in self-efficacy to teach statistics by the teachers, specific gains related to the course and its objectives were seen. After examining the data by SETS items, large gains were seen on topics related to the course objectives of inferential statistics, sampling distributions, and variation. Additionally, many teachers' mentioned course activities as reasons for their increase in self-efficacy to teach these topics. This speaks positively to the design of these activities and suggests that some course activities can have powerful impacts on teachers' confidence to teach statistics. These seem to be serving as a key mastery experiences.

However, not only did the areas that were emphasized in the course get impacted. There was an average increase on all items on the SETS survey including those that were not specifically stressed in the course. One possible source for this growth could be the course projects that allowed students to investigate topics of their choosing.

In general, these results point to specific activities that work to increase self-efficacy to teach statistics with teachers. Further research needs to be conducted to better understand what type of activities and how these activities are impacting teachers' self-efficacy to teach.

References

American Statistical Association. (2005). *Guidelines for assessment and instruction in statistics education: College report*. Alexandria, VA: Author. [Online: http://www.amstat.org/education/gaise/]

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.

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

Bandura, A. (1997). Self-efficacy: The exercise of control. New York: Freeman.

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

Conference Board of the Mathematical Sciences. (2012). *The mathematical education of teachers II*. Providence, RI: American Mathematical Society.

Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington D.C.: Author.

Czerniak, C. M., & Chiarelott, L. (1990). Teacher education for effective science instruction: A social cognitive persepective. *Journal of Teacher Education, 41*(1), 49-58.

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

Finzer, W. (2005). *Fathom Dynamic Data Software.* (Version 2.1) [Computer Software]. Emeryville, CA: Key Curriculum Press.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., et al. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) report: A Pre-K-12 curriculum framework*. Alexandria, VA: American Statistical Association. [Online: http://www.amstat.org/education/gaise/]

Friel, S. N., O'Connor, W., & Mamer, J. D. (2006). More than "meanmedianmode" and a bar graph: What's needed to have a statistical conversation? In G. Burrill (Ed.), *Thinking and reasoning with data and chance: Sixty-eighth yearbook* (pp. 117-137). Reston, VA: National Council of Teachers of Mathematics.

Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.

Greshman, G. (2008). Mathematics anxiety and mathematics teacher efficacy in elementary pre-service teachers. *Teaching Education, 19*(3), 171-184.
Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education, 38*(5), 427 - 437.
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.
Konold, C., & Miller, C. (2011). *TinkerPlots* [v. 2.0, computer software]
National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
Palmer, D. (2011). Sources of efficacy information in an inservice program for elementary teachers. *Science Education, 95*(4), 577-600.
Pfannkuch, M., & Ben-Zvi, D. (2011). Developing teachers' statistical thinking. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-Challenges for teaching and teacher education: A joint ICMI/IASE study*. Voorburg, The Netherlands: Springer.
Rossman, A. (2008). Reasoning about informal statistical inference: A statistician's view. *Statistics Education Research Journal, 7*(2), 5-19.
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.
Smith, J. P. (1996). Efficacy and teaching mathematics by telling: A challenge for reform. *Journal for Research in Mathematics Education, 27*(4), 387-402.
Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). New York: Macmillan.
Tschannen-Moran, M., Hoy, A. W., & Hoy, W., K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2), 202-248.
Wyatt, M. (2014). Towards a re-conceptualization of teachers' self-efficacy beliefs: Tackling enduring problems with the quantitative research and moving on. *International Journal of Research & Method in Education, 37*(2), 166-189.