USING THE SETS LEVEL A ITEMS TO CLASSIFY PRE-SERVICE TEACHERS' SELF-EFFICACY TO TEACH STATISTICS: AN APPLICATION OF THE MIXTURE RASCH MODEL

Leigh M. Harrell-Williams[1], Jennifer N. Lovett[2], M. Alejandra Sorto[3], Rebecca L. Pierce[4], Lawrence M. Lesser[5], Teri J. Murphy[6]
[1]University of Memphis, USA
[2]Middle Tennessee State University, USA
[3]Texas State University, USA
[4]Ball State University, USA
[5]The University of Texas at El Paso, USA
[6]University of Cincinnati, USA
Leigh.Williams@memphis.edu

Recent studies using the Self-Efficacy to Teach Statistics instruments found pre-service teachers (PSTs), as a group, exhibit higher self-efficacy for certain statistical topics among the GAISE Level A subscale items than for other statistical topics. Hence, this study explores latent classes for statistics teaching self-efficacy using Rasch mixture modeling of item-level response data. Four classes were identified for n = 588 middle and secondary mathematics PSTs. One class had consistently high efficacy; one had consistently low efficacy. Four items related to graphical displays contributed to differences between the middle two classes, contradicting previous findings that PSTs, on average, are more confident about teaching graphical displays than other topics. Hence, evaluation of statistics teaching self-efficacy should possibly include examination of specific content areas along with subscale scores.

BACKGROUND

The additional emphasis on statistics in grades 6-12 state mathematics standards in the United States has brought more attention to K-12 mathematics teacher preparedness and self-efficacy to teach statistics. Research on global and domain-specific teaching self-efficacy, such as mathematics and science, provides evidence that teaching self-efficacy influences classroom processes, student motivation and achievement, and teacher psychological well-being (Zee & Kooman, 2016). Guided by the American Statistical Association's Guidelines for Assessment and Instruction in Statistics Education (GAISE) Pre-K-12 document (Franklin et al., 2007) and state standards for mathematics in grades 6 through 12 (i.e., ages 10-18), the Self-Efficacy to Teach Statistics instruments (SETS-MS and SETS-HS; Harrell-Williams, Sorto, Pierce, Lesser, & Murphy, 2014a; Harrell-Williams, Lovett, Lee et al., 2017) were developed to be used in related research in statistics education.

Recent studies using the SETS instruments have explored what impacts pre-service mathematics teachers' (PSTs) self-efficacy to teach statistics (Harrell-Williams, Sorto, Pierce, Lesser, & Murphy, 2014b; Lovett, 2016), which concepts PSTs feel more efficacious about teaching (Harrell-Williams, Sorto, Pierce, Lesser, & Murphy, 2015; Lovett & Lee, 2017), and how PSTs' self-efficacy to teach statistics relates to PSTs' statistics content knowledge (Lovett, 2016). However, no study yet has investigated possible statistics teaching efficacy latent classes using the SETS items. This paper employed a Rasch mixture model analysis to explore how the model identifies and distinguishes between latent classes of PSTs using responses to the SETS-MS and SETS-HS Level A items.

PARTICIPANTS

The 588 PSTs in this study came from two separate studies using the SETS instruments. The first study involved "Middle Grades" pre-service teachers, with a convenience sample of 309 PSTs whose intended licensure/certification included at least one year from the middle grades (generally ages 11 – 14). Data were collected across four different large-sized public institutions of higher education in four states in the United States. Approximately 78% of the PSTs were female. Most self-identified as Caucasian (88%). The second study included 290 PSTs in secondary mathematics teacher education programs across 20 universities in 14 states in the U.S. The PSTs

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

were predominantly female (70.3%). Approximately 82% self-identified as Caucasian. The percentages follow current teacher demographics in the United States (Taie & Goldring, 2017).

INSTRUMENTS: SETS-MS AND SETS-HS

There are two grade-level specific SETS instruments. The SETS-MS (Harrell-Williams et al., 2014a), focusing on middle grades, has 26 items that ask teachers to rate their confidence to teach students skills necessary to complete specific statistics tasks, using a 6-point Likert scale, with 1 = *not at all confident* and 6 = *completely confident*. The items are divided into two subscales relating to the levels of the GAISE Pre-K-12 document (Franklin et al., 2007): “Reading the Data (Level A)” and “Reading Between the Data (Level B).” The SETS-HS (Harrell-Williams, Lovett, Lee et al., 2017) includes the same 26 items as the SETS-MS and an additional 18 items related to the two data analysis strands of the High School Common Core State Standards for Mathematics (National Governors Association and Council of Chief State School Officers, 2010). These additional 18 items create a third subscale for the SETS-HS: “Reading Beyond the Data (Level C).” For this study, the focus is on the first 11 items on both SETS instruments, comprising the “Reading the Data (Level A)” subscale. These items cover topics such as recognizing variability in data and results, collecting data to answer a specific statistical question, generalizing from a smaller group to a larger group, creating graphical displays to summarize distributions, and using features identified on graphical displays to describe distributions or relationships. The SETS-MS items are listed in Harrell-Williams et al. (2015) and the SETS-HS items are listed in Harrell-Williams, Lovett, Lee et al. (2017).

ANALYSIS

Previous analyses using the combined sample showed that Level A subscale score means were not significantly different across middle and high school PSTs but item responses differed as middle grade teachers were more likely to use the higher response categories (5 or 6) with higher frequency than the high school PSTs (Harrell-Williams, Lovett, Pierce et al., 2017). Hence, a more thorough item-level approach to profile analysis to explore differences seemed necessary. Rasch mixture modeling combines latent class analysis with item and person parameter estimation, providing finer grained item-level detail than traditional cluster or latent class analysis. The mixture Rasch model analysis in this study employed the rating scale option in the Winmira software (von Davier, 2000), fitting models with one to six latent classes. Since there is no definitive rule regarding the number of classes to fit, six latent classes were chosen as a stopping point as model fit worsened with the addition of each additional class beyond six. Akaike’s Information Criteria (AIC; Akaike, 1973) and Bayesian Information Criteria (BIC; Schwarz, 1978) were used to determine the optimal number of classes, with the smallest value indicating the optimal number. Model results were used to compare differences in item responses across the PSTs using visual inspection of item difficulty estimates for each profile. Exploration of the frequency of use of each of the six response categories for each item showed that the lowest response category (“not at all confident”) was used as a response fewer than 10 times for seven of the 11 items, creating an issue with sparse data. Hence, the two lowest categories (“not at all confident” and “only a little confident”) were combined for this analysis.

MIXTURE LATENT CLASS MEMBERSHIP

Both AIC and BIC had the smallest value for the four-class model, indicating this as the optimal number of classes (Table 1). The posterior probabilities for class membership ranged from .926 to .964 (first four rows of Table 2), indicating a high certainty regarding class membership. Each of the four classes contained a reasonable number of PSTs (fifth row in Table 2), with a mix of middle and secondary in each latent class (bottom two rows in Table 2).

Table 1. Model Selection Criteria Results for Each Latent Class Model

	1 Class	2 Classes	3 Classes	4 Classes	5 Classes	6 Classes
AIC	18892	15072	14659	14568	14600	15143
BIC	18953	15208	14865	14843	14945	15532

Table 2. Posterior Probabilities, Overall Latent Class Size, and Class Breakdown by Teacher Type

	Membership in Latent Classes			
Prior Probabilities of Class Membership	Class 1	Class 2	Class 3	Class 4
Class 1 Prior Probabilities	.964	.032	.003	.001
Class 2 Prior Probabilities	.043	.926	.028	.003
Class 3 Prior Probabilities	.004	.044	.931	.021
Class 4 Prior Probabilities	.009	.014	.048	.929
Class Size - Overall Sample	184 (31%)	182 (31%)	148 (25%)	74 (13%)
Middle Grades PSTs	76 (42%)	103 (57%)	80 (53%)	26 (35%)
Secondary PSTs	108 (58%)	79 (43%)	68 (47%)	49 (65%)

Note: Overall sample percentages represent percentage of n=588 PSTs. Middle grades and secondary percentages represent the percent within each latent class.

MIXTURE ITEM PARAMETER ESTIMATES

Lower values for item location parameters indicate that the item was easier to endorse with higher response categories (such as "5" or "6") indicating higher confidence about teaching skills related to a specific statistics task. Higher values indicate that the item was easier to endorse with lower categories (such as a "1" or "2"). As shown in Figure 1, one latent class (Class 4, yellow line at top of figure) had consistently high efficacy while another (Class 1, blue line at bottom of figure) had consistently low efficacy. Item location parameters indicated that the four items related to creating/using graphical displays (items 4-7) contributed to differences as large as 1.5 logits between the middle two latent classes, with Class 3 (green line, middle of figure) having higher efficacy on these four items than Class 2 (red line, middle of figure).

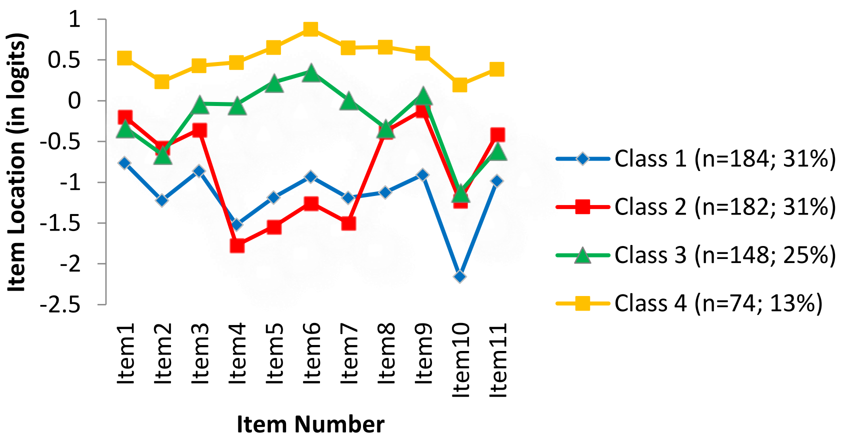

Figure 1. Difficulty Item Parameter Estimates by Latent Class

Note: A lower item location estimate for a latent class corresponds to higher self-efficacy ratings.

DISCUSSION

While previous studies found that items 4-7 were among the items that PSTs rate with higher efficacy (Harrell-Williams et al., 2015; Harrell-Williams, Lovett, Lee et al., 2017), this study indicates a subclass of teachers exists who do not rate these items highly. While this study used PSTs only from the United States, a similar distinction might be found among PSTs from other countries. Similar to Harrell-Williams et al. (2015) and Harrell-Williams, Lovett, Lee et al. (2017), Watson (2001) found that, overall, secondary teachers had higher confidence to teach measures of central tendency and graphical representations such as bar graphs. In terms of implications for teacher preparation and professional development programs, the results indicate that those using the SETS-MS or SETS-HS instruments for course and program evaluation purposes should look beyond just the current GAISE level subscores for more information regarding teaching efficacy for specific statistical topics. Additionally, future SETS measurement work should include evaluation of potential statistical content area subscales, such as graphical displays.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov & F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Budapest: Akademiai Kiado.

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

Harrell-Williams, L.M., Lovett, J.N., Lee, H.S., Pierce, R.L., Lesser, L.M., & Sorto, M.A. (2017). Validation of scores from the high school version of the Self-Efficacy to Teach Statistics (SETS-HS) instrument using pre-service mathematics teachers. *Journal of Psychoeducational Assessment*. Advanced online publication. doi: 10.1177/0734282917735151

Harrell-Williams, L.M., Lovett, J.N., Pierce, R.L., Sorto, M.A., Lee, H.S., & Lesser, L.M. (2017). The middle grades SETS instrument: Psychometric comparison of middle and high school pre-service mathematics teachers. In E. Galindo & J. Newton (Eds.), *Proceedings of the 39th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1064-1067). Indianapolis, IN: Hoosier Association of Mathematics Teacher Educators.

Harrell-Williams, L.M., Sorto, M.A., Pierce, R.L., Lesser, L.M., & Murphy, T.J. (2014a). Validation of scores from a measure of teachers' self-efficacy to teach middle grades statistics. *Journal of Psychoeducational Assessment, 32*(1), 40-50.

Harrell-Williams, L.M., Sorto, M.A., Pierce, R.L., Lesser, L.M., & Murphy, T.J. (2014b). 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. (2015). Identifying statistical concepts associated with high and low levels of self-efficacy to teach statistics to middle grades. *Journal of Statistics Education, 23*(1), 1–20.

Lovett, J.N. (2016). The preparedness of preservice secondary mathematics teachers to teach statistics: A cross-institutional mixed methods study. Unpublished doctoral dissertation. NC State University. Raleigh, NC.

Lovett, J.N., & Lee, H.S. (2017). New standards require teaching more statistics: Are preservice secondary mathematics teachers ready? *Journal of Teacher Education, 68*(3), 299-311.

National Governors Association and Council of Chief State School Officers (2010). *Common Core State Standards for Mathematics.* Washington, D.C.: Author.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461-464.

Taie, S., & Goldring, R. (2017). *Characteristics of public elementary and secondary school teachers in the United States: Results from the 2015–16 National Teacher and Principal Survey First Look* (NCES 2017-072). Washington, DC: National Center for Education Statistics.

von Davier, M. (2000). Winmira 2001 – A Microsoft Windows program for analyses with the Rasch model, with the latent class analysis and with the mixed Rasch model [Computer Software].

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.

Zee, M., & Koomen, H.Y. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research, 86*(4), 981-1015.