Research Questions

(1) Does a 5-point bipolar scale ("Strongly Agree" to "Strongly Disagree") or a 4- or 5-point unipolar scale ("Completely Prepared" to "Not at All Prepared") provide a better measure of the construct of preparedness for teaching?

(2) Is the free (or two-parameter) graded response model (GRM) or the equal a (one-parameter) GRM a better fitting measurement model for each of the instruments data sets?

(3) How could the program completers survey be improved based on item level analysis?

Research Method

Both the 2015-16 and 2016-17 surveys were administered as web-based surveys to all teacher candidates who completed their teacher education program in the semester of the survey’s administration. For the 2015-16 academic year, 155 of the 229 program completers (68%) completed the survey. For the revised 2016-17 academic year survey, 114 of the 194 program completers (59%) completed the survey.

Univariate, bivariate, and unidimensionality analysis with SPSS v22

1. Univariate, reliability, & unidimensionality analysis with SPSS v22

2. Unidimensional item response theory (IRT) utilizing Samejima’s Graded Response Model (GRM) with free a and equal a parameters with IRTPRO v4

3. Item level analysis to inspect item functioning

Results

Both the 2015-16 and 2016-17 survey scales indicated strong reliability (Cronbach’s alpha: 0.915 and 0.952, respectively) and both were within acceptable values for meeting the assumptions of unidimensionality and local independence.

Free a vs Equal a Models: For both the 2015-16 and 2017, the free a (two-parameter) model was the better fitting model. For the 2015-16 survey, only the RMSEA value for the free a model indicated goodness of fit (RMSEA = 0.80). Therefore, the free a model was selected as the best fitting model. For the 2017 survey, the RMSEA values for both models indicated goodness of fit (RMSEA = 0.8). The GRM free a (two parameter) model had a better fit than the GRM equal a (one parameter) model as shown by lower MIA, AIC, and BIC values in comparison with those of the equal a model.

Considerations for Writing Survey Items

Is the question too demanding? (i.e. Is the question long or potentially confusing?)

How could the program completers survey be improved based on item level analysis?

The use of surveys of current students and program completers in order to evaluate the outcomes and effectiveness of programs is a well-established method for program improvement and accreditation. In response to these newly developed CQP specifications, the researcher at the College of Education (COE), University of Hawai’i at Mānoa, revised the COE program completers survey items from a bipolar agreement 5-point scale to a unipolar preparedness 4-point scale. This study aimed to address the issue of accreditation mandated survey scale specifications and provide evidence addressing the justification for such mandates.

Additional Item Analysis

2015-16 Survey: The least discriminating items were 7 (α = 1.99) and 14 (α = 1.91). The most discriminating items were 9 (α = 5.06) and 23 (α = 4.80). The most difficult item to endorse was 9 whose b values started at b1 = -1.24, followed by item 23 whose b values start at b1 = -1.74. The least difficult item to endorse b1 was 17 (b = 3.34). Item 16 had low probability, factor loading, and discrimination values.

2016-17 Survey: The least discriminating items were 15 (α = 1.22) and 17 (α = 1.33). The most discriminating items were 4 (α = 5.99) and 20 (α = 5.52). The most difficult item to endorse was 10 whose b values started at b1 = -1.71, followed by item 8 whose b values at b1 = -1.85. The least difficult item to endorse b1 was 15 (b = 3.71). As with the 2015-16 results, item 16 had low probability, factor loading, and discrimination values.

Findings & Implications

Results indicated that, overall, the 5-point scale was a better measure of the construct of preparedness to teach than the 4-point scale; however, next step would be to investigate if a 3- or 4-point unipolar preparedness scale, as opposed to the 5-point bipolar agreement scale provides further improvement to the instrument. These findings provide evidence that the CQP specification to use even numbered scales may not provide better measurement. Note: In January 2017, COE again changed their criteria requirements for survey instruments and eliminated the specification of how many response options should be in survey scales used as accreditation evidence.

The free a (two-parameter) model, including both item discrimination and item difficulty, provided a better fit than the equal a model for both the old and revised surveys.

How to improve the instrument: Add items that measure preparedness for program completers with higher beta values (higher reported preparedness). Further investigate item functioning and whether the use of an alternative wording for item 16 could improve the quality of the items as results indicated that it was not functioning well on both versions of the survey.

Accreditation-Driven Research: Psychometric Analysis of Program Completer Surveys

Background

The use of surveys of current students and program completers in order to evaluate the outcomes and effectiveness of programs is a well-established practice in teacher education. In 2016, the Council for the Accreditation of Educator Preparation (CAEP) began specifying criteria requirements for survey instruments, including the number of response options in scales, developed for teacher preparation programs seeking accreditation. In response to these newly developed CQP specifications, the researcher at the College of Education (COE), University of Hawai’i at Mānoa, revised the COE program completers survey items from a bipolar agreement 5-point scale to a unipolar preparedness 4-point scale. This study aimed to address the issue of accreditation mandated survey scale specifications and provide evidence addressing the justification for such mandates.